

ABSTRACTS

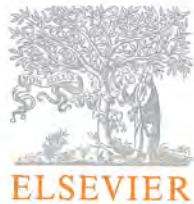
MAY 15- 19. 2026

VSS
2026

ST. PETE BEACH, FLORIDA



THANK YOU TO OUR SPONSORS



VPiXX Technologies Inc.
www.vpixx.com



frontiers
in Cognition



Rogue Research Inc.



Adaptive
Sensory
Technology





VIEWPixx3

Research Essentials



VPixx Technologies proudly presents our fully integrated research system:

- NEW VIEWPixx3 /OLED Display
- TRACKPixx3 2 kHz Eye Tracker
- RESPONSEPixx Button Box
- LabMaestro Experiment Builder
- LCD Console Monitor
- X-Rite i1Display Pro3
- Trigger Cable
- Chinrest

Providing tools for precision visual neuroscience for over 25 years.
Visit us at vpixx.com/research-essentials to learn more.



The Future of EEG is Here

Introducing the
**Geodesic EEG System 500
GES 500**

**Elevate your discoveries with the
industry's leading EEG technology**



The future of EEG is here, introducing EGI's transformative new HD-EEG platform. An intelligent eco-system for precision brain stimulation, discovery and advancing care.

Next-generation platform seamlessly integrates versatile hardware and intuitive software for EEG, TMS and other technologies, improving ease of use and allowing new brain stimulation innovations.

EGI is a global leader and gold standard in high density EEG, offering unmatched data quality and comfort.



Scan to
find out more

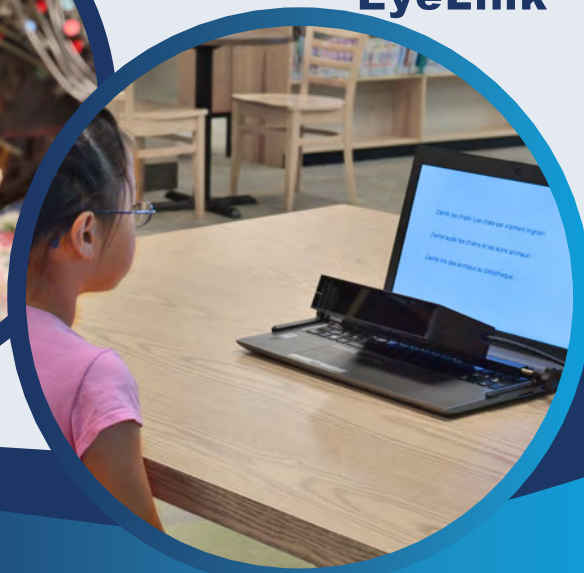
egi.com
magstim.com

Four neurotechnology brands.
One shared mission.



welcony





SR Research

EyeLink® 3

*Eye and Head-Tracking
in One*

- Full six-degrees of freedom head tracking data: X, Y, and Z position coordinates and pitch, yaw, and roll rotational angles
- Eye and head data synchronized up to 1000 Hz
- High frame rate (up to 500 FPS) eye video recordings
- Large trackable area
- Novel pixel-based eye and head data
- Integrates with Experiment Builder and Data Viewer

EyeLink® 1000 PLUS

*Multiple Eye-Tracking
Solutions in One*

- Remote, head free-to-move and head-stabilized eye tracking
- Multiple mounts for specialized environments
- Extend use to fMRI, MEG, EEG, TMS, & more
- Track infants, patients, seniors, & non-human primates
- Portable options to track off-site in hospitals, schools, homes, etc.
- Outstanding technical specifications - world leading

Fast, Accurate, Reliable Eye Tracking

EyeLink® PORTABLE DUO

*Portable Eye Tracking
Solution*

- Everything fits in a carry-on bag
- Quick and simple set up and calibration
- Dual-use head-stabilized or remote head free-to-move tracking with any display - laptop, desktop, projectors...
- Robust, accurate, and precise data for even the most demanding scientific and medical applications
- Smoothly integrates with existing software and hardware tools

www.sr-research.com
 info@sr-research.com
 +1-866-821-0731
 +1-613-271-8686

EyeLink is a registered trademark of SR Research Ltd.

Schedule Overview

Wednesday, May 13, 2026 ▲			
9:00 am - 6:00 pm	Computational and Mathematical Models in Vision (MODVIS)	Satellite	Blue Heron
Thursday, May 14, 2026 ▲			
9:00 am - 6:00 pm	Computational and Mathematical Models in Vision (MODVIS)	Satellite	Blue Heron
4:00 - 7:00 pm	Registration Open	Registration	Grand Palm Colonnade
Friday, May 15, 2026 ▲			
7:00 - 8:00 am	Morning Coffee & Continental Breakfast	Break	Garden Courtyard
7:00 am - 6:00 pm	Registration Open	Registration	Grand Palm Colonnade
7:00 am - 10:00 pm	VSS Quiet Lounge	Lounge	Compass
7:00 am - 10:00 pm	VSS Social Lounge	Lounge	Royal Tern
8:00 - 10:00 am	Intuitive physical reasoning in brains, minds, and machines	Symposium	Talk Room 1
8:00 - 10:00 am	Rhythms of Vision: How Neural Oscillations Structure Perception and Attention	Symposium	Talk Room 2
10:00 - 10:30 am	Coffee Break	Break	Garden Courtyard
10:30 am - 12:30 pm	What makes a task naturalistic? Towards a theory-inspired approach to studying visually guided behaviour	Symposium	Talk Room 1
10:30 am - 12:30 pm	Reimagining the binding problem(s) for the 21st century	Symposium	Talk Room 2
12:30 - 1:45 pm	Lunch (on your own) <i>Lunch Available for Purchase in Garden Courtyard 11:30 - 2:30</i>	Break	Garden Courtyard
12:45 - 1:45 pm	US Funding Workshop	Workshop	Banyan/Citrus
1:45 - 3:45 pm	Beyond Prediction: Interpretable Brain Encoding Models for Understanding Vision	Symposium	Talk Room 1
1:45 - 3:45 pm	The quest for 'the average person' in vision science: promise and pitfalls	Symposium	Talk Room 2
3:45 - 4:15 pm	Coffee Break	Break	Garden Courtyard
3:45 - 6:00 pm	Friday Afternoon Posters Visual Working Memory: Performance, influences Multisensory Processing: Motor Multisensory Processing: Visual-tactile Eye Movements: Cognition Eye Movements: Pursuit, vergence Face and Body Perception: Social cognition 1 Temporal Processing: Duration and timing perception Functional Organization of Visual Pathways: Cortical visual processing 1	Poster Session	Banyan Breezeway
4:15 - 5:45 pm	Attention 1	Talk Session	Talk Room 1
4:15 - 5:45 pm	Perceptual Organization: Neural mechanisms	Talk Session	Talk Room 2
6:00 - 7:30 pm	Lemonade and a Scavenger Hunt: A Newcomer and Undergrad Social	Social	Banyan/Citrus
7:30 - 9:30 pm	Opening Night Reception	Social	RumFish Beach
Saturday, May 16, 2026 ▲			
7:00 am - 10:00 pm	VSS Quiet Lounge	Lounge	Compass

7:00 am - 10:00 pm	VSS Social Lounge	Lounge	Royal Tern
7:30 am - 6:45 pm	Registration Open	Registration	Grand Palm Colonnade
7:45 - 8:45 am	Morning Coffee & Continental Breakfast	Break	Garden Courtyard & Pavilion
8:00 am - 5:30 pm	Exhibits Open	Exhibits	Pavilion
8:15 - 9:45 am	Eye Movements: Neural dynamics, perceptual changes	Talk Session	Talk Room 1
8:15 - 9:45 am	Perceptual Organization: Segmentation and grouping	Talk Session	Talk Room 2
8:30 am - 12:30 pm	Saturday Morning Posters Perceptual Training, Learning and Plasticity: Psychophysics Visual Working Memory: Interactions with long-term memory Visual Working Memory: Interference, attention Scene Perception: Models, natural image statistics Scene Perception: Intuitive physics Undergraduate Just-In-Time 1	Poster Session	Banyan Breezeway
8:30 am - 12:30 pm	Saturday Morning Posters Object Recognition: Reading 3D Shape and Space Perception: Cues, integration 3D Shape and Space Perception: Surfaces, objects Attention: Neural Functional Organization of Visual Pathways: Cortical visual processing 2 Face and Body Perception: Models Decision Making: Perception 1 Action: Navigation, locomotion Action: Miscellaneous	Poster Session	Pavilion
9:45 - 10:30 am	Coffee Break <i>Sponsored by The Center for Visual Science (CVS)</i>	Break	Garden Courtyard & Pavilion
10:45 am - 12:30 pm	Spatial Vision: Neural mechanisms, modeling	Talk Session	Talk Room 1
10:45 am - 12:30 pm	Face and Body Perception: Mechanisms	Talk Session	Talk Room 2
12:30 - 2:30 pm	Lunch (on your own) <i>Lunch Available for Purchase in Garden Courtyard 11:30 - 2:30</i>	Break	Garden Courtyard
12:30 - 2:30 pm	Canadian Vision Science Social	Satellite	Palm/Sabal/Sawgrass
12:45 - 2:15 pm	Virtual Reality & Eye Tracking for Research <i>Organized by WorldViz VR</i>	Satellite	Blue Heron
12:45 - 2:15 pm	AI in Science: Opportunities and Boundaries <i>Organized by the VSS Student-Postdoc Advisory Committee (SPC)</i>	Workshop	Banyan/Citrus
2:30 - 4:15 pm	Functional Organization of Visual Pathways 1	Talk Session	Talk Room 1
2:30 - 4:15 pm	Temporal Processing	Talk Session	Talk Room 2
2:45 - 6:45 pm	Saturday Afternoon Posters Spatial Vision: Clinical Perceptual Training, Learning and Plasticity: Rehabilitation Visual Search: Search strategies, clinical Visual Search: Attention, memory, decision-making Pre-Data-Collection Poster Session 1	Poster Session	Banyan Breezeway
2:45 - 6:45 pm	Saturday Afternoon Posters Attention: Endogenous, exogenous Attention: Features, objects	Poster Session	Pavilion

	Perceptual Organization: Features, parts, wholes, objects Visual Memory: Objects, features Color, Light and Materials: Material perception Color, Light and Materials: Affect, cognition Theory Multisensory Processing: Recalibration, temporal		
4:15 - 5:00 pm	Afternoon Coffee & Snack	Break	Garden Courtyard & Pavilion
5:15 - 7:00 pm	Scene Perception	Talk Session	Talk Room 1
5:15 - 7:00 pm	Visual Working Memory: Influences, models	Talk Session	Talk Room 2
7:30 - 8:30 pm	Keynote Address given by Michael Chiang AI, data science, vision research, and eye care: perspectives from the National Eye Institute	Keynote	Talk Room 1-2
8:30 - 9:30 pm	Dan Ts'o Commemoration	Social	Pirate Island
Sunday, May 17, 2026 ▲			
7:00 am - 10:00 pm	VSS Quiet Lounge	Lounge	Compass
7:00 am - 10:00 pm	VSS Social Lounge	Lounge	Royal Tern
7:30 am - 6:45 pm	Registration Open	Registration	Grand Palm Colonnade
7:45 - 8:45 am	Morning Coffee & Continental Breakfast	Break	Garden Courtyard & Pavilion
8:00 am - 5:30 pm	Exhibits Open	Exhibits	Pavilion
8:15 - 9:45 am	Attention 2	Talk Session	Talk Room 1
8:15 - 9:45 am	Object Recognition	Talk Session	Talk Room 2
8:30 am - 12:30 pm	Sunday Morning Posters Color, Light and Materials: Adaptation, contrast, lightness, brightness Perceptual Organization: Ensembles Perceptual Organization: Individual differences, aesthetics Scene Perception: Neural mechanisms Scene Perception: Spatiotemporal factors Undergraduate Just-In-Time 2	Poster Session	Banyan Breezeway
8:30 am - 12:30 pm	Sunday Morning Posters Functional Organization of Visual Pathways: Subcortical, clinical Functional Organization of Visual Pathways: Neuroimaging 3D Shape and Space Perception: Miscellaneous Eye Movements: Individual differences, visual preference Eye Movements: Mechanisms, perception, fixational Eye Movements: Saccades Face and Body Perception: Wholes, parts, configurations, features Decision Making: Perception 2	Poster Session	Pavilion
9:45 - 10:30 am	Coffee Break	Break	Garden Courtyard & Pavilion
10:45 am - 12:30 pm	Multisensory Processing	Talk Session	Talk Room 1
10:45 am - 12:30 pm	Visual Working Memory: Space, features	Talk Session	Talk Room 2
12:30 - 2:30 pm	VISxVISION Workshop	Satellite	Banyan/Citrus
12:30 - 2:30 pm	Lunch (on your own) <i>Lunch Available for Purchase in Garden Courtyard 11:30 - 2:30</i>	Break	Garden Courtyard

12:30 - 2:30 pm	Neurodiversity in Academia: A "Fireside Chat" with Brian Anderson	Networking	Palm/Sabal/Sawgrass
12:45 - 2:15 pm	In-silico replications	Satellite	Blue Heron
12:45 - 2:15 pm	Visual Snow Meet-Up	Satellite	Horizons West
2:30 - 4:15 pm	Development	Talk Session	Talk Room 1
2:30 - 4:15 pm	Face and Body Perception: Emotion, identity	Talk Session	Talk Room 2
2:45 - 6:45 pm	Sunday Afternoon Posters Visual Memory: Long-term memory Visual Memory: Encoding and retrieval, capacity Spatial Vision: Natural images, texture Spatial Vision: Neural mechanisms Spatial Vision: Crowding, eccentricity	Poster Session	Banyan Breezeway
2:45 - 6:45 pm	Sunday Afternoon Posters Color, Light and Materials: Neural mechanisms Eye Movements: Models, remapping Eye Movements: Pupillometry Visual Search: Features, scenes, real-world stimuli Visual Search: Neural mechanisms, models, eye movements Attention: Inattention, attentional blindness Attention: Spatial Attention: Individual differences	Poster Session	Pavilion
4:15 - 5:00 pm	Afternoon Coffee & Snack <i>Sponsored by Neurobridge</i>	Break	Garden Courtyard & Pavilion
5:15 - 7:15 pm	Perceptual Training, Learning and Plasticity	Talk Session	Talk Room 1
5:15 - 7:15 pm	Action	Talk Session	Talk Room 2
7:30 - 9:00 pm	CSHL Vision Courses Meet-up	Satellite	Snowy Egret
7:30 - 9:00 pm	FoVea Workshop: What Is This Job Anyway?	Satellite	Banyan/Citrus
Monday, May 18, 2026 ▲			
7:00 am - 2:30 pm	VSS Quiet Lounge	Lounge	Compass
7:00 am - 2:30 pm	VSS Social Lounge	Lounge	Royal Tern
7:45 - 8:45 am	Morning Coffee & Continental Breakfast	Break	Garden Courtyard & Pavilion
7:45 am - 2:30 pm	Registration Open	Registration	Grand Palm Colonnade
8:00 am - 12:30 pm	Exhibits Open	Exhibits	Pavilion
8:15 - 9:45 am	Attention 3	Talk Session	Talk Room 1
8:15 - 9:45 am	3D Shape and Space Perception	Talk Session	Talk Room 2
8:30 am - 12:30 pm	Monday Morning Posters Object recognition: Categories Object Recognition: Features, parts Face and Body Perception: Social cognition 2 Face and Body Perception: Individual differences Face and Body Perception: Bodies Functional Organization of Visual Pathways: Retinotopy, population receptive fields Pre-Data-Collection Poster Session 2	Poster Session	Banyan Breezeway
8:30 am - 12:30 pm	Monday Morning Posters Spatial Vision: Binocular vision	Poster Session	Pavilion

	Eye Movements: Natural, complex tasks Eye Movements: Clinical Action: Reaching Action: Grasping, affordances Decision Making: Actions, metacognition Motion: Illusions Motion: Optic flow, in-depth, biological, higher-order Perceptual Training, Learning and Plasticity: Statistical learning Perceptual Training, Learning and Plasticity: Category learning		
9:45 - 10:30 am	Coffee Break	Break	Garden Courtyard & Pavilion
10:45 am - 12:15 pm	Visual Search: Attention, mechanisms, models	Talk Session	Talk Room 1
10:45 am - 12:15 pm	Visual Memory: Encoding, retrieval, capacity	Talk Session	Talk Room 2
12:30 - 2:00 pm	VSS Awards Session Davida Teller Award, Ken Nakayama Medal, and Young Investigator Award	Award	Talk Room 2
2:00 - 3:00 pm	3MT Competition for Students and Postdocs	Student	Offsite
2:00 - 3:30 pm	Demystifying Peer Review: Editorial Insight and Perspectives <i>Organized by the VSS Student-Postdoc Advisory Committee (SPC)</i>	Workshop	Banyan/Citrus
2:00 - 5:30 pm	re:vision: A community-driven replication initiative for condition-rich neuroscience of visually-presented stimuli	Satellite	Blue Heron
2:00 - 6:00 pm	PURSUE Workshop	Satellite	Horizons West
3:00 - 4:00 pm	Public Lecture: Bevil Conway What is beauty?	Other	Offsite
4:00 - 5:30 pm	Meet the Professors	Student	Banyan Breezeway
7:00 - 10:00 pm	Demo Night Sponsored by Meta	Social	Talk Room 1-2
Tuesday, May 19, 2026 ▲			
7:00 am - 10:00 pm	VSS Quiet Lounge	Lounge	Compass
7:00 am - 10:00 pm	VSS Social Lounge	Lounge	Royal Tern
7:45 - 8:45 am	Morning Coffee & Continental Breakfast	Break	Garden Courtyard & Pavilion
7:45 am - 7:30 pm	Registration Open	Registration	Grand Palm Colonnade
8:00 am - 5:30 pm	Exhibits Open	Exhibits	Pavilion
8:15 - 9:45 am	Visual Search: Real-world scenes, objects	Talk Session	Talk Room 1
8:15 - 9:45 am	Functional Organization of Visual Pathways 2	Talk Session	Talk Room 2
8:30 am - 12:30 pm	Tuesday Morning Posters Perceptual Training, Learning and Plasticity: Neural mechanisms Perceptual Training, Learning and Plasticity: Neuroimaging, neurostimulation Perceptual Organization: Intuitive physics Perceptual Organization: Grouping Scene Perception: Ensemble Scene Perception: Virtual reality Scene Perception: Categorization, memory Visual Working Memory: Models, neural	Poster Session	Banyan Breezeway

8:30 am - 12:30 pm	Tuesday Morning Posters Attention: Capture 1 Attention: Models Temporal Processing: Neural mechanisms, models Motion: Mechanisms, models Action: Pointing, tracking Action: Perception, recognition Multisensory Processing: Cross-modal interactions Face and Body Perception: Development, clinical Development	Poster Session	Pavilion
9:45 - 10:30 am	Coffee Break	Break	Garden Courtyard & Pavilion
10:45 am - 12:15 pm	Visual Memory: Individual differences, imagery	Talk Session	Talk Room 1
10:45 am - 12:15 pm	Decision Making	Talk Session	Talk Room 2
12:15 - 1:15 pm	VSS Business Meeting	Business	Talk Room 2
1:00 - 2:45 pm	phiVis: Philosophy of Vision Science	Satellite	Banyan/Citrus
1:15 - 2:45 pm	Connect with Industry	Networking	Spotted Curlew
1:15 - 2:45 pm	Connect with Industry	Networking	Blue Heron
1:15 - 2:45 pm	Connect with Industry	Networking	Snowy Egret
1:15 - 2:45 pm	Synchronizing and Coordinating Multiple Data Acquisition Devices <i>Organized by SR Research</i>	Satellite	Horizons West
1:15 - 2:45 pm	Lunch (on your own) <i>Lunch Available for Purchase in Garden Courtyard 11:30 - 2:30</i>	Break	Garden Courtyard
2:45 - 4:30 pm	Theory	Talk Session	Talk Room 1
2:45 - 4:30 pm	Eye Movements: Mechanisms, naturalistic viewing	Talk Session	Talk Room 2
2:45 - 6:45 pm	Tuesday Afternoon Posters Visual Memory: Imagery Visual Memory: Mechanisms, models, individual differences Face and Body Perception: Neural mechanisms Face and Body Perception: Emotion Undergraduate Just-In-Time 3	Poster Session	Banyan Breezeway
2:45 - 6:45 pm	Tuesday Afternoon Posters Visual Working Memory: Spatial Visual Working Memory: Objects, features Object Recognition: Models Perceptual Organization: Neural mechanisms, models Attention: Capture 2 Attention: Temporal Attention: Reward Multisensory Processing: Audiovisual	Poster Session	Pavilion
4:30 - 5:15 pm	Afternoon Coffee & Snack	Break	Garden Courtyard & Pavilion
5:15 - 7:15 pm	Color, Light and Materials	Talk Session	Talk Room 1
5:15 - 7:15 pm	Motion: Neural mechanisms, models and illusions	Talk Session	Talk Room 2
8:30 - 10:00 pm	Visibility Social	Social	Banyan/Citrus
9:00 pm - 2:00 am	Chill Vision	Social	Palm/Sabal/Sawgrass

Symposium Sessions

SYMPOSIUM: FRIDAY, MAY 15, 2026, 8:00 – 10:00 AM, TALK ROOM 1

Intuitive physical reasoning in brains, minds, and machines

Organizers: RT Pramod¹, Nancy Kanwisher¹; ¹MIT

Presenters: Tomer Ullman, Shari Liu, Kelsey Allen, SP Arun, RT Pramod

How do we perceive, predict and plan actions in the physical world like pouring milk, catching a ball, or playing jenga? Inferring the physical state of the world to make predictions and generate explanations is called 'intuitive physics'. From infancy, humans expect objects to be solid, continuous, and subject to gravity; and they distinguish between things and stuff, recognizing that rigid objects maintain shape while substances deform or pour. They also understand relational concepts such as containment, support, and collision, revealing an early foundation for causal reasoning about the physical world. In adults, research has begun to reveal both the sophistication and the limits of intuitive physics, inspiring computational frameworks that model how humans reason flexibly about the physical world. Competing accounts debate whether intuitive physics relies on simple perceptual heuristics and pattern recognition, or on structured internal models that perform approximate simulations for prediction—analogue to the physics engines used in video games. In the latter view, intuitive physical reasoning requires a rich understanding of the physical scene, including the shapes, physical properties, relationships and dynamics of objects, and the ability to use this information to predict future states through approximate probabilistic simulations. These theoretical distinctions have motivated new empirical investigations into how such computations might be implemented in the brain. Over the past decade, neuroimaging studies have identified a network of fronto-parietal regions in the human brain, called the Physics Network, that encodes object properties, spatial relationships, and predicted future states. This network has been proposed as a neural substrate for simulation-based intuitive physical reasoning, yet key questions remain: What computations do they perform, and how do these computations depend on the participant's current goals? Are the regions not only involved in but necessary for intuitive physical reasoning? This symposium brings together researchers from cognitive science, neuroscience, and computational modeling to

synthesize emerging findings, highlight convergent themes across disciplines, and chart a path forward for understanding how brains, minds, and machines reason about the physical world. The organizers will kick off the symposium by introducing the topic and outlining the symposium (5 mins). Tomer Ullman will lay out how the human mind might be solving intuitive physics through various approximations under constraints of time, memory, and computation. Shari Liu will present behavioral evidence for interaction between intuitive physics and intuitive psychology and propose a computational framework that captures the interaction. Kelsey Allen will describe computational modeling studies that probe how people use intuitive physics to solve novel problems. SP Arun will share some exciting results on the representation of physical scene properties in the non-human primate brain using wireless recordings in freely moving monkeys. Finally, RT Pramod will review neuroimaging evidence for the 'physics engine' hypothesis in the human brain. Each speaker will present for 20 minutes with 3 additional minutes for Q&A. This symposium should be of interest to VSS members from various subfields including 3D perception, Object and Scene perception, Spatial vision, Material perception, Motion, Events, Relations, Perception and Action, Visual neuroscience, and Computational models of vision.

TALK 1

GOOD ENOUGH: APPROXIMATIONS IN MENTAL SIMULATION AND INTUITIVE PHYSICS

Tomer Ullman¹; ¹Harvard University

From spreading mayo on toast to dodging an errant frisbee, people handle the everyday physical world with remarkable ease. Without a sense of intuitive physics, every day would be a series of small disasters. How do people do it? One current model of intuitive physics supposes that people are carrying out a kind of mental simulation, moving objects in the mind step by step. While successful in several cases, even the people who champion this idea recognize that humans can't possibly be running a perfect simulation. But what sort of approximations might humans be using? Engineers concerned with 'good enough' simulations use principled short-cuts and workarounds to get around constraints of time, memory, and computation. I argue that these provide a good starting point for considering how the human mind, which is trying to create good-enough simulations under constraints of time, memory, and computation, may implement mental simulation. In this talk, I consider specifically approximate bodies in tracking, partial simulation, lazy evaluation in imagery, and bounds on

the number of objects that can be simulated at once. I will also discuss the computational models that capture these approximations, which include a proposed fundamental split between physics and graphics.

TALK 2

NAIVE PSYCHOLOGY DEPENDS ON NAIVE PHYSICS

Shari Liu¹; ¹Johns Hopkins University

Across the cognitive sciences, researchers have studied naive psychology (making sense of other people's actions in terms of their mental states) and naive physics (making sense of physical events in terms of their underlying mechanics and dynamics), as two separate processes. In this talk, I'll review evidence for these domain-specific systems for reasoning about the psychological and physical world, with distinct computational goals, representations, and neural substrates. At the same time, I propose that from early on in human development people navigate the social world by using two distinct but interacting systems for reasoning about other agents' immaterial minds and their material bodies. I'll review research from developmental psychology and cognitive neuroscience that provides evidence for this interaction: (1) human minds and brains represent the bodies of animate agents as objects, and their actions as physical events, and (2) we use physical knowledge to make inferences about other minds, including what other people want, feel and know, how hard they are trying, and how much danger they are in. I will also discuss Bayesian computational models of theory of mind, which articulate a formal hypothesis for this interaction. The talk will end by discussing key future empirical tests of this proposal.

TALK 3

COMPUTATIONAL MODELS OF HUMAN INTUITIVE PHYSICS

Kelsey Allen¹; ¹University of British Columbia

Intuitive physics lets us do more than predict how the world will unfold, it also allows us to solve problems. In this talk, I will describe my lab's work investigating the computational underpinnings of how people use intuitive physics to select and create novel objects to solve videogame puzzles. Across multiple timescales of learning, we find that people adapt the ways that they sample possible solutions, guided by their predictions of how those possibilities will affect the game environment. Finally, I will conclude by drawing comparisons with modern large scale machine learning methods applied to the same puzzles to illustrate the importance of structured reasoning in human behavior.

TALK 4

REPRESENTATION OF INTUITIVE PHYSICS PROPERTIES IN MONKEY INFEROTEMPORAL CORTEX

SP Arun¹; ¹Indian Institute of Science

We see and recognize objects every day but also form expectations about how they will handle when touched, or how they interact with other objects when moved. Where are these expectations created in the brain? One strong contender is the inferior temporal cortex, a region known for its highly object-selective neurons that are invariant to properties like size, position and viewpoint. While we know that IT neurons encode interesting visual features, whether they integrate physics expectations into their representations is unknown. Here, I will describe some studies in which we have performed wireless brain recordings from high-level sensory and motor regions in freely moving monkeys, to find that (1) After monkeys physically interact with objects, their IT neurons start encoding their mass; (2) When monkeys watch objects bouncing at various angles, their IT neurons show congruent tuning for incidence and bounce angle; and (3) When monkeys watch a ball disappear into an object or bounce off from it, their IT neurons dynamically infer that the object is hollow or solid. Taken together, these observations indicate that high-level visual cortex not only encodes visual properties, but more broadly encodes all object properties and even carries expectations about how objects will behave in the real world.

TALK 5

INTUITIVE PHYSICS IN THE HUMAN BRAIN

RT Pramod¹; ¹MIT

Intuitive physics has long fascinated researchers studying human behavior, yet its neural basis is still not well understood. In this talk, I will focus on a set of regions in the fronto-parietal cortex of the human brain, called the 'Physics Network (PN)', and review neuroimaging (fMRI) evidence for a variety of physical scene properties. Specifically, I will show that the PN encodes information about object properties such as mass and material, their configurations (including stability and contact relationships), and their predicted future states. In addition, I will present evidence that these regions respond not only to visual but also auditory and linguistic stimuli, indicating multi-modal processing. Our findings also reveal that the PN responds strongly to causal relationships (versus non-causal ones) specifically in the physical (versus social) domain, suggesting that these regions are representing the causal structure of the physical world. Finally, I will provide anatomical, functional, and network-level evidence that the PN is at least partially dissociable from the neighboring Multiple Demand (MD) system, which supports domain-general reasoning. Altogether, these results are indicative of the PN being the brain's 'Physics Engine', representing latent physical properties and causal structure of the physical scene, and enabling predictions through forward simulation.

Rhythms of vision: How neural oscillations structure perception and attention

Organizers: April Pilipenko¹, Jason Samaha¹; ¹University of California, Santa Cruz

Presenters: David Melcher, Joey Zhou, April Pilipenko, Luca Ronconi, Ian Fiebelkorn, Ayelet Landau

A growing body of work demonstrates that perception is not continuous but fluctuates rhythmically, shaped by dynamic interactions between oscillatory and aperiodic neural activity across the visual system. This symposium brings together six complementary research programs that collectively advance our understanding of how endogenous brain rhythms structure perception, attention, and decision-making. Together, these talks highlight how oscillations in the theta (3-8 Hz), alpha (8-12 Hz), and beta (13-30 Hz) frequency bands, along with aperiodic neural dynamics, govern temporal integration, sensory encoding, visuospatial processing, attentional selection, and the resolution of neural competition. The first talk examines how individual differences in alpha oscillations and aperiodic activity jointly determine the temporal binding windows that govern whether visual events are integrated or segmented. These results suggest a framework in which oscillatory sampling, neural noise regulation, and perceptual history together shape the temporal resolution of visual experience. The second presentation extends this perspective by revealing that alpha is not a unitary phenomenon: network-level MEG analyses show that multiple, functionally distinct alpha networks dynamically recur during perceptual decisions, each exerting different influences on cortical excitability and sensory evidence. Next, a series of EEG and psychophysics studies demonstrates how alpha power and phase shape early visual encoding. High alpha power shifts perceptual criterion, while optimal alpha phases maximize sensitivity by reducing trial-to-trial variability and sharpening feature tuning, providing a mechanistic account of how spontaneous alpha rhythms regulate sensory fidelity. Complementing these findings, new evidence highlights a central role for beta oscillations in dorsal-stream visuospatial processing, motion perception, and oculomotor control. Beyond basic mechanisms, beta-targeted neurostimulation in adults with dyslexia yields improvements in reading fluency and visuospatial working memory, pointing to translational applications of oscillatory research. The final two talks focus on the rhythmic nature of selective attention. One line of work shows that attention alternates at theta frequencies between states that enhance sensory sampling and states that promote attentional shifting, creating periodic fluctuations in perceptual sensitivity and susceptibility to distraction both in perception and in working memory. The concluding talk extends this

rhythmic perspective to attentional competition, demonstrating that when multiple objects vie for representation, attentional sampling slows from ~8 Hz to ~4 Hz, reflecting the division of processing resources across competing neural populations. By integrating insights across spatial scales, frequency bands, and methodological approaches, this symposium showcases how oscillatory dynamics provide a unifying framework for understanding visual processing as an intrinsically rhythmic and dynamically structured phenomenon.

TALK 1

THE ROLE OF ALPHA OSCILLATIONS AND APERIODIC NEURAL ACTIVITY IN THE TEMPORAL ORGANIZATION OF VISUAL PERCEPTION

David Melcher¹, Michele Deodato¹, Gianluca Marsicano¹; ¹NYU Abu Dhabi

Visual perception requires integrating sensory input over time, as well as segmenting the continuous flow into discrete events. Across a number of studies, we have investigated the role of oscillatory and aperiodic neural activity in this integration/segregation process. In particular, we have focused on individual differences and trial-by-trial variability in the temporal binding windows which define whether two stimuli are combined into a single, coherent percept or segmented into unique events. Here, we will present converging results from multiple paradigms, ranging from presentation of brief flashes to continuous streams of stimuli and judgments of causality from Michotte-style collision events. Using either resting-state EEG or event-related EEG during the psychophysical paradigm, we consistently find that alpha oscillatory activity and aperiodic activity combine to predict individual differences in perceptual judgments. Faster individual alpha rhythms were associated with narrower temporal integration windows, and a steeper aperiodic spectrum (lower neural noise) predicted greater perceptual precision (Deodato & Melcher, 2024). Moreover, individuals with slower alpha frequencies and flatter spectra showed stronger reliance on prior perceptual judgments. Subjective confidence was also linked to peak alpha frequency, with faster rhythms supporting higher certainty. Together, these findings suggest that temporal binding arises from the joint influence of the alpha rhythm and aperiodic neural activity. Faster alpha rhythms may promote greater perceptual resolution by generating more sensory samples per time unit, resulting in higher temporal precision, reduced reliance on prior perceptual experience, and greater subjective confidence in perceptual decisions.

TALK 2

DUAL ALPHA MECHANISMS IN VISUAL PERCEPTION

Joey Zhou¹; ¹Shenzhen University

Spontaneous neural oscillations, particularly in the alpha band (8-14 Hz), create rhythmic fluctuations in cortical excitability that influence conscious perception. However, the underlying neural mechanisms through which these oscillations modulate perception remain unclear.

Previous work has shown mixed results on how such excitability changes modulate sensory signals of interest versus noise, and what the consequences are for subsequent perception. A probable reason for these conflicting findings is that different alpha networks dominate across different studies. In this presentation, I will address this by introducing a network-based approach to analyze MEG data from visual decision-making tasks. Our results demonstrate that multiple, functionally distinct alpha networks coexist and dynamically recur over time, with each influencing perceptual decision-making through different mechanisms. These findings have important methodological and theoretical implications for characterizing the role of oscillatory activity in perception and decision making.

TALK 3

HOW SPONTANEOUS ALPHA OSCILLATIONS MODULATE SENSORY ENCODING

April Pilipenko¹, Jason Samaha¹; ¹University of California, Santa Cruz

Alpha-band (8-12 Hz) neural oscillations recorded from posterior brain regions are strongly associated with visual processing due to their ability to predict trial-to-trial variability in a range of perceptual reports from low-level detection to higher level illusions. Yet, our understanding of the mechanisms by which alpha shapes vision is limited. In this talk, I will present a series of recent EEG studies that integrate psychophysical modeling, signal detection theory, and reverse correlation to characterize the mechanisms by which alpha oscillations modulate early visual encoding. By presenting a range of stimuli that sample an individual's psychometric function linking stimulus contrast to detection, our first study demonstrates how spontaneous alpha power leads to a liberal shift in criterion during states of weak alpha power rather than an increase in perceptual gain, suggesting an indiscriminate boost in the baseline firing activity of sensory neurons. Our second study focused on the phase of alpha oscillations, revealing that perceptual sensitivity is optimized at specific phases. Double-pass and reverse correlation techniques reveal that optimal alpha phases are associated with a reduction in trial-to-trial variability and sharper perceptual tuning for stimulus-relevant features. Together, these findings provide converging evidence that alpha-band oscillations rhythmically shape visual encoding by dynamically modulating the excitability and synchronization of underlying neural populations. This work advances our understanding of how endogenous neural rhythms shape the fidelity of sensory representations and offers a mechanistic account of alpha's role in coordinating visual information processing.

TALK 4

NEURAL OSCILLATORY MECHANISMS SUPPORTING THE DORSAL VISUAL STREAM FUNCTIONALITY AND THEIR RESTORATION IN READING DISORDERS

Luca Ronconi¹; ¹University of Trento

Neural oscillations in the alpha (8–12 Hz) and beta (13–30 Hz) frequency bands play a fundamental role in coordinating feedback/re-entrant loops and large-scale cortical interactions. While alpha rhythms have been extensively associated with temporal and featural integration in visual perception, increasing evidence points to beta oscillations as crucial for spatial organization, motion processing, and visuo-spatial attention. Particularly, findings suggest that beta activity within poster parietal regions would sustain the functionality of the dorsal (“where”) visual pathway, providing fast preconscious spatial guidance over the slower ventral (“what”) stream. Crucially, this hierarchical dynamic might be crucial in guiding the extraction of letters/words identity and the oculomotor processes required for fluent reading. I will summarize recent evidence highlighting the role of beta oscillations as a neural code for dorsal pathway functions, integrating findings from human psychophysics, electrophysiology and neurostimulation. I will then present the results of a randomized clinical trial conducted in adults with developmental dyslexia (DD), a condition characterized by visuo-spatial and oculo-motor dysfunctions. The study employed bi-focal beta-band transcranial alternating current stimulation (tACS) over parietal areas, combined with a visuo-attentional training. As compared to a placebo/sham intervention, beta-tACS led to significant improvements in reading speed, oculomotor control, and visual motion perception, as well as long-term gains in visual/auditory working memory. Together, these findings provide converging evidence – from basic research to clinical application – supporting beta-band synchronization as a key mechanism for visuo-spatial and oculo-motor integration and a promising target for rehabilitating reading disorders.

TALK 5

UPDATING THE RHYTHMIC THEORY OF ATTENTION

Ian Fiebelkorn¹, Zach Redding¹, Paul Cavanah¹; ¹University of Rochester

Selective attention is a collection of filtering mechanisms through which behaviorally important information receives preferential processing. Recent research has shown that attention-related changes in neural activity (e.g., spiking activity) and behavioral performance (e.g., accuracy) fluctuate during attentional deployment, rather than being sustained over time. We and others have linked such temporal fluctuations in attention-related effects to the phase of frequency-specific neural activity. For example, different phases of theta-band activity (3–8 Hz) are associated with (1) either better or worse perceptual sensitivity at a to-be-attended location and (2) increased spiking among different, functionally defined cell types (i.e., either sensory or sensorimotor neurons). The Rhythmic Theory of Attention proposes that phasic changes in behavioral performance and neural activity reflect alternating attentional states associated with either attention-related sampling (i.e., sensory functions) or an increased likelihood of attention-related shifting (i.e., motor functions). A periodically occurring ‘shifting state’ could be beneficial, e.g., by preventing us from becoming overly focused on any single stimulus.

However, a periodically occurring 'shifting state' might also make us more susceptible to distracting information (i.e., behaviorally irrelevant information). Here, I will present recent work that demonstrates such rhythmic susceptibility to distractors. Increased susceptibility to distractors occurs during the same phase associated with decreased perceptual sensitivity (i.e., during the proposed 'shifting state'). I will also present recent work demonstrating that the theta-rhythmic process for selective sampling of environmental information is similarly engaged during selective sampling of internally stored information (e.g., information being held in working memory).

TALK 6

RHYTHMIC ATTENTION NEGOTIATES COMPETITION ALONG THE VISUAL HIERARCHY

Ayelet Landau¹; ¹Hebrew University of Jerusalem

Navigating the environment involves engaging with multiple objects, each activating specific neuronal populations. When objects appear together, the respective neuronal populations compete. Classical attention theories suggest that selection involves biasing one population over another. Recent research shows that perception fluctuates over time. When a single object is processed over time a ~8 Hz fluctuation seems to govern its perception. When attention is distributed over two different objects a 4 Hz fluctuation is measured, possibly due to the division of the 8 Hz rhythm between competing objects. In my talk I will explore these rhythmic phenomena, coined attentional sampling, across the visual hierarchy. I will argue that sampling is a selection mechanism that negotiates neuronal competition. It manifests as early as eye channels and extends to complex features higher in the visual hierarchy and potentially beyond the visual modality. Finally, I will discuss the cognitive significance of this mechanism and its potential neuronal implementation.

SYMPOSIUM: FRIDAY, MAY 15, 2026, 10:30 AM – 12:30 PM, TALK ROOM 1

What makes a task naturalistic? Towards a theory-inspired approach to studying visually-guided behaviour

Organizers: Jolande Fooker¹, Constantin A. Rothkopf¹; ¹Centre for Cognitive Science, Technical University Darmstadt, Germany
Presenters: Kendrick Kay, Jody Culham, Lukas T. Oesch, Caroline Robertson, Kathryn Bonnen, Constantin A. Rothkopf

As vision scientists, we aim to understand how behaving systems sense, process, and act upon visual information in service of their goals. Imagine making yourself a peanut butter jelly sandwich. This seemingly simple action requires the integration of perceptual, sensorimotor, and cognitive processes. Traditional laboratory paradigms have advanced the field by isolating separate process-

specific components and studying their impact on perception. Yet, a major challenge today is determining how these insights generalize to more naturalistic settings, where multiple processes operate simultaneously and where factors unknown to the experimenter can shape the visual computations that unfold moment to moment. An important step to bridging the gap between the lab and the wild, is to systematically measure natural behaviour in everyday tasks to understand the functional demands placed on the visual system. Moreover, empirical studies should be complemented by computational approaches that formalize the structure of natural visual tasks and identify the underlying neural processes that support them. The goal of our symposium is to encourage vision scientists to consider what it means for a task to be naturalistic. Specifically, we will pose three questions: Does the way a task is designed influence neural and behavioural responses? Can task goals be formalized by carefully studying diverse behaviour? Can we find theory-driven approaches to studying natural behaviour? The symposium will include six talks from speakers at different career stages. Speakers will present various experimental techniques, including neural imaging in humans, fluorescence microscopy in mice, virtual reality, motion capture, eyetracking, and computational modelling. To start things off, Jolande Fooker will give a brief introduction (3 minutes) on how we came to ask the question: "What makes a task naturalistic?". The introduction will be followed by six symposium talks (15 + 2 minutes questions). First, Kendrick Kay will highlight the importance of considering visual task design when investigating how the brain transforms visual inputs into representations. When studying neural responses in natural viewing, we not only need to consider task demands but also how natural actions, such as reaching and grasping, can be studied in the constrained space of the brain scanner. Second, Jody Culham will highlight how 3D displays and active video games can improve naturalism in human fMRI. Leaving the scanner, Lukas Oesch will demonstrate the neural control of decision monitoring in freely behaving mice. From perceptual decisions in mice, we will switch to cognitive processing in humans. Fourth, Caroline Robertson will show that immersive virtual reality with in-headset eye-tracking offers a methodological bridge between traditional vision science paradigms and natural behavior. We will then leave the virtual and dive into the real world. Fifth, Kathryn Bonnen will present how humans adapt visuomotor priorities and control strategies when walking in complex environments. Lastly, to tie things together, Constantin Rothkopf will describe how we can formalize the regularities observed in natural behaviour through theory-driven computational models. Each individual talk will conclude with a take-home message about key features of a naturalistic task. These take-home messages will set up the concluding panel discussion (15 minutes).

TALK 1

A BROAD SAMPLING OF HOW DIFFERENT TASKS INFLUENCE VISUALLY EVOKED RESPONSES IN THE HUMAN BRAIN

Kendrick Kay¹, Clay Curtis^{2,3}, Adrian Wong¹, Amy Poole¹, Eline Kupers⁴; ¹Center for Magnetic Resonance Research (CMRR), Department of Radiology, University of Minnesota, Minneapolis, Minnesota, USA, ²Department of Psychology, New York University, New York, NY, USA, ³Center for Neural Science, New York University, New York, NY, USA, ⁴Department of Psychology, York University, Toronto, Ontario, Canada

Visual neuroscientists seek to understand how the brain transforms visual inputs to representations that support behavior. In many visual neuroscience experiments, neural responses to visual stimuli are measured and the task performed by the observer is often neglected or assumed to have secondary importance to the scientific questions being tackled. However, since task demands influence visually evoked responses, it is clear that a complete characterization of visual processing must refer to both the stimulus and the task. Here, we describe ongoing efforts to collect a large-scale 7T fMRI dataset, termed the Visual Cognition Dataset (VCD), where extensive whole-brain fMRI responses (> 40 hours of data per participant) are measured from a small set of observers (n = 6) while they engage in a variety of visual tasks on a common set of stimuli. We sample both classic laboratory tasks (e.g., engaging in fixation, attentional, decision-making, and working memory tasks on controlled artificial stimuli such as Gabors and random-dot kinematograms) as well as more naturalistic tasks (e.g., making 'what', 'where', and 'how' judgments on complex natural scenes). Importantly, we impose shared constraints on the tasks with respect to timing and trial structure to facilitate interpretation and direct comparison of tasks. We anticipate that VCD will serve as a rich resource for gaining insights into task-dependent neural responses and, more generally, how the brain executes visual tasks.

TALK 2

BRINGING THE REAL WORLD INTO THE BRAIN SCANNER USING 3D DISPLAYS AND VIDEO GAMES

Jody Culham¹; ¹University of Western Ontario, London, ON, Canada

Vision science often uses simple stimuli like images along with simple stimulus-response tasks, particularly in studies using functional magnetic resonance imaging (fMRI), where more complex stimuli and behaviors are limited by methodological constraints. In this talk, I will review two new approaches in brain imaging for investigating more realistic stimuli and tasks and show how they can provide novel findings. First, I will discuss how the contributions of real-world geometry (physical/virtual size and distance) can be explored using tangible objects and binocular simulations with a 3D projector. Results suggest that real-world distance is represented throughout the visual system, especially for physical (vs. simulated) displays. Second, I will discuss how video games can be used to study brain activation in "closed-loop" scenarios where continuous visual feedback is used to guide actions. Active video game play evokes different brain regions and interactions compared to a control condition with matched visual

input and motor output. Together, new approaches show that a move toward naturalistic neuroimaging can reveal new aspects of visual processing that may have been neglected with conventional approaches.

TALK 3

ANTERIOR CINGULATE NEURONS COMBINE OUTCOME MONITORING OF PAST DECISIONS WITH ONGOING MOVEMENT SIGNALS

Lukas T. Oesch¹, Makenna C. Thomas¹, Davis Sandberg¹, João Couto¹, Anne K. Churchland; ¹Department of Neurobiology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles

In dynamic environments, animals must closely monitor the effects of their actions to inform switches in behavioral strategy. Anterior cingulate cortex (ACC) neurons track decision outcomes in these environments. Yet, it remains unclear whether ACC neurons similarly monitor behavioral history in static environments and, if so, whether these signals are distinct from movement representations. We recorded large-scale ACC activity in freely moving mice making visual evidence-accumulation decisions. Many ACC neurons exhibited nonlinear mixed selectivity for previous choices and outcomes (trial history) and were modulated by movements. Trial history could be stably decoded from population activity and accounted for a separable component of neural activity than posture and movements. Trial history encoding was conserved across different subjects and was unaffected by fluctuating behavioral biases. These findings demonstrate that trial history monitoring in ACC is implemented in a conserved population code that is independent of the volatility of subjects' task environment.

TALK 4

WHAT MAKES VISION NATURALISTIC? INSIGHTS FROM IMMERSIVE VR AND ACTIVE VISUAL PERCEPTION

Caroline Robertson¹; ¹Dartmouth College, NH, USA

Natural behavior unfolds as a continuous stream of self-directed perceptual decisions—what to look at, when, and why. A central challenge for naturalistic vision science is that these decisions are shaped not only by sensory input, but also by the ongoing process of constructing semantic understanding across the visual scene. Capturing the emerging process requires situating vision in active paradigms that engage perceptual-mnemonic interactions. In this talk, I will argue that immersive virtual reality with in-headset eye-tracking offers a way forward—a methodological bridge between traditional vision paradigms and natural behavior. This approach allows participants to explore rich environments with natural eye, head, and body movements while preserving full experimental control over stimulus content and timing. By enabling dense sampling of gaze across large, diverse sets of naturalistic stimuli within standardized

conditions, immersive VR provides an exceptionally powerful substrate for computational modeling of natural visual behavior. I will highlight evidence from recent work revealing three aspects of naturalistic visual processing that emerge in active viewing. In immersive environments, memory anticipates upcoming input across head turns, generating predictions that speed perceptual judgments. Active viewing also heightens the influence of semantic meaning on attention, producing richer, more differentiated patterns of meaning-guided sampling than in matched passive conditions. When participants freely navigate the motor and semantic landscapes of real-world scenes, individual differences become more pronounced—exposing stable conceptual priorities that generalize across environments and sessions. Together, these findings illustrate how immersive VR can help us understand how memory, meaning, and individual differences shape natural visual behavior.

TALK 5

LEVERAGING SPATIOTEMPORAL STRUCTURE IN NATURAL TASKS: EVIDENCE FROM VISUALLY-GUIDED WALKING

Kathryn Bonnen¹; ¹School of Optometry, Indiana University

In our everyday lives, perception and action unfold continuously in time, constrained by the dynamics of both the environment and the body. Walking is a critical naturalistic behavior that depends on precisely timed visual sampling to support foothold selection, navigate obstacles, and plan routes. These tasks rely on distinct spatiotemporal structures of gaze and movement, each reflecting different visuomotor priorities and control strategies, and walkers seamlessly transition between these tasks adapting their gaze and actions as environmental demands change. This complexity presents a challenge: how do we identify the meaningful units and patterns of behavior that reveal how vision supports movement in the real world? We address this question by measuring gaze, head, and body motion during walking in complex natural terrains. We use information about the task, environmental structure, and motor behavior to characterize the spatiotemporal organization of gaze. This approach leads to the identification of recurring visuomotor motifs such as step-by-step foothold-fixation sequences, forward-looking navigation checks, exploratory gaze forks, and a rhythmic “gaze cycle” in low-demand conditions. Each appears to serve a distinct functional role, and walkers sequence these visuomotor behaviors to meet the evolving demands of their environment as they move through it. Together, these results demonstrate how information about task features, environmental structure, and motor behaviour can be used to decompose visually-guided walking into meaningful visuomotor components, revealing how vision and movement are coordinated during walking.

TALK 6

COMPUTATIONAL ELEMENTS OF NATURAL VISION

Constantin A. Rothkopf¹; ¹Centre for Cognitive Science, Technical University Darmstadt, Germany

Natural, everyday tasks such as navigation, food preparation, and sports give rise to complex, sequential behavior of the whole body, including the eyes, head, and limbs. During such goal-directed behavior, vision has been described as tightly integrated with cognitive processes, including attention, memory, decision-making, planning, action selection and generation, and learning. Observing behavior in such scarcely controlled tasks appears to yield highly variable behavioral measurements with broad interindividual differences, rendering analysis with statistical methods developed for trial-based experiments challenging. How can such behavior be understood quantitatively? In this talk, I will present several naturalistic tasks together with a computational model of sequential decisions that incorporates sensory uncertainty, internal model uncertainty, and motor variability. These so-called Partially Observable Markov Decision Processes (POMDP) models can be understood as a generalization of signal detection theory to sequential tasks with additional sources of noise. The value of these models lies in their ability to provide parsimonious explanations of behavioral strategies, their associated errors and variability, and to quantify individual differences and explain their sources. For exemplary tasks including navigation, pouring a cup of coffee, and catching a fly ball, these models provide a subject-by-subject, trial-by-trial, and moment-by-moment explanation of how patterns of goal-directed sensorimotor behavior arise from the continuous and dynamic interactions of uncertainties in perception, cognition, and action, and how these uncertainties are actively shaped during natural behavior.

SYMPOSIUM: FRIDAY, MAY 15, 2026, 10:30

AM – 12:30 PM, TALK ROOM 2

Reimagining the binding problem(s) for the 21st century

Organizers: JohnMark Taylor¹, Seda Karakose-Akbiyik²; ¹Columbia University, ²Stanford University

Presenters: Peter Tse, JohnMark Taylor, Seda Karakose-Akbiyik, Ana Chica, Anne Sereno, Jake Quilty-Dunn

Research on the binding problem stands at an exciting juncture: decades of psychophysical research have illuminated the dynamics of feature binding, increasingly detailed neural measurements have mapped how features are represented across early, ventral, and dorsal visual regions, and deep neural networks (DNNs) have emerged as powerful “model organisms” for testing candidate models of visual processing, as well as valuable tools for real-world tasks requiring successful binding. However, this wealth of research spread across studies employing different paradigms or studying different parts of the brain—and often studied in separate research traditions—has given rise to the “methodological binding problem”: synthesizing

these disparate findings into unified neurocognitive models of how the brain integrates visual features. Indeed, just as the presence of multiple objects raises the challenge of associating them with the correct features, there may be multiple binding problems, and it is important to establish when different subfields are referring to the same one. This symposium tackles both the classical and methodological binding problems head-on by bringing into dialogue new developments from neurophysiology (spanning early, ventral, and dorsal visual regions), neuropsychology, psychophysics, computational modeling, and philosophy across six talks and a subsequent panel discussion. With the goal of renovating the binding problem for the 21st century, all six talks either present evidence that challenges us to reconceptualize core components of the binding problem as traditionally posed, or leverage new methods to make progress on important avenues inaccessible to earlier researchers. One talk presents psychophysical and neural evidence that visual search for specific feature conjunctions can approach pop-out with training, challenging the traditional view that only elementary features, not conjunctions, can be encoded pre-attentively. Another talk presents neuroimaging evidence against the neuroanatomical segregation of different features, such as color and shape, that has often been used to motivate the existence of a binding problem in the first place, and compares how the human brain and DNNs integrate visual features. New neuropsychological findings illuminate how the ventral and dorsal streams may interact during feature integration, a crucial but underexplored piece of the puzzle. Another line of work measures brain activity using fMRI and EEG during stimulus conditions designed to induce binding errors, identifying oscillatory signatures and occipito-parietal activation patterns that distinguish correct and incorrect binding (overcoming the longstanding methodological challenge of identifying neural signatures reflecting conscious binding rather than the mere co-occurrence of features). Another talk uses DNNs to identify key computational benefits of segregated feature processing (illuminating why the binding problem may arise to begin with), and provide powerful new tests of longstanding hypotheses about how space constrains feature binding. Finally, a talk from philosophy argues for a novel taxonomy of multiple forms of binding, developing much-needed conceptual clarity and setting the stage for a concluding panel discussion that seeks to relate different findings, establish theoretical common ground among research traditions, and articulate remaining gaps in our understanding. The symposium aims at a broad audience, reflecting its interdisciplinary nature and goal of “binding” insights across different subcommunities in VSS.

TALK 1

LEARNING TO MASTER OUR WORLD: CONJUNCTION LEARNING IN THE BRAIN AND MIND

Peter Tse¹, Sebastian Frank², Eric Reavis³, Mark Greenlee²;
¹Dartmouth University, ²University of Regensburg, ³UCLA

We do not come to the world recognizing the conjunction of redness and the octagonal shape that together comprise a stop sign. Yet, through exposure and with practice, we automatize this conjunction into a unit that speeds processing. Such conjunction learning happens both motorically and perceptually. Serial search for a conjunction-defined target typically involves slower reaction times in visual search the more distractors there are. We find that with practice, search slopes flatten such that conjunction search comes to resemble visual pop-out. This learning is not like traditional forms of perceptual learning in that it transfers to locations, sizes and orientations and even to other conjunctions that differ from those used in training. In this talk I will ask what actually is learned when a conjunction is learned, and how such learning happens in the brain, both cortically and subcortically.

TALK 2

JOINT POPULATION CODING OF COLOR AND FORM IN THE HUMAN VENTRAL PATHWAY AND CONVOLUTIONAL NEURAL NETWORKS

JohnMark Taylor¹, Yaoda Xu²; ¹Columbia University, ²Yale University

Both the human ventral visual pathway and CNNs trained to recognize objects have been shown to encode color and form information about stimuli, but how exactly are these features jointly represented in distributed population codes, and how might the nature of this representation vary between biological and artificial visual systems? This talk examines population coding of color and form across multiple experiments, using both controlled artificial stimuli and naturalistic stimuli. In the human ventral pathway, color and form information were found to be decodable using fMRI in every region in the ventral pathway for artificial stimuli, even in regions defined based on their univariate sensitivity to one feature or the other, with the exception of a color-sensitive region anterior to V4. However, despite this anatomical intermingling, color and form information tended to remain representationally independent, with brain regions encoding information about the two individual features, but generally not about particular conjunctions of features. By contrast, in CNNs, we found in a study using naturalistic colored stimuli that color and form tended to be encoded in an increasingly conjunctive format as processing proceeded. That said, further analyses suggest that the ventral pathway and CNNs trained to recognize objects are similar in their relative emphasis of these two features, with both systems increasingly emphasizing color relative to orientation information, but increasingly deemphasizing color relative to curvature information. Collectively, this work thus identifies both similarities and divergences in how these systems solve the problem of simultaneously encoding multiple features across distributed neural populations.

TALK 3

INSIGHTS ABOUT FEATURE BINDING FROM NEUROPSYCHOLOGICAL DEFICITS OF COLOR AND SHAPE PROCESSING

Seda Karakose-Akbiyik¹, Alfonso Caramazza²; ¹Stanford University, ²Harvard University

Color and shape are classically thought to be processed via partially segregated cognitive and neural mechanisms, giving rise to a binding problem for generating our unitary perceptual experience. Spatial attention is suggested to play a role in integrating visual features, but the mechanisms underlying this process are not fully understood. In addition, recent neurophysiological evidence on joint encoding of visual features in the ventral visual stream has called into question the existence of a binding problem. One way to get a handle on this debate is to study deficits of visual perception that arise after brain damage. In this talk, we will review deficits in perceiving visual features and their conjunctions that arise after brain damage, using color and shape processing as a case study. We will discuss the theoretical implications of these deficits for our accounts of visual feature encoding and binding. We will review work on visual agnosia, achromatopsia, and attentional failures as well as our recent findings from a patient with neglect (DH) who is able to report color, orientation, and location of shapes on the left half of objects despite lack of awareness of the shapes themselves. We propose that previous neuropsychological evidence on color and shape processing and DH's novel behavioral profile provide intriguing insights regarding visual feature encoding and integration, interactions between dorsal and ventral visual regions, and the mechanisms of spatial attention in this process.

TALK 4

BRAIN OSCILLATIONS AND OCCIPITO-PARIETAL CONTRIBUTIONS TO COLOR-SHAPE INTEGRATION

Ana Chica¹, Maria Cobos¹, Pablo Rodríguez-San Esteban¹; ¹University of Grenada

Our sensory system constantly receives information from both the environment and our own bodies. Despite our impression to the contrary, we are largely unaware of most of this information and often cannot report it accurately. Thus, perception—a seemingly ordinary process—can be complex and requires significant computational effort, which can lead to errors such as incorrect feature integration. We will present data from two studies that used functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) to investigate the neural mechanisms associated with correct and incorrect feature integration. fMRI data showed that a distributed set of brain regions exhibited increased activity during correct feature integration compared to incorrect integration, including enhanced functional coupling between occipital, parietal and frontal regions. In contrast, incorrect feature integration was linked to heightened occipital (V1-V2) responses in the early stages of processing, along with a general decrease in coactivation between occipital, parietal and frontal regions. EEG data indicated that alpha and beta power were enhanced during correct as compared to incorrect feature integration. This result was associated with cognitive processes such as pre-

stimulus preparation, distractor inhibition or the reliability of perceptual information representation within the cortex. Our findings underscore the role of parietal regions in feature integration and highlight the importance of occipital-parietal interactions, which oscillate in low-frequency bands, in perceptual processing.

TALK 5

INSIGHTS INTO THE REASONS FOR SEGREGATED PROCESSING AND THE PRIVILEGED ROLE OF SPACE IN BINDING, FROM A COMPUTATIONAL PERSPECTIVE

Anne Sereno¹, Zhixian Han²; ¹Purdue University, ²Indiana University

Much neuroscience research supports segregation of visual processing. Using artificial neural networks (ANNs), we previously showed that segregated streams for identity and space actively retain information about both identity and space, independently and differently; and further, that ANNs with two pathways, compared to a single pathway, have better performance (higher accuracy, less training time); with these advantages increasing with binding difficulty (i.e., increasing number of objects), suggesting segregation itself may help ease the binding problem. However, if segregated-pathway networks have advantages over single-pathway networks, why are not all attributes segregated in the brain? Simulations suggest that when there are a greater number of classes or greater discrimination is needed for a given attribute, the advantage of segregated pathways is greater. These ANNs used a relative location map to successfully bind identity and absolute location of objects. However, the brain needs to combine many different attributes (e.g., color, identity) and any of these attributes could be used in a similar fashion to constrain the binding problem. Using visual objects with multiple attributes (identity, luminance, orientation, location) to be recognized, we tried to find the best attribute (among an identity, luminance, orientation or location map) to constrain the binding problem. When attributes are independent, a location map is always the best choice for constraining the binding problem. A computational approach can provide insights into why segregation and the binding problem arise and the best way to constrain the binding problem. Implications of the findings for neuroscience and modeling will be discussed.

TALK 6

TWO WAYS TO SEE A RED SQUARE

Jake Quilty-Dunn¹; ¹Rutgers University

Vision scientists and philosophers share an interest in how representations are combined in visual perception. Much of the work on this topic presupposes that features are bound by being attributed to the same object. However, this “object-based” account of binding fails to explain cases of binding that occur outside attentive, object-centric processing. The philosophical arguments for the object-based account also fail to establish that attributing features to objects is necessary for binding. This talk explores a range of findings and argues that the visual system has multiple different principles for

binding features. Object-based binding is especially useful for interfacing with systems outside vision, including verbal report. Other forms of binding are more useful for processes internal to the visual system, including ensemble coding. The resulting picture is that vision constructs redundant representations of feature conjunctions for different purposes.

SYMPOSIUM: FRIDAY, MAY 15, 2026, 1:45 – 3:45 PM, TALK ROOM 1

Beyond prediction: Interpretable brain encoding models for understanding vision

Organizers: Zitong Lu¹, Apurva Ratan Murty²; ¹Massachusetts Institute of Technology, ²Georgia Institute of Technology

Presenters: Hossein Adeli, Katharina Dobs, Iris I.A. Groen, Zitong Lu, Andrew Luo, Apurva Ratan Murty

A central challenge in vision science is to understand how visual inputs are transformed into neural representations – and critically, how computational models can reveal the mechanisms underlying this transformation. Although neural encoding models have advanced rapidly with deep learning and large-scale datasets, purely performance-driven approaches offer limited scientific insight. To genuinely advance vision science, encoding models must be interpretable - constructed in ways that make their internal representations, architectural assumptions, and computational operations transparent and meaningful for understanding the brain. This symposium brings together early-career researchers using interpretable brain encoding models as mechanistic tools, highlighting both how such models are constructed and how they can be used – beyond traditional cognitive neuroscience methods – to reveal the mechanisms of visual processing. Across fMRI, EEG, and cross-subject datasets, the speakers will demonstrate how carefully designed architectures, structured feature spaces, representational analyses, and in-silico manipulations can uncover aspects of visual computation that conventional approaches cannot isolate. Together, the talks illustrate converging directions within this emerging landscape. Some approaches embed biologically meaningful structure, such as retinotopic maps or category-selective pathways, enabling models to recapitulate known functional organization while generating new hypotheses about information flow. Others interrogate internal representations through model–brain alignment to infer how visual categories and fine-grained feature dimensions may be encoded in the human visual system. Several studies leverage generative or perturbation-based encoding frameworks, which allow direct manipulation of internal representations to identify candidate computations underlying specific neural responses. Still others explore cross-subject generalization, using interpretable alignment strategies to reveal shared representational structure despite individual

variability. The six presentations span these perspectives in a coherent progression. Hossein Adeli will begin by introducing an interpretable attention-based routing model that links retinotopic features to category-selective regions. Also focusing on category-selectivity, Katharina Dobs will show how large-scale comparisons between ANNs and human cortex reveal the organizing principles of functional specialization. Extending beyond object vision, Iris Groen will demonstrate how encoding models can uncover scene affordance and dynamic video perception through model-brain alignment. Shifting from fMRI to EEG, Zitong Lu will present Img2EEG, a generative encoding framework enabling mechanistic exploration of temporal dynamics through simulated EEG signals. Then, Andrew Luo will illustrate how meta-learned cross-subject encoding and decoding models reveal generalizable representational structure across individuals. Finally, Apurva Ratan Murty will show the strengths and limits of current ANN-based encoding models by conducting cognitive experiments in them. All these talks present a unified vision: interpretable encoding models are becoming scientific instruments – tools that help reveal the computations that give rise to it. By integrating architectural constraints, representational transparency, and model-based experimentation, these approaches open new pathways for probing visual mechanisms in naturalistic settings. This symposium, featuring a brief introduction followed by six 16–20 minute talks with discussion, is designed to engage a broad audience across experimental and computational domains. By highlighting how interpretable brain encoding models can overcome limitations of traditional methods and provide mechanistic insight, the session aims to inspire new ways of using models to understand how the brain constructs visual experience.

TALK 1

DYNAMIC TRANSFORMER ROUTING FROM RETINOTOPIC CORTEX EXPLAINS ENCODING IN CATEGORY-SELECTIVE CORTICAL AREAS

Hossein Adeli¹, Nikolaus Kriegeskorte¹; ¹Columbia University

A major goal of neuroscience is to understand brain computations during naturalistic visual perception. A dominant approach is to use image-computable deep neural networks as a basis for linear encoding models. However, this approach ignores the structure of the feature maps both in the brain and the models, which could be used to build more interpretable models. In this work, we employ the transformer attention mechanism to model how retinotopic visual features can be dynamically routed to category-selective areas in high-level visual processing. We show that this computational motif not only is significantly more powerful than alternative methods in predicting brain activity during natural scene viewing, it is also inherently more interpretable because the attention routing signals can be easily visualized for any input image. Analyzing the attention signals in our model replicates the categorical selectivity of known regions (e.g. image patches with faces are routed to predict fMRI voxels in the Fusiform

Face Area, FFA) for face, body, word, and place regions and reveals structured between-area visual and semantic similarities that reflect their underlying computations. We also show how this state-of-the-art encoding model can serve as a digital twin, on which we perform in silico experiments to discover categorical selectivity across the whole brain. Our results suggest that the human visual hierarchy employs a dynamic routing mechanism that guides relevant information from image-specific retinal locations to category-selective regions. Our work shows how interpretable encoding models can help us understand visual processing in the brain.

TALK 2

INTERPRETING VISION THROUGH THE LENS OF CATEGORY SELECTIVITY IN BRAINS AND ANNS

Katharina Dobs¹; ¹Justus Liebig University Giessen

Functional specialization is a hallmark of visual processing. In the human visual cortex, distinct regions respond selectively to specific categories, such as faces, bodies, or scenes. Strikingly, artificial neural networks (ANNs) for vision show emergent patterns of category selectivity reminiscent of those observed in the brain. While this functional alignment offers a unique opportunity to probe the origins of category selectivity, many questions remain open: To what extent do ANNs and the human visual cortex rely on shared organizing principles, and can ANNs help reveal functional specializations that have not yet been discovered in the brain? Here, we systematically compare category selectivity in ANNs and human visual cortex across a large and diverse set of categories. We find that the degree of selectivity for a given category in ANNs predicts its selectivity in the visual cortex, particularly for ANN units and cortical regions showing clear functional tuning. By quantifying the correspondence between category selectivity in artificial and biological vision systems, these results offer an interpretable framework to investigate the computational principles underlying functional specialization and to uncover previously unknown selectivities in the visual cortex.

TALK 3

LEVERAGING DNN-BRAIN ALIGNMENT TO UNDERSTAND HOW HUMANS PERCEIVE SCENES AND VIDEOS

Iris I.A. Groen¹; ¹University of Amsterdam

The unprecedented ability of deep neural networks (DNNs) to predict neural responses in the human visual cortex has created excitement in vision research about these models' potential to capture human visual perception. However, many initial observations relating the inner representations of DNNs to those in the human brain primarily confirmed existing ideas about visual cortex (e.g., the presence of a low-to-high complexity object feature hierarchy). Can we leverage brain-DNN alignment to learn something new about how human vision works? In this talk, I will show how we use DNN-based representational alignment and model comparison to move from object vision towards new domains, including the neural representation of

scene affordances and spatial and temporal neural dynamics during video perception. Moreover, I will showcase how DNN-brain alignment can be leveraged to generate new hypotheses about neural representations in human visual cortex via brain-guided image diffusion, as implemented in our newly developed Brain Activation Control Through Image Variation (BrainACTIV) method.

TALK 4

EXPLORING HUMAN VISION THROUGH Img2EEG: AN ENCODING FRAMEWORK GENERATING HIGH-RESOLUTION TEMPORAL EEG SIGNALS FROM VISUAL INPUTS

Zitong Lu^{1,2}; ¹The Ohio State University, ²Massachusetts Institute of Technology

Understanding the complex interplay between visual stimuli and brain activity has been a focal point in cognitive neuroscience. The recent advent of image-to-brain encoding models presents an unparalleled opportunity for deeper explorations and broader applications in experimental and computational research. Building on this momentum, we developed "Img2EEG", a pioneering encoding framework consisting of visual, semantic and integration modules based on state-of-the-art deep learning models from computer vision and natural language processing. Trained on a large-scale EEG dataset of natural images at the individual subject level, Img2EEG effectively generates highly realistic EEG signals given any image input. We propose Img2EEG as an innovative and internally manipulable tool for investigating visual mechanisms. Through internal and external ablation experiments on Img2EEG, we unfold the detailed temporal dynamics underlying visual processes. Moreover, through in silico experiments on Img2EEG, by feeding it novel image sets outside its training data, we not only reproduced the classic N170 component in the model-generated EEG signals but also leveraged its internal manipulability to identify the key features driving this response. Furthermore, our Img2EEG achieves more accurate decoding, outperforming current state-of-the-art EEG decoding models (e.g., 200-class classification accuracy on THINGS EEG2). Finally, we applied Img2EEGs across all ImageNet images, creating a large stimulated EEG dataset, 'ImageNet-SimEEG', providing potential applications for decoding images from EEG and advancing more brain-like visual models. Overall, Img2EEG mapping from visual inputs to high temporal resolution brain signals offers novel and powerful approaches to probe human visual representations.

TALK 5

META-LEARNING IN CONTEXT ENABLES TRAINING-FREE CROSS-SUBJECT ENCODERS AND DECODERS

Andrew Luo¹; ¹University of Hong Kong

Understanding the relationship between sensory stimuli and neural responses is a fundamental goal in neuroscience. A significant

obstacle has been developing models that generalize across individuals due to high inter-subject variability in neural representations, traditionally necessitating bespoke, subject-specific models. To address this, we propose a unified framework that leverages in-context learning for both predicting neural responses (encoding) and interpreting them (decoding), without requiring subject-specific fine-tuning. Our approach utilizes a transformer-based architecture, meta-optimized to rapidly infer an individual's unique neural encoding patterns by conditioning on a small set of stimulus-response examples. For encoding, this allows the model to accurately predict voxelwise neural responses to entirely novel stimuli, outperforming existing methods in low-data regimes. For decoding, the framework performs hierarchical functional inversion to achieve robust semantic interpretation of brain activity. The model demonstrates strong generalization across different subjects and acquisition parameters, a critical advance that eliminates the need for anatomical alignment or stimulus overlap. This work marks a significant step away from single-subject analysis towards a generalizable foundation model for non-invasive brain signal analysis. By enhancing interpretability and predictive power, our framework advances the broader effort to predict and understand neural representations in the human brain.

TALK 6

USING COGNITIVE TESTS TO REVEAL THE STRENGTHS AND LIMITS OF BRAIN ENCODING MODELS

Apurva Ratan Murty¹; ¹Georgia Institute of Technology

If artificial neural networks (ANNs) are to serve as mechanistic hypotheses of the visual system, they must also recapitulate decades of cognitive experiments. Neural predictivity, the current modeling standard, is useful but does not tell us the computations a model implements. Cognitive neuroscience studies, by contrast, were designed precisely as probes of perceptual computations, yet we almost never use them to evaluate brain models. In this work, we revisited 20 experiments across face, scene, and body perception that localized and characterized responses in human category-selective cortex. Many of these studies used manipulated stimuli that fall outside the training domain of standard ANNs. We ran these experiments directly through models ($N = 150$) without retraining them on any of the specific images or fMRI data from previous studies. As expected, ANNs recapitulated many hallmark univariate and multivariate signatures. However, the cognitive tests also exposed clear differences between brain regions, between specific cognitive phenomena, and separated the 150 models better than prediction scores alone. Together, these findings show how cognitive neuroscience experiments can clarify what current ANN models explain and reveal their limits as in-silico models of the brain. All models are released for researchers to run their own cognitive tests.

SYMPOSIUM: FRIDAY, MAY 15, 2026, 1:45 – 3:45 PM, TALK ROOM 2

The quest for 'the average person' in vision science: Promise and pitfalls

Organizers: Jeremy Wilmer¹, Michael Herzog²; ¹Wellesley College,

²Swiss Institute of Technology (EPFL)

Presenters: Michael Herzog, Jeremy Wilmer, Jonathan Winawer, Anna Kosovicheva, Michael Webster, Alasdair Clarke

In everyday language, average means typical, usual, and ordinary. In data analysis, the average is a mathematically defined summary statistic. A good summary statistic should match the ordinary meaning of average, capturing the main tendencies while “dispensing with needless details” (Oxford Languages). But does the scientific practice of describing the human mind and brain through the average reasonably capture what is important? Human variation is ubiquitous. It demands explanation, gives life color, and inspires scientific curiosity. Moreover, as soon as one considers multiple dimensions—even highly correlated ones—the “average person” quickly ceases to exist (Rose, 2015; Downey, 2024). In this sense, the average person may, in fact, represent no one at all. At the same time, data descriptions that retain all details can be uninterpretable; they may lead one to miss the forest for the trees. In this symposium, we critically examine the scientific reliance on averaging across individuals in vision research. When does this approach succeed? When does it fall short? And what complementary methods might it require? The talks in this symposium tackle these questions via diverse methodologies (neural, behavioral, computational) and across multiple content areas (faces, illusions, color, scenes, graphs, eye movements). Herzog and colleagues show that the classic N1 component of visual evoked potentials is an artifact of averaging, and that there is no common factor in visual illusions; in both cases, the average response is atypical and deviations from it constitute stable individual traits. Wilmer and colleagues show that the average viewer’s reading of plotted averages masks massive variability and misperception, concluding that such graphs are a suboptimal method of visual data communication. Winawer and Benson show that standard cortical atlases, derived from population averages, often have large errors when applied to individual brains, and explore parcellation methods that balance efficiency with individual-level accuracy. Kosovicheva and colleagues show that even processes as simple as localization vary considerably between individuals and are tied to other forms of spatial judgment, suggesting that results relying on the average observer do not reflect individuals’ perceptual reality. Webster and colleagues review the individual differences manifest at different levels of color vision, the neural coding of color, and how they can be compensated to increase a common color experience among

observers. Finally, Clarke and colleagues considers what we can infer from correlational analyses on averaged data and proposes new modeling solutions for studying individual differences. Taken together, these talks reveal a wide array of rich insights embedded in non-average-ness; that is, in deviations from the average. They show, for example, that patterns of covariance across individuals can reveal mechanisms; that proper stratification before averaging can reap benefits of averaging, such as increased signal to noise ratio, without compromising naturalism; and that the existence and magnitude of stable variation can be a phenomenon worthy of study in its own right. We hope that this symposium will stimulate ongoing discussion about the benefits and risks of averaging data across individuals, and about the complementarity between traditional average-based approaches and approaches that embrace individual differences.

TALK 1

INTER-INDIVIDUAL VARIABILITY IS (OFTEN) NOT NOISE

Michael Herzog¹, Melissa Faggella¹, Simona Garrobo¹; ¹Swiss Institute of Technology (EPFL)

Classically, vision scientists treat inter-individual differences as a nuisance and eliminate it by aggregating individuals (grand-group average). Behind this procedure is the assumption that there is one “true” waveform, hidden in noise: vision research is ergodic, i.e., conclusions based on the mean are also valid for the single participants and vice versa. Here, we tested this assumption. We analysed visual evoked potentials (VEPs) recorded during a backward masking task and found that different observers had different numbers of peaks in their waveforms. When averaging, we obtained a single peak, the classic N1 component. We tested the same participants 5 and 10 years later and obtained virtually the same waveforms, i.e., a participant with 3 peaks had 3 peaks also after a decade. Hence, differences in waveforms are traits, not noise, and the N1 is an artifact of averaging. In fact, many participants did not show an N1. Similar results hold true for behavioural tests. Participants performed a battery of illusions. We found almost no correlations between the illusions. An observer may experience a strong Ebbinghaus and a weak Müller-Lyer illusion. Hence, there is no common factor for illusions. Still, in each participant, the respective illusion magnitudes were stable across a year. We will discuss the implications of heterogeneity for vision research, including how to measure such heterogeneity properly, and we will show how inter-individual differences may pave new avenues in vision research.

TALK 2

GRAPHED AVERAGES SOW DISAGREEMENT: THE AVERAGE READING OF AN AVERAGE IS ATYPICAL

Jeremy Wilmer¹, Sarah Kerns¹, Ken Nakayama²; ¹Wellesley College, ²University of California, Berkeley

A key aim when graphing data is that, as far as possible, different viewers should gain a comparable conception of the data from the

same graph. The average (mean) has long been the most common computed and plotted value in vision research. Yet little is known about how plotted averages are perceived and understood, especially by viewers who may vary in statistical training. Here, we first developed an accessible yet information-rich drawing-based measure of how viewers read graphs of averages. We then gathered readings from a broad array of participants. Surprisingly, roughly 1 in 5 viewers, across a wide range of education levels, misinterpreted the bar in a bar chart of averages as depicting the range of the data, with the average frequently placed near the middle of the bar. The average read of the average, therefore, was inside the bar, a location with few to no real responses. A second surprising finding was that viewers both wildly underestimated and disagreed upon the amount of variation around individual averages. For example, for a graph taken directly from the most popular Introductory Psychology textbook, a gender difference was falsely perceived, by the average viewer, to have zero overlap, as if all men outperformed all women. Moreover, across many graphs and levels of statistical training, viewers' conceptions of variation and overlap differed so much that, again, the average response reflected virtually no actual responses. We conclude that average-only graphs routinely fail at their basic tasks of consistent, and correct, data communication.

TALK 3

COMMON PATTERNS AND INDIVIDUAL DIFFERENCES IN VISUAL CORTICAL MAPS

Jonathan Winawer¹, Noah Benson²; ¹New York University, ²University of Washington

About a quarter of the human cerebral cortex is visual, consisting of multiple retinotopic maps and other areas that are highly responsive to certain classes of stimuli such as faces. Identifying these areas in individual participants has many applications in neuroscience and medicine: interpreting visual disorders; tracking cortical changes over development and aging; localizing intracranial electrodes in patients; characterizing how different areas respond to stimuli or tasks; tracing EEG or MEG signals back to specific cortical maps. Functional MRI experiments can identify most of these areas in any healthy individual, but this can be time-consuming and costly. An alternative way to identify visual areas is by standardized atlases, which take advantage of the tendency for specific visual areas to align with particular sulci and gyri. The atlases rely on two levels of averaging: one for the cortical surface itself, and another for the location of visual maps on this surface. Such atlases are now used in many research studies. However, atlases without functional localization at the individual level raise a number of questions: How often do they mislabel cortical areas? When they err, how large is the error? In short, how much do we lose by treating each individual as if they are the average? We argue that for some applications, cortical atlases without individual functional localizer data are insufficiently accurate. We then consider new methods that have some of the efficiency advantages of atlases while still respecting individual differences.

TALK 4

INDIVIDUAL DIFFERENCES IN LOCALIZATION REVEAL LINKS BETWEEN FUNDAMENTAL VISUAL PROCESSES

Anna Kosovicheva¹; ¹University of Toronto Mississauga

Approaches to examining individual differences in visual perception often emphasize high-level processes like object recognition, but what variability can we observe in fundamental visual tasks, and are these between-observer differences simply random noise? In a striking example of variability, several studies demonstrate that individuals exhibit consistent, idiosyncratic patterns of directional error in position judgments. In a series of laboratory studies, participants report the positions of briefly flashed targets in the periphery, and the directional errors in these tasks have been shown to be highly stable over several months and consistent across measurement methods (Kosovicheva & Whitney, 2017). These responses are weakly correlated between observers, such that the “average observer” shows accurate responses. Importantly, these directional biases have been linked to individual differences in other fundamental processes, including perceived size and acuity (Wang, Murai, & Whitney, 2020), as well as the strength of visual crowding (Haseeb, Wolfe, & Kosovicheva, 2023). More recently, we have shown that these idiosyncrasies exist not only in laboratory samples but in much larger, less-controlled, tasks (with over 9,400 observers and 4.5 million trials), demonstrating that stable individual variation is a meaningful, underacknowledged element in understanding localization. In this online sample, we found that directional localization errors were, on average, weakly correlated between all possible pairs of participants, yet were highly consistent and stable within each observer. Focusing on the average observer has revealed much about vision, but understanding visual function requires tackling the complexity and variability of individuals to reveal important links between different processes.

TALK 5

INTERPRETING AND ACCOUNTING FOR INDIVIDUAL DIFFERENCES IN COLOR VISION

Michael Webster¹, Kara Emery², Camilla Simoncelli¹; ¹University of Nevada, Reno, ²New York University

The concept of a “standard” or average observer is central to color science and application, but masks the marked variability even among observers with “typical” trichromatic color vision. These differences arise independently at many levels and affect diverse aspects of color coding, from genetics and sensitivity to color perception and cognition. As a result, two observers viewing the same stimulus may have very different experiences. Measurements of these variations have provided important and often surprising clues about the mechanisms and representational structure of color vision. Differences can be amplified in modern wide-gamut devices, and there is increasing recognition and interest in accounting for them to better display and communicate information about color. One approach is to introduce

individual differences into the stimulus by tailoring it to the individual, so that two observers, viewing their own stimuli, have more standard color experiences.

TALK 6

TALK CANCELLED - GENERATIVE MODELS CAN EXPLAIN (OR AT LEAST HELP US UNDERSTAND) INDIVIDUAL DIFFERENCES IN COGNITION

Alasdair Clarke¹, Anna Hughes¹, Amelia Hunt²; ¹University of Essex, ²University of Aberdeen

Nearly all experiments in cognition and visual perception involve taking multiple measurements from each participant. This includes taking multiple recordings of the same concept (e.g., recording response time on a number of trials), and taking multiple measurements per trial (e.g., response time and accuracy). The number of measurements expands as we add in eye-tracking metrics, EEG statistics, and so on. Researchers interested in individual differences typically assume that these measurements are i) mutually independent and ii) free from confounds, before taking simple summary statistics and then applying correlation procedures. A general theme of research on individual differences in our field has been a lack of correlations. I argue that, on reflection, this shouldn't be surprising, as the measurements we put into our correlations are imprecise composites of multiple causal factors. For example, it is well established that reaction times are influenced by learning and inter-trial serial dependencies. This means that participants in a hypothetical visual search experiment may all have identical “skill” in search, but could exhibit wildly different RTs due to differences in learning rates and serial dependency effects. I present simulated data to demonstrate how such processes can both mask correlations and produce spurious correlations. A potential solution to my hypothetical problem is the use of generative models with latent parameters. I provide an example in the form of FoMo, which decomposes performance in visual foraging into a small number of per-participant parameters.

Talk Sessions

**TALK SESSION: FRIDAY, MAY 15, 2026, 4:15
– 5:45 PM, TALK ROOM 1**

Attention 1

Moderator: Alex White, Barnard College

TALK 1, 4:15 PM, 15.11

VISUAL SEARCH IS DISRUPTED BY PERCEIVED LIGHTING CHANGES IN DYNAMIC AND STATIC DISPLAYS

*Gabriel Conn¹, Zephyr Markley¹, Katie Jobson², Kayla Sansevere³,
Katherine Moore¹; ¹Arcadia University, ²University of Pennsylvania,
³Tufts University*

Visual search depends on internal search templates that guide attention toward likely targets while filtering distractors. These templates, stored in activated long-term memory, can shift flexibly depending on task demands and display context. One factor influencing template formation is color constancy—the perceptual adjustment that maintains stable color appearance across changes in lighting. We investigated whether perceived lighting changes automatically reshape visual search templates, even when such adjustments hinder performance. Participants searched a rapid serial visual presentation for a target presented while background and distractor colors were systematically shifted in hue to simulate lighting changes. There were two critical distractors for each target, which were generated by applying red and green filters to the target. Participants were instructed to reject these distractors. Across four experiments and six unique targets using a dynamic display, participants consistently showed greater difficulty rejecting distractors whose color shifts matched the background tint, i.e., distractors that appeared as a target would have had there been a true lighting change. Target identification also suffered on tinted backgrounds relative to white. These effects were amplified when the background color persisted across trials as opposed to when it changed randomly, suggesting that establishing a lighting context strengthens template shifts. We found similar effects in static displays of realistic visual scenes. These findings demonstrate that color constancy principles influence visual search automatically, constraining strategic control over search templates. This reflexive adjustment, while sometimes counterproductive as in the case of this study, is reflective of an adaptive mechanism for object recognition in natural environments.

TALK 2, 4:30 PM, 15.12

SHORT-TERM EFFECT OF EYE-SPECIFIC ATTENTION TRAINING IN AUGMENTED REALITY

Yizhi Wang^{1, 2} (18369329216@163.com), Jinyou Zou³, Peng Zhang^{1, 2}; ¹State Key Laboratory of Cognitive Science and Mental Health, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China, ²University of Chinese Academy of Sciences, Beijing, China, ³Aier Academy of Ophthalmology, Central South University, Changsha, China

Introduction: Top-down attention can modulate visual processing in specific monocular channels. Here, we investigated whether short-term (30 min) training in altered reality (AR) can influence eye-specific attention ability. Methods: During AR intervention, images were captured in real time through a HD high-speed camera built in a pair of LED goggles. Participants played action video game with the non-deprived eye (NDE) presented with original videos while the deprived eye (DE) received phase-scrambled pink noise. A modified binocular rivalry paradigm was used to assess eye-specific attention effect before and after training. In the central visual field, one eye was presented with a radial grating while the other eye with dynamic Mondrian noise. Two monocular attentional cues, one in each eye, were presented to the left and the right side of the rivalry stimuli. Participants were required to pay attention to the monocular cue in one eye while monitoring the perceptual state of the rivalry stimuli. They pressed a button when the radial grating break interocular suppression from the dynamic noise. Results: At baseline, suppression time was significantly longer when the attended eye was presented with the dynamic noise than with the radial grating, suggesting an eye-specific attention effect. After training, eye-specific attention effect disappeared for the NDE, while no significant change was observed for the DE. Conclusion: Depriving phase information in one eye for 30 min in altered reality significantly reduced the eye-specific attention ability of the non-deprived eye. This effect suggests an adaptation of eye-based attention to a monocular channel.

TALK 3, 4:45 PM, 15.13

ATTENTIONAL EYE SELECTION MODULATES INTEROCULAR SUPPRESSION ACROSS THE VISUAL CORTICAL HIERARCHY

Chuan Hou¹, Junxian Rao¹; ¹Smith-Kettlewell Eye Research Institute

Previous work revealed that attentional modulation in V1 from the amblyopic eye is degraded, and that this degradation correlates with the magnitude of behaviorally measured visual suppression in amblyopia, suggesting a tight link between selective visual attention and interocular suppression (Hou et al., 2016). In this study, we investigated whether attentional eye selection can causally modulate interocular suppression in the visual cortex. We employed dichoptic multiple-object tracking stimuli, modulated at distinct temporal frequencies in each eye, in observers with normal vision. The different

frequencies allowed us to quantify spectral response components associated with each eye's inputs: 3.75 Hz (F1) for target dots in the tracking eye and 3 Hz (F2) for distractor dots in the other eye. Neural activity and behavioral performance were recorded using a 128-channel EEG system under three conditions: (1) tracking dots in the dominant eye, (2) tracking dots in the non-dominant eye, and (3) passive viewing with no tracking. Regions of interest (ROIs) were defined with structural and functional MRI, including V1, extrastriate cortex (V3A, hV4, hMT+, and lateral occipital cortex), and the intraparietal sulcus (IPS). Our results showed that during natural viewing (passive condition), there was little or no neural bias between the two eyes. However, selectively attending to one eye (i.e., tracking target dots with either the dominant or non-dominant eye) produced stronger responses to the attended eye's signal compared to the distractor eye. This attentional response bias was most prominent in extrastriate cortex and posterior IPS. These findings demonstrate that attentional eye selection can modulate interocular suppression across the visual cortical hierarchy. This mechanism may have clinical implications for understanding and treating visual disorders characterized by interocular sensory imbalance, such as the pathological suppression observed in amblyopia.

This research was supported by NIH grant R01EY035346 to C. H.

TALK 4, 5:00 PM, 15.14

PRE-SACCADIC ATTENTION AND THE RIGHT VISUAL FIELD ADVANTAGE FOR WORD RECOGNITION

Mariam Latif¹, Devon Lack¹, Alex L White¹; ¹Barnard College, Columbia University

Background: Written words are generally easier to recognize in the right visual field (RVF) than in the left visual field (LVF). This asymmetry is typically measured while participants maintain fixation and judge words at unpredictable locations, unlike how readers frequently shift attention and move their eyes. Here we manipulate spatial attention in order to elucidate its role in the hemifield asymmetry. The question is whether English words are still easier to recognize in the RVF even when attention is focused to the left during preparation of a leftwards saccade. Methods: On each trial, we presented an English word on one side of fixation and a pseudoword on the other. The primary task was semantic categorization of the word. On neutral trials, participants maintained fixation and attended to both sides equally. On saccade trials, a central cue prompted participants to immediately saccade to the left or right. The word appeared just before the saccade, either at the cued side (2/3 of trials) or the opposite side (1/3 of trials). Results: On neutral trials, semantic categorization accuracy was higher when the word was in the RVF than the LVF. On saccade trials, as time approached the saccade onset, accuracy increased for words at the cued side and decreased at the opposite side. Thus, pre-saccadic attention modulated the magnitude of the hemifield asymmetry. However, the asymmetry never reversed: even just before a leftwards saccade, accuracy was

still better for words in the RVF. Conclusion: The hemifield asymmetry is not due to an attentional bias to the right, as it persists when attention is focused to the left. Rather, we propose that it is due to the left lateralization of reading-related brain areas. Future directions include testing readers of right-to-left scripts (e.g., Arabic) who have more experience attending leftwards while reading.

TALK 5, 5:15 PM, 15.15

DUAL TASKS IN ANIMATED DATA VISUALIZATION: EVALUATING GLOBAL STATISTICS UNDER MULTIFOCAL ATTENTION

Ouxun Jiang¹, Steven Franconeri¹; ¹Northwestern University

When presented with an animated data visualization, viewers often complete multiple tasks simultaneously. For example, in an animated bubble chart, viewers might track a few bubbles representing data points of interest while also evaluating whether the average size of the bubbles is increasing or decreasing. Under such dual-task conditions, how do the visual tasks interfere with each other? If the tracking task impairs the evaluation of average features, it would suggest that the 'multifocal' attention used during tracking is insufficient for extracting global statistics; otherwise, it would suggest that global features remain accessible outside focal attention. To test this, we asked participants to perform a multiple-object tracking task while evaluating the average feature (size) changes of either the tracked targets or the non-targets. Results showed that during tracking, participants could reliably evaluate the average size change for targets but not for non-targets. It suggests that the multifocal attention demanded by tracking tasks relies on more of a focal selection, consistent with the idea that feature information degrades outside visual focus, even for global statistics. We also tested whether a practical design intervention could ease processing demands for animated data visualizations by highlighting target objects. Highlighting improved performance in tracking and in evaluating average size change of non-targets. These findings provide insights for designers to consider the limits of demanding multiple tasks within one animated data visualization, and practical design interventions that could mitigate such limits.

TALK 6, 5:30 PM, 15.16

A HEADWINDS BIAS IN VISUAL ATTENTION: SELECTIVE DETECTION OF SELF-DISADVANTAGING CHANGES IN A GAME OF PONG

Loren Matelsky¹, Colleen Macklin², Benjamin van Buren; ¹The New School, ²Parsons School of Design

People are better able to recall circumstances that have disadvantaged them (headwinds) compared to those that have been advantageous (tailwinds). For example, people believe that their parents' actions more frequently favored their sibling, and political party members believe that the political system is biased against their own party. Does the tendency to overweight headwinds events first

emerge during recall, or might it originate earlier, in visual attention and perception? If so, when playing a game of Pong against a computer-controlled paddle, people might be more likely to notice changes that impede their goals, relative to changes that help them. In a preregistered experiment, 160 online players completed a 60-second game of Pong with an embedded measure of visual change detection/blindness. Every five seconds, the screen flickered gray for 150ms, masking a 5% length change in one paddle. In Headwinds games, the changes hindered the player (either the player's paddle shortened, or the computer paddle lengthened); in Tailwinds games, the changes helped the player (either the player's paddle lengthened, or the computer paddle shortened). After the game, we measured players' awareness of the visual changes and their impressions of whether the game was fair. Although paddle length changes were perfectly matched across conditions, players were substantially more likely to notice Headwinds changes vs. Tailwinds changes (65% vs. 35%) and to judge Headwinds vs. Tailwinds games as unfair (74% vs. 46%). Players selectively noticed changes that hindered them, and were blind to changes that helped them. We propose that the headwinds/tailwinds evaluative bias is rooted in a visual information processing bias, shaping not just what we can recall, but what we even see in the first place.

TALK SESSION: FRIDAY, MAY 15, 2026, 4:15 – 5:45 PM, TALK ROOM 2

Perceptual Organization: Neural mechanisms

Moderator: Daniel Kaiser, Justus Liebig University Giessen

TALK 1, 4:15 PM, 15.21

CHARACTERIZING FIGURE-GROUND RESPONSE MODULATION IN HUMAN VISUAL CORTEX

Emil Olsson¹ (emilolsson94@gmail.com), Michael Epstein^{2,3}, Juneau Wang³, Karen Tian³, Brian Maniscalco¹, Jennifer Motzer³, Angela Shen¹, Minh Nguyen¹, Meera Sriram¹, Sydney Liu¹, Olenka Graham Castaneda¹, Maggie Zhang¹, Yuzheng Wu¹, Diana Gamboa³, Richard Brown⁴, Victor A. F. Lamme⁵, Hakwan Lau^{6,7}, Biyu J. He⁸, Jan W. Brascamp⁹, Ned Block^{10,11}, David Chalmers^{10,11}, Rachel Denison³, Megan Peters^{1,12,13,14,15}, ¹Department of Cognitive Sciences, University of California Irvine, ²Center for Brain Imaging, New York University, ³Department of Psychological & Brain Sciences, Boston University, ⁴Department of Humanities, LaGuardia Community College; The Graduate Center, City University of New York, ⁵Department of Psychology, University of Amsterdam, Amsterdam, ⁶Department of Biomedical Engineering and Department of Intelligent Precision Healthcare Convergence, Sungkyunkwan University, ⁷RIKEN Center for Brain Science, ⁸New York University Grossman School of Medicine, ⁹Department of Psychology, Michigan

State University, ¹⁰Department of Philosophy & Center for Neural Science, ¹¹Department of Psychology, New York University, ¹²Department of Logic & Philosophy of Science, University of California, ¹³Center for the Neurobiology of Learning & Memory, University of California, ¹⁴Center for Theoretical Behavioral Sciences, University of California, ¹⁵Program in Brain, Mind, & Consciousness, Canadian Institute for Advanced Research

Recurrent processing in early visual cortex underlies figure-ground segregation and has been proposed as a neural correlate of visual awareness. Human fMRI studies have demonstrated figure-ground modulation in early visual cortex, but how this modulation varies with stimulus strength and across visual areas V1-V3 is not well characterized. Here, we measure figure-ground modulation across human V1-V3 using carefully controlled stimuli. We collected fMRI data from 27 participants across two independent sites (BU n=15, UCI n=12). Using an event-related design, participants viewed briefly-presented (250 ms) texture-defined figures at five stimulus strength levels, behaviorally calibrated per participant. Critically, the stimuli control for low-level confounds, with matched luminance across strength levels and elimination of T-junctions at figure edges. Using preregistered behavioral and analytic pipelines with pRF mapping and a spatial localizer, we identified voxels in V1-V3 responsive to figure locations and computed figure-ground modulation (response to figure-present minus figure-absent). Voxels were grouped as “figure-positive” (enhanced response to figure presence) or “figure-negative” (suppressed response) based on this modulation. Across V1-V3, roughly half of the figure-responsive voxels showed figure-positive modulation and half showed figure-negative modulation, a ratio that was consistent across visual areas and robust to different classification criteria. We examined how figure-ground modulation varied as a function of stimulus strength and across the visual hierarchy. In both figure-positive and figure-negative voxels, figure-ground modulation responses remained stably present across areas V1-V3 (with numerically largest responses in V1), and increased modestly with stimulus strength, suggesting that figure-ground segregation involves coordinated and stable excitatory and inhibitory processes. This characterization of distinct figure-positive and figure-negative response populations, and their consistency across visual areas and stimulus strengths, sets the stage for future studies testing the functional role of figure-ground modulation in visual processing and experience.

Acknowledgements: Templeton World Charity Foundation 0567 to BH, JB, NB, DC, MP, and RD, startup funding from the University of California Irvine and support from the Canadian Institute of Advanced Research to MP, startup funding from Boston University to RD, and BU UROP funding to DG.

TALK 2, 4:30 PM, 15.22

INVESTIGATING NEURAL BASIS OF SERIAL DEPENDENCE USING EEG AND OCULAR TRACKING TASK

Bao Hong^{1,2} (bh2378@nyu.edu), Jing Chen², Li Li^{1,2}; ¹East China Normal University, Shanghai, China, ²New York University Shanghai

Serial dependence is a phenomenon where perception is biased toward recent experience. While this effect is well documented behaviorally, its neural processes remain unclear. To investigate this, we combined eye-tracking and EEG recordings while participants tracked the step-ramp motion of a moving target with its motion direction randomized on each trial. Behavioral analysis revealed robust serial dependence: pursuit directions at initiation were systematically biased toward previous-trial target motion direction, followed by a later repulsive adaptation effect. This dynamic pattern suggests that serial dependence in ocular tracking arises at early stages of visual processing. To uncover its neural basis, we conducted time-resolved multivariate pattern analysis (MVPA) on the EEG signals. While current-trial motion direction could be decoded from EEG signals, previous-trial motion direction also showed a higher-than-chance decoding accuracy, emerging around ~100 ms after stimulus onset and strongest for occipital-parietal electrodes. This suggests that perceptual history might be reactivated within visual cortex early in processing. We next analyzed EEG activity during both stimulus presentation and the inter-trial interval (ITI) using a time-resolved regression with motion direction as the predictor. During stimulus presentation, the regression weights increased sharply after stimulus onset and decreased at the end of trials. During the ITI, the weights showed a gradual rise and exhibited rhythmic fluctuations. Spectral analysis confirmed a significant alpha-band component peaked ~10.5Hz, strongest for occipital-parietal electrodes. Sliding-window analysis revealed increasing alpha amplitude during the ITI, suggesting that past motion information might be retained through rhythmic activity. Together, these findings provide converging behavioral and neural evidence that serial dependence in ocular tracking arises early in visual processing. The retaining of information from previous trials might be linked to rhythmic alpha-band activity in the visual cortex. These results offer insight into how the brain dynamically preserves and reactivates recent motion history to guide ongoing perceptual responses.

TALK 3, 4:45 PM, 15.23

SLOW CHANGE BLINDNESS FROM SERIAL DEPENDENCE

Haley Frey^{1,2}, David Whitney^{1,2,3}; ¹Herbert Wertheim School of Optometry and Vision Science Program, University of California, Berkeley, ²Department of Psychology, University of California, Berkeley, ³Helen Wills Neuroscience Institute, University of California, Berkeley

Slow change blindness, when attentive observers fail to notice large changes that happen gradually, raises questions about how dynamic visual information is combined across time. One plausible strategy for integration is serial dependence: using information from the recent past to inform current perception. Here, we investigate continuous serial dependencies in perception of a cartoon object that slowly changes hue. Across three experiments, we measure perception at each moment in a morph by accumulating one-shot hue reports from a large number of observers who each viewed a different extent of the total slow hue change. Across participants, the entire morph was probed. We find that overall reported hue perception is biased towards the past (Experiments 1 and 2). Additionally, and more importantly, we find that the bias increases as more of the morph is experienced, consistent with a dependence on the preceding information at each moment (Experiments 1 and 2). By explicitly asking whether observers noticed the change, we verify that the resulting illusion of stability produces slow change blindness (Experiment 2). We show that reducing serial dependence via a change in object identity reduces the bias in hue reports, confirming that the measured bias is indeed serial dependence, rather than hysteresis or averaging (Experiment 3). Lastly, in a multi-trial replication, we show that observers who experience repeated trials report a smaller yet still significant bias, suggesting that continuous serial dependence persists despite knowledge of the task (Experiment 4). Overall, we provide evidence that serial dependence actively biases perception during gradual changes, producing slow change blindness.

National Institutes of Health grant T32EY007043; National Institutes of Health grant R01CA236793

TALK 4, 5:00 PM, 15.24

FEEDBACK-MEDIATED PRIOR INTEGRATION UNIFIES GESTALT PERCEPTUAL ORGANIZATION

Tahereh Toosi¹ (tahereh.toosi@columbia.edu), Kenneth D. Miller²; ¹Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY USA, ²Kavli Institute for Brain Science, Columbia University Irving Medical Center, New York, NY USA

Gestalt principles such as closure, similarity, continuation, and figure-ground segregation have traditionally been characterized as distinct perceptual rules. We present evidence that these phenomena emerge from a single computational mechanism: iterative prior integration using cortical feedback pathways. We developed Prior-Guided Drift Diffusion (PGDD), an algorithm that repurposes the same feedback connections used during learning to refine neural activations during inference. PGDD iteratively updates representations by gradient ascend updates moving away from a noisy representation of the input, constrained to remain near the sensory evidence. Applied to pre-trained convolutional neural network trained for robust object recognition on ImageNet (e.g. ResNet50), PGDD reproduced classic Gestalt demonstrations without additional training or architectural modifications. For closure, networks presented with Kanizsa square

inducers generated clear illusory contours completing missing edges. We tested large scale recurrent models (cornet-based or predictive processing) as controls, and they showed no square completion. Similarity grouping and good continuation emerged through the same mechanism, linking elements sharing visual features or completing interrupted line segments across gaps. Figure-ground segregation experiments revealed prior-dependent interpretation: when presented with the Rubin face-vase illusion, object-trained networks generated vase-like patterns while face-trained networks produced face-like features through identical PGDD processing. Control experiments confirmed this specificity: networks trained on faces or scenes failed to generate contour completion for Kanizsa square input, while object-trained networks succeeded. These effects matched biological signatures of Gestalt processing, including delayed response profiles and feedback dependence. The results suggest that diverse Gestalt phenomena reflect a unified process of integrating learned statistical regularities with sensory input through feedback pathways. Rather than requiring specialized mechanisms for each principle, perceptual organization emerges from how neural networks dynamically access implicit priors acquired during training to interpret ambiguous or incomplete sensory evidence.

T.T. is supported by NIH 1K99EY035357-01 and NSF DBI-2229929. This work was also supported by the Simon Foundation and Gatsby Charitable Foundation GAT3708.

TALK 5, 5:15 PM, 15.25

CONSTANT CURVATURE OPERATORS IN HUMAN VISION

Austin Phillips¹, Philip Kellman¹; ¹University of California, Los Angeles

A central challenge in vision science remains understanding how early, rapidly changing signals—from orientation-tuned contrast detectors, for example—are transformed into stable, symbolic descriptions of object properties. Recent research has investigated this question in the perception of contours and 2D shape. Evidence suggests that an initial symbolic representation of contour shape consists of segments of constant curvature (Baker, Garrigan, & Kellman, 2021). These representations may be produced by responses from banks of constant-curvature filters built from oriented units, linked by constant turning angles, in early visual processing (Kellman, Garrigan, & Erikhman, 2013; Baker & Kellman, 2021). This initial symbolic encoding of arbitrary smooth contours into sequences of constant-curvature segments may provide a concise and versatile description of shape that provides a foothold for further operations. Together with neurophysiological findings regarding curvature sensitivity in V4 (e.g., Dumoulin & Hess, 2007; El-Shamayla & Pasupathy, 2016), this framework suggests that there may exist hard-wired detectors for constant curvature that operate in parallel across broad regions of the visual field. To investigate this hypothesis and explore properties of hypothesized curvature detectors in the visual pathway, we developed a variant of the path-detection task (Field, Hayes, & Hess, 1993).

Observers judged small arrays containing three-element Gabor targets. Triplets forming constant-curvature arcs were contrasted with (a) relatable but non-constant-curvature arrangements with uniform curvature polarity, (b) relatable triplets with alternating curvature polarity, and (c) “null” configurations in which elements were non-relatable. Detection performance was substantially higher for constant-curvature triplets relative to all other conditions. Both relatable types exceeded performance on null triplets, and this pattern held at 600, 1000, and 1400ms exposures. These findings support the presence of dedicated curvature mechanisms that efficiently encode smooth contour structure, and may provide the earliest symbolic descriptions of contour shape.

TALK 6, 5:30 PM, 15.26

A RAPID AND AUTOMATIC NEURAL CORRELATE OF LIKING FOR VISUAL ARTWORKS

Daniel Kaiser¹, Sanjeev Nara², Gustavo Menegon³, Vaishali Goyal³, Philipp Fieger¹; ¹Justus Liebig University Giessen, ²IIT Mandi, ³Lusofona University, University of Barcelona, and University of Paris Cité

When we appreciate visual art, we often take our time: We visit museums or galleries and deliberately engage with artworks and their meaning. This deliberate engagement in turn shapes whether we like an artwork or not. Yet, our liking might at least partly be shaped by incidental encoding of the artworks' visual features, a process initially independent from sustained engagement. Here, we show that liking for visual art can be predicted from rapid and automatic responses in the visual brain. We recorded EEG responses while participants viewed 999 artworks from the Vienna Art Picture System (VAPS), which features complex artworks that span different categories (from still lifes to portraits and landscapes) and art styles (from renaissance to impressionism and expressionism). Artworks were shown with a 300ms stimulus-onset-asynchrony (150ms on, 150 off), and participants responded to occasional target images. Using representational similarity analysis, we predicted the geometry of neural representations from the artworks' similarity in liking (based on ratings from a different group of observers contained in the VAPS), while simultaneously controlling for similarities in other properties (category, art style, valence, arousal, and familiarity). Crucially, neural responses emerging within the first 150ms were robustly connected to liking, even though the task did not afford any liking judgments. Rapidly emerging visual representations thus predict liking for visual artworks, highlighting the contribution of basic visual feature processing. Strikingly, this early neural correlate of liking was also observed in a second experiment, in which we presented the same 999 artworks with a 50ms stimulus-onset-asynchrony. This shows that even under ultra-fast presentation conditions, automatic neural responses rapidly indicate liking. While fully appreciating an artwork may still need deliberate and prolonged engagement, the perceptual basis of whether we like or dislike an artwork seems to be laid in a split second.

DK is supported by the DFG (KA4683/6-1, project number 536053998) and an ERC Starting Grant (PEP, ERC-2022-STG 101076057). This work is further supported by the DFG under Germany's Excellence Strategy (EXC 3066/1 "The Adaptive Mind", project number 533717223).

TALK SESSION: SATURDAY, MAY 16, 2026, 8:15 – 9:45 AM, TALK ROOM 1

Eye Movements: Neural dynamics, perceptual changes

Moderator: James Bisley, UCLA

TALK 1, 8:15 AM, 21.11

THE PHYSIOLOGICAL CORRELATES OF CONSCIOUS AND UNCONSCIOUS VISUAL AND AUDITORY PROCESSING: AN MEG AND EYE-TRACKING STUDY
Plyfaa Suwanamalik-Murphy¹ (plysuwana@gmail.com), Sharif Kronemer¹, Victoria Gobo^{1,2}, Javier Gonzalez-Castillo¹, Amaia Benitez-Andonegui³, Anna Namyst³, Peter Bandettini^{1,4}; ¹Section on Functional Imaging Methods, Laboratory of Brain and Cognition, National Institute of Mental Health (NIMH), National Institutes of Health (NIH), Bethesda, Maryland (MD), United States of America (USA), ²Baylor College of Medicine, Houston, Texas, USA, ³MEG Core Facility, NIMH, NIH, Bethesda, MD, USA, ⁴Functional Magnetic Resonance Imaging Core Facility, NIMH, NIH, Bethesda, MD, USA

Distinguishing conscious from unconscious processing is a major challenge for uncovering the neural mechanisms of conscious perception and guiding the assessment of conscious level in patient groups with disordered consciousness. Here, we explore neural and non-neural physiological metrics and their signal properties that may differentiate conscious and unconscious sensory processing in vision and audition. We explored eye metrics (pupil size, blinking, microsaccades) and cortical dynamics measured with magnetoencephalography (MEG). Two major signal characteristics were assessed: (1) magnitude and (2) timing (e.g., the latency of change from stimulus onset and duration of response). Healthy adult participants performed a visual (N = 33) and auditory (N = 37) stimulus detection task with simultaneous eye-tracking, pupillometry, and MEG. Sub-, near-, and supra-threshold stimuli were presented on the left or right side of a screen (face images) or in the left or right ear (tones). Participants indicated when they perceived a stimulus and its location (on-screen or in which ear). Our results indicate strong correspondence across behavioral, eye, and MEG responses and between visual and auditory modalities. As stimulus intensity increased, perception rate increased, and reaction time decreased. Perceived stimuli elicited increased pupil size and blink rate and decreased microsaccade rate in auditory and visual modalities, relative to unperceived stimuli. Eye metric responses changed

systematically with stimulus intensity regardless of conscious perception: as intensity increased, response amplitudes increased, and latencies decreased. Similarly, for visual and auditory stimuli, MEG activity showed larger magnitude and more prolonged responses for perceived versus unperceived stimuli. However, regardless of perception, MEG signal magnitude and latency were similarly modulated by stimulus intensity in occipital (visual) and temporal (auditory) scalp regions: larger and more rapid MEG responses as stimulus intensity increased. These findings help clarify perception-dependent and independent components of eye and cortical signals in conscious and unconscious vision and audition.

NIH Intramural Research Program ZIAMH002783

TALK 2, 8:30 AM, 21.12

FREQUENCY SELECTIVE POST-MICROSACCADIC PERCEPTUAL ENHANCEMENTS IN THE FOVEA

*Zoe Stearns^{1,2}, Krishnamachari Prahalad^{1,2}, Martina Poletti^{1,2};
¹Center for Visual Science, ²University of Rochester*

Microsaccades compensate for non-uniform sensitivity within the foveola by precisely redirecting the preferred locus of fixation (PLF) and by enhancing fine spatial vision at their target location before their onset. Yet, whether and how visual perception is also modulated upon microsaccade landing remains unclear. To address this, we tested post-microsaccadic visual discrimination of different spatial frequencies at the landing site and further away. Eye movements were recorded with a high-resolution Dual-Purkinje Image eye tracker as observers freely viewed an array of five gabors spanning 2.5° (0.5° center-to-center) at the center of the display. At variable times all the gabors changed phase and one also changed orientation ($\pm 45^\circ$). Subjects (N=10) reported the direction of the brief orientation change (2AFC). Stimulus contrast was adjusted so that performance was at threshold when the orientation-change occurred during fixation at the PLF. Subjects spontaneously performed frequent microsaccades during free viewing. Visual discrimination performance was then evaluated at baseline when subjects were maintaining fixation and 0-130ms after microsaccade landing. Performance was assessed both for stimuli near (within 15 arcmin) and far (20–40 arcmin) from the PLF and for low (4 cpd) and high (16 cpd) spatial frequencies. Visual discrimination was enhanced relative to baseline immediately after microsaccade landing (rm-ANOVA: $F(1,9)=8.19$, $p=0.019$). This enhancement was spatially localized around the PLF ($F(1,9)=11.83$, $p=0.007$) and selective for higher spatial frequencies ($F(1,9)=12.69$, $p=0.006$). For high spatial frequencies, the post-microsaccadic gain increased over time, reaching significance ~40 ms upon landing, whereas the performance for low spatial frequencies remained constant in the period following microsaccade landing. These results show that microsaccades are associated with post-microsaccadic perceptual enhancements, these modulations are spatial-frequency selective and are localized at the microsaccade landing site, indicating

that they are likely driven by extraretinal signals rather than temporal modulations of the visual input brought by the saccadic transient.

funded by Meta and NIH grant EY001319 to the Center for Visual Science.

TALK 3, 8:45 AM, 21.13

MOTOR PRIORITY ACCELERATES RECOVERY FROM GLOBAL SACCADIC INHIBITION

Paul Schmitthäuser^{1,2} (paul.schmitthaeuser@hu-berlin.de), Martin Rolfes¹, Nina M. Hanning¹; ¹Humboldt-Universität zu Berlin, ²Freie Universität Berlin

We generate saccadic eye movements at a baseline rate of approximately 2 to 4 per second. Sudden visual transients reliably interrupt this stream of eye movements, producing a sharp suppression of saccade initiation—a phenomenon known as “saccadic inhibition”—that is followed by a rebound that briefly exceeds baseline. Although saccadic inhibition has been demonstrated in many visual tasks, it remains elusive how this inhibition interacts with the unfolding dynamics of saccade planning in naturalistic free viewing, and whether it depends on the spatial congruency between a transient and the current motor plan. Participants performed an unrestricted search task in which they freely explored an array ($r \sim 14^\circ$) of 19 differently colored tiles until they found a hidden target. Tile colors signaled target probabilities and encouraged systematic, predictable gaze patterns. Brief luminance probes (8.3 ms) were flashed either at the current fixation location, the predicted saccade target, or a non-target location. The natural variability in probe timing relative to saccade onset allowed us to probe different phases of saccade preparation. Saccadic inhibition followed probe onsets at any location, regardless of congruency with the saccade target. Critically, however, recovery was faster when probes coincided with the upcoming saccade target than when they appeared at other locations. These results suggest that visual transients interact with naturalistic visual exploration in two temporally distinct ways: an initial, global pause that interrupts saccade initiation, followed by a spatially selective recovery that is facilitated when the transient is congruent with the current saccade plan. This pattern aligns with an architecture in which global inhibition temporarily halts saccade execution, while spatially selective motor priority is preserved and shapes the speed of recovery.

This research was supported by the European Research Council (ERC) under the European Union’s Horizon Europe research and innovation programme (project 101165343 – FREE-VIEW to NMH).

TALK 4, 9:00 AM, 21.14

TOP-DOWN MOTIVATION BOTH INCREASES AND DECREASES FEATURE INTERFERENCE FOLLOWING A SACCADIC

Tzu-Yao Chiu¹ (chiu.315@osu.edu), Julie D. Golomb¹; ¹The Ohio State University

Spatial attention is rarely static, but instead highly dynamic in real life. Recent studies have revealed that shifts of attention and eye movements incur consequences for ongoing feature perception, such as systematic feature-binding (swapping and mixing) errors. Here we investigated if these consequences are sensitive to top-down control, such that they could be flexibly mitigated when participants are motivated to improve their performance, or if feature interference reflects a more automatic byproduct of attentional remapping mechanisms following a saccade. In a pre-registered experiment, 40 healthy adults reported the color of an item appearing at a pre-cued spatiotopic (world-centered) location. Following the cue, participants made a saccade, after which an array of four colored items briefly appeared. One of the colored items appeared at the cued spatiotopic target location, one at the cue’s retinotopic (eye-centered) location, and two at control non-target locations. Crucially, across blocks we provided two different levels of performance-contingent reward (monetary incentives for precise spatiotopic target reports) to induce differences in top-down motivation. Compared to baseline lower-reward blocks, during higher performance-contingent reward (“bonus”) blocks, participants showed improved general performance and reduced subtle retinotopic mixing errors immediately after the saccade. However, we found the opposite effect on retinotopic swapping errors, with higher reward/motivation associated with increased retinotopic swapping errors early after the saccade, indicating that retinotopic swapping errors may reflect a structural limitation of the attentional remapping process. Overall, these results uncover the role of motivation in distinct attentional mechanisms underlying the remapping of spatial attention across saccades, and provide novel insights into the role of top-down motivation in dynamic spatial attention and visual perception more generally.

NIH R01-EY025648 (JG), NSF 1848939 (JG)

TALK 5, 9:15 AM, 21.15

QUESTIONING THE FUNCTIONAL VALIDITY OF PRE-SACCADIC REMAPPING IN LIP

James Bisley¹, Yelda Alkan¹, Alanna Morris¹; ¹UCLA

Despite constantly shifting our gaze, our perception remains stable. It has been proposed that predictive remapping of receptive fields (RFs) in the lateral intraparietal area (LIP) may play a role in maintaining this stability. A number of studies have shown that remapped responses can begin before the saccade: pre-saccadic remapping. To improve our understanding of remapping, we investigated how the remapped response relates to saccade length. We recorded LIP activity in 2 animals performing a saccade task in which a task irrelevant probe appeared in the post-saccadic RF before the animals made a 7, 14 or 21 degree saccade. RFs were mapped using flashed stimuli and a memory guided saccade task. Despite the care taken to map the RFs, we found that some units responded significantly to the probe appearing in the post-saccadic RF prior to the 7 deg saccade. For the main experiment, these conditions were excluded because the probe

was not presented outside the RF. Surprisingly, we found that pre-saccadic remapping only occurred in the excluded condition. No units for which the probe was presented clearly outside of the RF showed pre-saccadic remapping, and units that had pre-saccadic responses only had so for the excluded condition – none showed pre-saccadic remapping in the 14 and 21 deg saccade conditions. The few prior studies that have shown pre-saccadic remapping in LIP often do not describe the lengths of saccades nor provide detailed receptive field mapping methods, or only use RF “hot spots”. And remapping studies looking at other cortical areas either have not aligned data by saccade onset or have used slow infrared eye tracking systems and large analysis windows that do not allow for accurate timing estimates. We propose that pre-saccadic responses may be a result of within neuron processing rather than a fundamental feature of LIP remapping.

This work was supported by the National Eye Institute (R01 EY032863).

TALK 6, 9:30 AM, 21.16 - REMOTE PRESENTATION

OBJECT FILES ARE LOST AT SACCADIC SPEEDS

Melis Ince^{1,2} (melis.ince@hu-berlin.de), Carolin Hübner³, Martin Rolfs^{1,2}; ¹Humboldt-Universität zu Berlin, ²Berlin School of Mind and Brain, ³Technische Universität Chemnitz

Object motion that follows the velocity-duration-amplitude relationship of saccades can support perceived object continuity even when the motion itself is not consciously perceived (Ince et al., 2025). Here we tested whether object files—temporary representations of objects and their features (Kahneman et al., 1992)—survive motion at saccadic speed. In an adapted object-reviewing paradigm (Sasi et al., 2023), trials started with a brief preview of two digits displayed inside two Gabor patches. The patches then blanked and reappeared at the same (stationary) or new locations (apparent motion), or moved smoothly to their new locations (continuous motion). The duration of the blank/motion phase was set to match saccade kinematics for the movement distance (6 dva) used in the motion conditions (mean velocity: 39 ms, 154.3 dva/s; peak velocity: 25 ms, 239.4 dva/s), or a previously used lower velocity (300 ms, 20 dva/s; Sasi et al., 2023). After the blanking/motion, a target digit appeared in one of the two Gabors, which participants (n = 19) had to rapidly identify. This target digit was either one of the previewed digits (from the same or different object) or a new digit. At long durations (slow velocity), we replicated Object-Specific Preview Benefits (OSPBs)—shorter response times for digits previously shown in the same compared to a different object (Sasi et al., 2023). Consistent with object-file theory, these OSPBs were evident in the static and continuous conditions, but not in apparent motion, where the same and different objects could not be distinguished. OSPBs were absent for motion at saccadic speeds (mean and peak velocities). These results suggest that—although motion at saccadic speed can disambiguate object correspondence through spatiotemporal continuity (Ince et al., 2025)—object-file

representations (that include object features) do not survive these rapid retinal position changes.

This research was supported by the European Research Council under the European Union’s Horizon 2020 research and innovation program (Grant 865715), and the Deutsche Forschungsgemeinschaft (grants RO3579/8-1, RO3579/10-1).

TALK SESSION: SATURDAY, MAY 16, 2026, 8:15 – 9:45 AM, TALK ROOM 2

Perceptual Organization: Segmentation and grouping

Moderator: Jeremy Wilmer, Wellesley College

TALK 1, 8:15 AM, 21.21

ENSEMBLE FEATURE DISTRIBUTIONS SUPPORT VISUAL STABILITY

Vladislav Khvostov¹, Julie Golomb¹; ¹Department of Psychology, The Ohio State University

To bypass attentional/working memory limitations, the visual system can represent statistical information about object groups (ensembles). Recent work revealed that observers explicitly represent feature distributions, i.e., relative frequencies of each feature value in the visual field. However, these feature distribution representations are abstract: by pooling features within the entire visual field, the visual system loses information about individual objects’ locations. Might this ability help explain the visual stability paradox (people’s perception of the world as stable despite drastic changes in the retinal positions of visible objects due to eye/head movements)? To test this, we had observers perform a transsaccadic change detection task. Observers viewed a display containing 36 colored disks (500ms), then executed a saccade, during which the display changed. Observers reported the number of disks that changed color between the two displays. In the Distribution Change condition, the color changes altered the shape of the color distribution (along the Gaussian-Bimodal spectrum, identical mean/range). In the Spatial Swap condition, we preserved the exact ensemble color distribution but swapped the spatial positions of individual colors. These conditions were matched for the number of disk changes and the total item-wise color difference. In the Spatial Swap condition, observers reported very few changes, even in the most extreme case, where 78% of the disks changed color, indicating that as long as the ensemble feature distribution is preserved across two scenes, participants perceive the environment as stable. In contrast, observers were highly sensitive to transsaccadic changes in the Distribution Change condition. We conclude that ensemble feature distribution representations play an important role in visual stability: because people make frequent eye/head movements, it may be more beneficial for the visual system to represent ensembles as feature

distributions and not store individual members' locations. As a result, perception of visual stability relies on feature distribution representations.

Supported by grant NIH R01-EY025648 (JG)

TALK 2, 8:30 AM, 21.22

PRINCIPLES OF LOCAL AND GLOBAL GROUPING THAT UNDERLIE SEGMENTATION OF NATURAL TEXTURE IMAGES

Wilson Geisler¹, Abhramil Das^{1,2}; ¹University of Texas at Austin, ²Brain Institute UFRN

Humans have a remarkable ability to segment natural scenes into physically meaningful regions. This sophisticated process contains many components, some "high level" (exploiting recognition of materials, objects and/or scene context), and others "low level" and largely independent of specific prior experience. The low-level components are essential in that they are required for initiating recognition processes, and for learning to recognize new materials, objects, and contexts. Our aim is to develop hierarchical Bayesian observer (HBO) models of natural texture segmentation that are biologically plausible, account for the statistical properties of natural scenes, and do not depend on prior experience. The current HBO model consists of five grouping steps consistent with Gestalt principles: 1. local similarity grouping with local normalization, 2. mutual similarity grouping (local grouping is strengthened if the neighboring regions are similar to the same set of other regions), 3. transitive grouping (good continuation), 4. confidence grouping (neighboring regions far from the same-different decision boundary guide grouping of regions near the boundary), and 5. region grouping (similarity grouping of the regions from the initial segmentation). We find that the local similarity grouping process, trained to maximize accuracy based on natural scene statistics and Bayesian decision processes, predicts human local similarity grouping accuracy. We then find that the four additional steps are able to accurately segment images with randomly-shaped regions containing arbitrary natural textures. The success of the model depends on all of the steps, but especially on local-similarity grouping and transitive grouping. We also find that the transitive grouping allows correct segmentation of non-stationary (e.g., slanted in depth) texture regions. Further, we find that when illumination varies across the image, local normalization enables both correct texture segmentation and estimation of the illumination change. Finally, we show that state-of-the-art deep-network models fail on these stimuli where our model succeeds.

TALK 3, 8:45 AM, 21.23

REPRESENTATIONS OF SINGLE OBJECTS, HOMOGENEOUS ENSEMBLES, AND HETEROGENEOUS ENSEMBLES

Patxi Elosegi¹, Boyang Hu¹, Marvin Chun¹, Yaoda Xu¹; ¹Yale University

Visual scenes are inherently hierarchical, containing discrete objects (e.g., a bird) and collections of objects that can be efficiently summarized using ensemble statistics (e.g., the average motion of a flock of birds). Although prior psychophysical work shows that objects and ensembles can be perceived in parallel, it remains unclear how they are represented together in the brain, especially in comparison to heterogeneous ensembles with mixed categories of objects. In Experiment 1, we used a block-design fMRI paradigm and presented participants either a single object from one of three categories or an 8-item homogeneous ensemble comprised of one of three object categories. Participants performed an orthogonal image-jitter-detection task while viewing the stimuli. Using multivoxel-pattern analysis, we examined responses from functionally-defined early visual areas, object-selective areas, scene-selective areas, and posterior parietal cortex. Across all ROIs, decoding of single object categories generalized strongly to ensemble categories and vice versa, suggesting shared representations. Strikingly, adding a single outlier to an otherwise homogeneous ensemble significantly altered the representation compared to those of single objects and homogeneous ensembles, as revealed by limited cross-decoding from the latter two to the former. To further characterize the neural representation of heterogeneous ensembles, in Experiment 2, we presented stimuli containing nine ratios of objects from two categories, including homogeneous ensembles (8:0 and 0:8) and all heterogeneous combinations (7:1, 6:2, 5:3, 4:4, 3:5, 2:6, and 1:7). Across all ROIs, we replicated the main findings of Experiment 1. Critically, a U-shaped representational structure emerged for the heterogeneous ensembles, indicative of an orthogonal encoding of object category and ratio. Together, these findings show that single objects and homogeneous ensembles share neural representations, whereas heterogeneous ensembles utilize a coding scheme that preserves information about both the constituent object categories and their relative proportions.

Supported by NIH Grant R01EY030854 to YX.

TALK 4, 9:00 AM, 21.24

FROM REPRESENTATION TO RECONSTRUCTION: TEMPORAL SIGNATURES OF GENERATIVE ASSEMBLY

Yaxin Liu¹ (yl1913@georgetown.edu), Yuval Hart², Adam Green¹;

¹Georgetown University, ²Hebrew University of Jerusalem,

³Georgetown University

Humans rapidly reconstruct object parts and relations from mere silhouettes. Yet this "inverse graphics" problem is computationally intractable given the combinatorial explosion of part configurations. How do humans efficiently solve this vast search space in real time? Prior work has mainly examined how humans and models infer a

single global shape or category, leaving open how multi-step generative reconstruction occurs. We hypothesized that efficient generation relies on detecting high-constraint parts (e.g., sharp vertices with unique geometry) that act as "anchor points" to reduce the search space. By contrast, low-constraint and ambiguous parts may demand internal mental simulation or offload computation to continuous sampling. To test this, 90 participants reconstructed persons, animals, objects, and geometric shapes from silhouettes using a computerized Tangram task—a generative assembly task. Participants used translations, rotations, and flips to match the target contours. We recorded high-resolution mouse trajectories to analyze the fine-grained temporal profile of real-time reconstruction. For each target contour, we computed a constraint index capturing the number of valid part-to-contour mappings (high constraint: anchors with near 1:1 mappings; low constraint: ambiguous regions with many valid decompositions). We found that high-constraint regions triggered short-latency, ballistic bursts of actions clustered in time. Conversely, low-constraint regions elicited threefold longer pauses and non-clustered moves, consistent with a shift to slower mental simulation over multiple competing assemblies. These findings suggest that generative reconstruction may be neither purely feedforward nor purely analysis-by-synthesis. Rather, high-constraint anchors are used to rapidly prune the hypothesis space, allowing slower, simulation-based mechanisms to resolve ambiguities without the need for continuous physical sampling.

TALK 5, 9:15 AM, 21.25

WHEN TWO BECOME ONE: A "SAME-OBJECT ADVANTAGE" THAT SPANS ACROSS SHADOWS AND THEIR CASTING OBJECTS

Albert Z. K. Li¹, Dominic Alford-Duguid¹, Joan Danielle K. Ongchoco¹;
¹University of British Columbia

Visual experience is populated by "things" (solid objects) and "stuff" (liquids or granular substances). But then there are shadows. They are informative, but immediately discounted. They can carry the structure of their objects, but are deformable and often distorted. Perhaps most puzzling: shadows may be segmented as bounded regions that give rise to object-based effects, yet ultimately, cannot exist in the physical world without their casting objects. What is the relationship between shadows and their objects in visual processing? To explore this, we used a classic test of object-based representations: the same-object advantage. Observers viewed 3D-rendered scenes with a light, a casting object, and a receiving surface. Two probes flashed quickly, either with the same or different orientations, and observers reported whether they were identical. Observers were more accurate when probes appeared on the same object (versus when split between the object and surface) and on the same shadow (versus when split between the shadow and surface), replicating the traditional same-object advantage. But curiously, the usual cost of switching between object representations was not present when probes were split between the shadow and its object — a "shadow-object advantage" —

which should not occur if shadows and objects are being treated as "equals" (i.e., separate bounded regions). This shadow-object advantage held even when the shadow was incongruent with its object (perhaps because once tagged as the object's shadow, its exact shape is discounted), and remarkably, even when it was physically separated from its object (by being projected onto a wall). And when we "break" the shadow (by replacing it with an outlined dark mask of the identical shape and position), this shadow-object advantage then disappears. Altogether, these results suggest a new visual category: shadows as *derivative individuals* may be bound together in visual processing with their casting objects.

TALK 6, 9:30 AM, 21.26

GRAPHS OF AVERAGES EXAGGERATE AND SOW DISAGREEMENT

Jeremy Wilmer¹ (jwilmer@wellesley.edu), Sarah Kerns^{1,2}, Lily Widdup³, Ken Nakayama^{3,4}; ¹Wellesley College, ²Dartmouth College, ³Harvard University, ⁴UC Berkeley

Scientific communication must be both valid (accurate) and reliable (consistent). Graphs are among the most widely used tools for conveying scientific evidence, yet we show that a very common practice—displaying only group averages—systematically fails on both counts. Using a novel drawing-based response measure, participants sketched the data points they believed underlay plotted averages. Across 40 replications spanning plot types, data domains, and levels of quantitative expertise, interpretations of average-only graphs were both severely exaggerated and strikingly inconsistent. Perceived effect sizes frequently exceeded plausible scientific bounds, and disagreement among viewers interpreting the same single graph was greater than that typically observed across entire research literatures. In most tested cases, interpretations were systematically less accurate than blind guesses, indicating that graphs of averages can satisfy the formal criteria for misinformation. We identify a simple and effective remedy: displaying individual data points alongside the mean. Contrary to concerns that increased visual detail would overwhelm or confuse viewers, data-showing graphs produced substantially greater accuracy, stronger consensus among viewers, and more positive subjective evaluations. Together, these results demonstrate that a widely used visualization practice can actively distort scientific understanding and that routinely showing raw data offers a practical, evidence-based improvement to the validity and reliability of scientific communication.

TALK SESSION: SATURDAY, MAY 16, 2026, 10:45 AM – 12:30 PM, TALK ROOM 1

Spatial Vision: Neural mechanisms, modeling

Moderator: David Brainard, University of Pennsylvania

TALK 1, 10:45 AM, 22.11

A NONLINEAR, SPATIO-TEMPORO-CHROMATIC, IMAGE-COMPUTABLE MODEL OF THE ON-CENTER MIDGET RETINAL GANGLION CELL MOSAIC ACROSS THE RETINA

Nicolas Cottaris¹, Brian Wandell², David Brainard¹; ¹University of Pennsylvania, ²Stanford University

Our previous work introduced a framework for synthesizing biologically plausible mosaics of linear spatio-chromatic receptive fields (RFs) for ON-center midget Retinal Ganglion Cells (mRGCs). We validated the spatial properties of these linear RF mosaics against in vivo and in vitro physiological data from macaques, and we further showed that a computational observer with access to the model output recapitulates aspects of human psychophysical spatio-chromatic contrast detection. To enhance realism and predictive power, we now extend the model by incorporating key non-linear and temporal mechanisms. Specifically, (a) the nonlinear cone phototransduction cascade, (b) mRGC temporal dynamics, and (c) static non-linearities (e.g., contrast gain control) at the mRGC RF center and surround components as well as the mRGC output stage. These additions preserve the model's agreement with published spatial tuning data in macaque mRGCs at the stimulus conditions under which the macaque data were collected (temporal frequency: 4 Hz, contrast: 25%, background luminance: 40 cd/m²). Crucially, the non-linear and linear models significantly diverge under conditions where adaptation effects are strong, e.g., low temporal frequency (< 2 Hz), high background luminance (> 100 cd/m²), and high contrast (> 75%). This divergence reveals the critical importance of incorporating non-linear mechanisms in predicting performance under diverse stimulation conditions. Furthermore, we propose a novel method for integrating mRGC temporal dynamics with the phototransduction cascade, showing that a triphasic temporal impulse response filter, placed after the cone outer segment photocurrent generation stage, is sufficient to reproduce measured mRGC stimulus-referred dynamics. The updated model represents a significant step toward predicting human performance across a wider range of spatio-temporo-chromatic psychophysical tasks.

TALK 2, 11:00 AM, 22.12

CONTRAST-DEPENDENT REORGANIZATION OF CENTER-SURROUND ORIENTATION TUNING

Lisa Schwetlick^{1,2}, Peter Neri^{2,3}; ¹EPFL, ²ENS, ³IIT

Visual systems flexibly adjust local processing in response to a variety of contextual signals. For example, orientation tuning is known to depend strongly on surrounding orientations. Here, we asked (1) how orientations in the surround influence detection of a central orientation target, and (2) how this center-surround interaction reorganizes as a function of relative contrast. Such shifts may reveal transitions between linear and nonlinear computational regimes. Observers

performed a 2-AFC orientation-discrimination task in which both the central target patch and a surrounding annulus were independently perturbed with orientation noise. Reverse-correlation analysis yielded tuning curves for the center and surround separately. By varying center and surround contrasts, we systematically examined not only the strength of contextual modulation but also how the shapes of the tuning curves change across contrast regimes. Tuning curves computed from the center noise show that target-aligned noise in the center biases decisions toward the target orientation, whereas orthogonal noise has the opposite effect. Introducing target-aligned or orthogonal surrounds causes suppression of the corresponding peak of the central tuning curve (i.e., same-orientation inhibition). Surround tuning curves at low contrasts similarly show a positive peak at target-aligned orientations. At high contrasts, however, this pattern reverses: tuning shifts toward a regime that emphasizes contrastive information in the surround. We also find considerable individual differences between observers. We compared the performance of several candidate models—including linear templates, energy-based nonlinear models, and divisive normalization models. No single strategy can account for behavior across all regimes. Instead, center-surround interactions appear to reorganize systematically with contrast, revealing dynamic transitions between linear and nonlinear processing strategies in early vision.

TALK 3, 11:15 AM, 22.13

ADAPTATION REVEALS A GAP BETWEEN CONTRAST DISCRIMINATION AND CONTRAST RESPONSE

Sanga Cho¹ (sangacho@bu.edu), Sam Ling¹; ¹Boston University

Perceptual discrimination is linked to nonlinearities in neural coding, central to Weber-Fechner law. For instance, contrast discrimination is thought to arise from the nonlinearity between stimulus contrast and the neural contrast response function (CRF), with the canonical dipper-like shape of the threshold-versus-contrast (TvC) function attributed to the inverse slope of the CRF. However, this standard model linking brain with behavior has been difficult to evaluate, as nonlinear CRFs have been challenging to measure with fMRI. Here, we leverage a recently developed adaptation paradigm that captures nonlinearity of CRFs, to test models linking population CRFs to the TvC. To do so, each fMRI run started with initial 60-sec adaptation period to full visual field stimuli at 5%, 8%, or 32% contrast. Participants then viewed a sequence of stimuli at contrasts that straddled the adaptor contrast, interleaved with top-up adaptors, yielding voxel-wise CRFs at three adaptation conditions. We then compared these CRFs to TvC functions at the same adaptor levels, in the same participants. TvC functions were estimated via a staircase procedure, with participants judging which of two sequential stimuli had higher contrast, across different pedestal contrasts at the same three adaptation contrasts. We found that higher contrast adaptors substantially shifted neural sensitivity toward higher contrasts, increasing the semi-saturation across V1-V3. Psychophysical TvC dips also shifted toward higher contrasts with adaptation. While both neural and behavioral measures

demonstrated ordinal alignment in how adaptation recenters sensitivity, the standard model still fell short. Model predictions based on ROI-averaged CRFs produced contrast sensitivity that was quantitatively misaligned with true psychophysical data, across adaptation conditions. Instead, voxel-wise analyses of population CRFs suggest that incorporating optimal pooling mechanisms may provide a principled bridge, linking neural contrast responses to contrast discrimination.

TALK 4, 11:30 AM, 22.14

ORTHOGONAL NEURAL GEOMETRIES OF BASIC VISUAL FEATURES IN MACAQUE V1

Cong Yu¹, Xin Wang², Shi-Ming Tang²; ¹Zhejiang University, ²Peking University

The classical ice-cube model of Hubel and Wiesel proposes that V1 neurons are spatially organized into orthogonal maps of orientation and ocular dominance to optimize wiring efficiency. However, extending this framework to include additional features such as spatial frequency imposes geometrical constraints on how these features can be spatially arranged on the two-dimensional cortical surface. A recent two-photon calcium imaging study of ours (Zhang et al., Cerebral Cortex, 2024) found that cellular-resolution maps of orientation, spatial frequency, and ocular dominance in macaque V1 lack consistent orthogonal or parallel spatial arrangements, raising the question whether these orthogonal features are instead represented in population activity space. We applied principal component analysis (PCA) to these and additional datasets. We found that population responses formed near-orthogonal geometries in representational space, supporting the idea that feature encoding relies more on population-level activity than spatial layout. This orthogonal structure remained robust to dimensionality increase and was absent in response-shuffled control data, in which feature axes collapsed to chance-level alignment. Furthermore, artificially disrupting orthogonality, either by aligning feature axes or randomizing trial positions in PCA space, severely impaired the decodability of stimulus features, demonstrating that orthogonal representations are critical for maintaining feature separability. The same orthogonality was also replicated in the representational geometry of color and orientation in macaque V1 interblob neurons, but not in blob neurons that showed high color selectivity but poor orientation selectivity. These findings suggest that V1 population responses follow an orthogonal encoding geometry, which makes orthogonal representation of multiple stimulus features possible. Moreover, population codes, rather than spatial maps, better capture feature representations. This principle may also serve as an important benchmark for V1-inspired deep neural networks.

TALK 5, 11:45 AM, 22.15

VISUAL MASKING REDUCES SPATIAL BUT INCREASES TEMPORAL FIELD ASYMMETRIES: A SPEED-ACCURACY TRADEOFF STUDY

Nicholas Crotty¹ (ndc5688@nyu.edu), Marisa Carrasco^{1,2};

¹Department of Psychology, New York University, ²Center for Neural Science, New York University

[INTRODUCTION] Visual perception is better and processing time is faster along the horizontal than the vertical meridian (horizontal-vertical anisotropy, HVA), and along the lower than the upper vertical meridian (vertical meridian asymmetry, VMA). Such asymmetries have been shown using speed-accuracy tradeoff (SAT) protocols, which conjointly measure discriminability and information accrual rate (processing time). According to “interruption theories” of visual masking, backwards-pattern masks disrupt the accrual of visual information shortly after onset. Thus, here we investigate for the first time whether and how visual masking alters differences in discriminability and processing time among the cardinal meridians. [METHODS] We introduce masking to the SAT protocol from Carrasco & McElree (2001), via a local patterned-mask following stimulus offset on half of the blocks. Participants performed both feature and conjunction search with stimulus arrays (set sizes of 1, 4, 8) presented at the cardinal locations for 40ms, and the mask subsequently presented. A tone at different intervals (40ms-2s) indicates the opening of the response window. We fit the 3-parameter SAT model (intercept, slope, asymptote) to the data using both a hierarchical nested-modeling and a novel Bayesian generalized-linear-mixed-effects (GLME) approach, quantifying the effects of masking and cardinal meridian on each SAT parameter. [RESULTS] The candidate HVA model from the nested approach indicated that masking reduced the asymmetry in discriminability, but increased the asymmetry in processing time. Both findings were confirmed by a Bayesian GLME model mirroring the candidate model. The candidate VMA model and its Bayesian equivalent indicated that masking reduced the discriminability asymmetry. Moreover, a fully-saturated Bayesian model indicated that masking also increased the VMA in processing time. [CONCLUSION] Both approaches revealed backwards pattern-masking reduced polar angle asymmetries in discriminability but strengthened them in information accrual. These findings highlight the importance of considering both spatial and temporal performance field asymmetries.

Supported by NSF GRFP to NC and NIH R01 EY027401 to MC

TALK 6, 12:00 PM, 22.16

WHY CORTICAL IMPLANT PATIENTS MAY NEVER SEE THE STARS

Eirini Schoinas¹ (eirini@ucsb.edu), Leili Soo², Alfonso Rodil

Doblado², Geoffrey M. Boynton³, Ione Fine³, Eduardo Fernández²;

¹University of California, Santa Barbara, ²University Miguel Hernández, Elche, Spain, ³University of Washington

The majority of visual cortical prosthetic devices under development rely on small penetrating electrodes, motivated by the assumption that increasing electrode density and reducing electrode size will produce high-resolution vision. However, neurons in the brain are not analogous to pixels on a screen - stimulation of a small number of cells by a tiny depth electrode will not consistently produce discrete, punctate phosphenes. Here, we compared data from CORTIVIS patients (Fernández et al. 2021) implanted with small depth electrodes to predicted phosphenes generated using a "virtual patient" simulation based on a simple model of V1 (Fine and Boynton 2024). Although stimulation with the tiny electrodes of the CORTIVIS implant occasionally produced small punctate "stars", patients reported a variety of percepts, including larger spots and irregularly shaped forms. These phosphenes were stable within electrodes but varied substantially across stimulation sites. Our model successfully predicted the wide variety of percepts reported by patients and further suggests that, for a subset of electrodes, stimulation may have activated multiple cortical locations — possibly as a result of axonal activation. This virtual patient model can also be used to simulate patient perceptual experiences for future cortical implants (such as Neuralink's). Our simulations of 'implants of the future' show that high resolution cannot be obtained simply by having a very large number of very small electrodes – it is necessary to develop an encoding model to stimulate the right electrode or combination of electrodes. Unfortunately, no one has yet suggested a sensible way to measure the receptive field of each individual electrode in a blind individual. Models like ours are an example of how virtual prototyping can offer more realistic expectations about what kind of vision bionic eyes might provide.

PDC2022-133952-100 & PID2022-141606OB-I00 Ministerio de Ciencia, Innovación y Universidades; CIPROM/2023/25 Generalitat Valenciana: Programa Iberoamericano: CYTED- NT4SM; EU Horizon 2020 Research and Innovation Programme No. 899287: NeuraViPeR) [EF]; NIH R01EY014645 [IF & GMB].

TALK 7, 12:15 PM, 22.17

A GENERATIVE DIFFUSION MODEL REVEALS V2'S REPRESENTATION OF NATURAL IMAGES

Neel Agarwal¹, Gabriel Yancy¹, Zahra Kadkhodaie², Justin D. Lieber¹, J. Anthony Movshon¹, Eero P. Simoncelli^{1,2}; ¹Center for Neural Science, New York University, ²Center for Computational Neuroscience, Flatiron Institute

Neurons in visual cortex represent the complex features that occur in natural environments. Characterization of their responses has been limited by our inability to create stimuli that are both naturalistic and precisely controlled. Traditional stimuli – like gratings, textures, or noise – capture only specialized elements, while photographic images are too unconstrained for systematic study. Here, we leveraged a

generative diffusion model to create stimuli, and used high-density electrode arrays to measure responses of populations of neurons in anesthetized macaque V2. We trained a diffusion model, using a denoising objective, to capture a prior probability distribution over a set of photographs of scenes from which natural images can be sampled. In the vicinity of any given image, this distribution may be approximately described as a low-dimensional manifold. We generated images both within and off this manifold, with controlled pixel-level distances (MSE) relative to a set of base natural images. Specifically, we generated off-manifold images by adding Gaussian white noise to base images, while we generated on-manifold images by adding noise and then forcing the noisy images back onto the learned manifold. On-manifold images drove population responses that were more diverse than distance-matched off-manifold images, without greatly changing the overall mean response. By analyzing responses to images that varied in distance from the base image, we found that the cosine similarity of the population response relative to the base image declined more steeply on-manifold than off-manifold. This occurred because some responses increased while others decreased along on-manifold trajectories, forming heterogeneous encoding axes. Our results suggest that V2 neurons provide an orderly representation of natural image structure. Diffusion models thus provide a powerful new stimulus engine to explore population coding in visual cortex.

Simons Foundation, NIH EY022428

**TALK SESSION: SATURDAY, MAY 16, 2026,
10:45 AM – 12:30 PM, TALK ROOM 2**

Face and Body Perception: Mechanisms

Moderator: Philippe Schyns, University of Glasgow

TALK 1, 10:45 AM, 22.21

FROM BLUR TO IDENTITY: TWO-STAGE NEURAL DYNAMICS IN PRIMATE FACE PROCESSING SYSTEM

Wanru Li¹ (wannuli@stu.pku.edu.cn), Yipeng Li¹, MiYoung Kwon², Pinglei Bao¹; ¹Peking University, Beijing, China, ²Northeastern University, Boston, MA, USA

Human observers can recognize faces under substantial blur, although this robustness requires additional viewing time. While coarse-to-fine dynamics are well documented in early visual cortex, far less is known about whether the macaque inferotemporal (IT) cortex exhibits a comparable temporal structure. Recent work indicates that IT cortex contains explicit spatial-frequency representations, yet how these representations contribute to face identity recognition under blur remains unclear. Using Neuropixels recordings from the macaque middle and anterior lateral face patches (ML and AL), we examined how face-selective neurons respond to 120 identities across 11 levels of blur. We found that face-selective neurons are not homogeneous in

their tuning. More than half (~60%) preferred moderately blurred faces, whereas the remainder preferred clear faces. Both subpopulations showed similar blur-induced increases in response latency, even after controlling for image contrast, indicating that blur broadly slows IT computations rather than merely reflecting low-level luminance differences. Analyses using high-pass and low-pass images further confirmed that clear-preferring neurons respond preferentially to high-frequency detail, whereas blur-preferring neurons emphasize coarse structure, revealing that IT face patches contain distinct spatial-frequency channels. At the population level, representational similarity analysis (RSA) revealed a robust two-stage dynamic across ML and AL face patches. After controlling for pixelwise similarity and contrast energy, blur-level information emerged rapidly as the dominant representational axis. Identity representations arose only later and became progressively delayed with increasing blur, revealing a temporal separation between coarse image structure and fine identity content. These results challenge models positing that invariant object recognition requires discarding low-level spatial frequency information. Instead, IT face patches retain multiple spatial-frequency channels and exploit these signals to extract identity from degraded sensory inputs. Together, our findings highlight a coarse-to-fine transformation in primate IT and provide a mechanistic account of the behavioral delays observed when recognizing blurred faces.

TALK 2, 11:00 AM, 22.22

CORTICAL FACE SELECTIVITY EXTENDS FAR INTO THE VISUAL PERIPHERY

Corey M Ziemba¹, Alex Roseman¹, Kenji W Koyano^{2,3}, Carly Gregg², Harish Katti², David A Leopold², Hendrikje Nienborg¹; ¹Laboratory of Sensorimotor Research, NEI, Bethesda, MD, ²Systems Neurodevelopment Laboratory, NIMH, Bethesda, MD, ³Advanced Neuroimaging Center, National Institutes for Quantum Science and Technology, Chiba, Japan

Although peripheral vision is less detailed than central vision, we can identify important stimuli, like faces, when presented far from the center of gaze. However, facial recognition and its neural mechanisms have mostly been studied near the fovea. Here, we combined neuronal population recordings from macaque cortical face patches, an advanced stimulus projection system, and naturalistic stimulus synthesis to investigate the strength and feature selectivity of face preference across nearly the entire visual field. We implanted microwire brush arrays in three fMRI-targeted face patches in two monkeys: The anterior medial face patch in inferotemporal cortex, the perirhinal face patch, and the prefrontal orbital face patch. We presented stimuli up to an eccentricity of 90 degrees using a projector, spherical mirror, and hemispheric dome screen. Monkeys fixated a central point while we briefly flashed images centered at different visual field locations spanning 0-80 degrees eccentricity. Stimuli consisted of isolated face and nonface object images. We also created scrambled counterparts for each image using a synthesis procedure that disrupts complex object features while maintaining low- and mid-

level features. These scrambled stimuli are known to elicit similar neural responses in early visual cortex and match peripheral appearance compared to natural images under some circumstances. We found that the face preference of neural populations from all areas was present in the periphery, sometimes beyond 50 degrees eccentricity. However, selectivity was generally weaker compared with stimulus presentation at the fovea. Further, we found little evidence that face selectivity in peripheral vision can be accounted for by lower-level features preserved by our scrambling procedure. This result contrasts with perceptual studies showing that the appearance of peripheral stimuli can be matched between natural images and their scrambled counterparts and indicates that the cortex maintains highly selective responses to faces presented far in the visual periphery.

We acknowledge funding from the National Eye Institute Intramural Research Program at the National Institutes of Health (NIH, grant no. 1ZIAEY000570-01 to H.N.)

TALK 3, 11:15 AM, 22.23

VISUAL DISTORTIONS REVEAL DISSOCIABLE REFERENCE FRAMES UNDERLYING FACE AND OBJECT RECOGNITION IN THE VENTRAL VISUAL PATHWAY

Antônio Mello¹, Daniel Stehr¹, Sarah Kerns¹, Chandana Kodiweera¹, Krzysztof Bujarski¹, Brad Duchaine¹; ¹Dartmouth College

A reference frame (RF) is the coordinate system used to spatially represent a visual stimulus. Neurophysiological and psychophysical research has shown that the visual system relies on several RFs (e.g., retino-centered, body-centered), but evidence for object-centered RFs, where spatial information is encoded relative to the object itself rather than the viewer, in the ventral visual pathway (VVP) is limited. Here, we test a rare neuropsychological case across six experiments to identify RFs underlying visual recognition. Nagel is a 40-year-old man with hemi-prosopometamorphopsia who perceives distortions on the right half (RH) of faces and, less frequently, on the RH of objects. In Experiments 1-4, we presented 432 images of faces and hands while varying visual field location, visible region, viewpoint, and picture-plane orientation of the stimuli. Fisher's exact tests revealed that Nagel's distortions do not depend on where stimuli appear in his visual field (i.e., distortions are not retino-centered). However, the distortion location differed by category. For faces, distortions followed the RH of the face across orientations, affecting the same features even when faces were upside down (object-centered distortions). For hands, distortions instead remained on the RH of the stimulus image, meaning the distorted region shifted as hands rotated (stimulus-centered distortions). This dissociation could depend on whether a category has a canonical orientation: faces are almost always viewed upright, whereas hands are often seen in many orientations. To test this hypothesis, Experiment 5 presented 180 rotated images of a face, a hand, and 13 different objects with established canonical orientations, as confirmed by human raters. Distorted regions of nonface objects again shifted with stimulus rotation (stimulus-centered

distortions), ruling out the canonical account. Experiment 6 replicated this pattern using 48 body images. Together, these findings provide robust evidence for object-centered RFs for faces and stimulus-centered RFs for objects within the VVP.

We thank the Hitchcock Foundation for funding this research.

TALK 4, 11:30 AM, 22.24

A NOVEL EXPLANATION OF THE INVERTED FACE EFFECT

Garrison Cottrell¹ (gary@ucsd.edu), Kira Fleischer¹, Nikita Kachappilly¹, Alexander Tahan¹, Samuel Lee¹, Xavier Chen¹; ¹UCSD

Vision researchers often assert that when shown an inverted face, subjects “revert to feature processing.” But how can they still use feature processing on inverted features? Eyebrows, eyes, noses and mouths are mono-oriented in everyday life. Furthermore, in the Thatcher illusion, subjects do not notice the features are right-side up relative to the face. This is a mystery: If subjects revert to feature processing, how would they use inverted features in one case, and not notice that they are upright in another? We suggest that, because of the log-polar mapping from the visual field to V1, in *both* cases, the features are *not* inverted when they enter the cortex. In this representation, rotation is just a vertical shift – features remain in the same orientation when the face is upright or inverted. So why are inverted faces difficult to recognize or remember? In face processing, small changes to configuration make the face appear to be someone else. Because the cortex is flat and not a torus, it can't represent that 270 degrees is continuous with -90 degrees. When the face is upright, the nose is above the left eye; inverted, the nose is below the right eye, disrupting the configuration of the features. We test this hypothesis with a DCNN, trained using a log-polar version of faces. The model is disrupted by inverted faces, but it still recognizes 50% of familiar faces. A standard DCNN is nearly at chance, unlike humans. This effect is much smaller for inverted objects, trained to be recognized at the basic level, where configuration doesn't matter. When the model is trained to be a dog expert, it again is disrupted by inversion (Diamond & Carey, 1986). Hence, we have a novel explanation of the inverted face effect, based on the transformation that occurs in V1.

This work was supported by NSF CRCNS grant #2208362

TALK 5, 11:45 AM, 22.25

FACE-TRAINED DEEP NEURAL NETWORK IS SEVERELY MISALIGNED WITH HUMAN PERCEPTUAL JUDGMENTS OF FACE SHAPE AND TEXTURE.

Virginia E Strehle¹ (virginia.e.strehle@vanderbilt.edu), Frank Tong¹; ¹Vanderbilt University

Recent studies have suggested human perception of facial similarity can be accurately predicted by face-trained deep neural network (DNN) models (e.g., Jozwik et al., 2022; Dobs et al., 2023). However,

it remains unclear whether human observers and DNN models are responding to the same dimensions of facial appearance. We explored this question by leveraging a 3D morphable model of face appearance (Paysan et al., 2009), which represented the shapes and textures of real human faces in separate principal component spaces. This model allowed us to independently and explicitly alter face shape and texture. In a forced-choice task, participants (n=26) viewed a target face and two modified faces that differed along a specific shape or texture principal component. Participants then chose the face that appeared more different. We conducted an analogous comparison on a ResNet-50 model trained on VGGFace2 by computing cosine distances between response patterns in the penultimate layer to those faces. Whereas humans perceived shape alterations as more salient, the model treated texture alterations as more distinct. Next, in a method-of-adjustment task, we generated pairs of faces that differed along multiple dimensions of shape and/or texture. Participants (n=23) traversed these dimensions by adjusting a probe face until they perceived a just-noticeable difference in identity from a reference face. We then compared human adjustment thresholds to the cosine distances between penultimate layer response patterns generated from the median human threshold faces and the associated reference faces. Both the human identity thresholds and the DNN response patterns were more sensitive to face shape than texture in this full dimensional space. However, human-to-human similarity far exceeded that of DNN-to-human similarity, especially in shape trials (human: $r = 0.60$; DNN: $r = -0.31$). Overall, our findings highlight major differences in how DNNs and humans process faces, particularly regarding face shape.

This research was supported by NEI grants R01EY035157 to FT and P30EY008126 to the Vanderbilt Vision Research Center.

TALK 6, 12:00 PM, 22.26

DEEP NEURAL NETWORKS PREDICT HUMAN SOCIAL TRAIT JUDGEMENTS FROM DYNAMIC FACES

Po-Yuan Alan Hsiao¹ (hsiao.alan36@gmail.com), Suvel Muttreja², Diego Rodriguez¹, Matteo Visconti di Oleggio Castello³, James V. Haxby⁴, Maria Ida Gobbi⁵, Guo Jiahui¹; ¹The University of Texas at Dallas, ²University of Southern California, ³University of California, Berkeley, ⁴Dartmouth College, ⁵University of Bologna

Humans constantly infer social traits from faces, and these subjective judgments play an important role in guiding social interactions. Dynamic faces offer information beyond physical facial features, providing contextual cues to aid judgements and creating ecologically valid experiences. However, most previous research has only focused on modeling social judgments from static faces. Thus, it is still unclear whether deep neural networks (DNNs) can also capture the information humans use to make social judgments from dynamic faces. To answer this question, we collected social judgments from a dynamic stimulus set consisting of 707 naturalistic four-second video clips of unfamiliar individuals. 171 subjects participated in the study

and rated faces on a continuous 1–100 scale. We also include a static stimulus set from the Profile Image Dataset (1224 images, 160 raters, 1–9 scale). We used three face-trained DNNs (InsightFace, Alex-Face, VGG-Face) and two object-trained DNNs (AlexNet, VGG-16) to extract features from dynamic videos and static photos. Features were reduced using PCA, and prediction performance was evaluated by correlating model-predicted and human-rated traits. We found that face-trained DNNs successfully predicted human social judgments across all three traits from dynamic faces, consistently outperforming object-trained DNNs. Variance partitioning analysis showed that face-trained networks explained substantial unique variance in human social judgments, with only limited unique contributions from object-trained networks. Furthermore, models trained on one trait showed little generalization to other traits, indicating that the model captured trait-specific information. Most interestingly, generalization between models trained on dynamic and static faces was limited, suggesting that the shared representations between humans and DNNs differed for dynamic and static faces. These findings demonstrate that human social judgments from dynamic faces can be modeled successfully using DNNs trained with faces. The shared information between humans and DNNs is face-specific, trait-unique, and different from judgments based on static images.

This work is supported by startup funds provided by the School of Behavioral and Brain Sciences at the University of Texas at Dallas.

TALK 7, 12:15 PM, 22.27

DYNAMIC TASK-DEPENDENT FEATURE COMPUTATIONS IN BRAIN PATHWAYS FOR FACE CATEGORIZATION

Yuening Yan¹, Jiayu Zhan², Yaocong Duan¹, Oliver G. B. Garrod¹, Robin A. A. Ince¹, Chen Zhou¹, Rachael E. Jack¹, Philippe G. Schyns¹; ¹University of Glasgow, ²Peking University

Understanding how the brain extracts meaning from complex, dynamic visual input is a central goal of vision sciences. Faces are among the most informative stimuli we encounter, conveying stable identity, transient emotional state and a wide range of social information. Yet the computations that enable the brain to route, represent, and integrate these different sources of facial information remain only partially understood. We combined generative 4D face modelling, participant-specific millisecond-resolved MEG, data-driven information-theoretic analyses and explicit modelling to uncover how the brain gates, routes, and integrates facial features according to task demands—whether participants categorized identity, emotion, or both. This multimethod approach allowed us to ask three core questions: (i) How does the brain gate task-relevant features in dynamic faces? (ii) Through which cortical pathways are facial identity and emotion features routed? (iii) Where and when do these features converge to form an integrated semantic code (e.g. “happy Mary”)? Participants viewed the same 3,600 facial animations (dense-sampled design) and categorized them by identity, emotion or both (N = 8 per condition; all analyses performed per participant; $p < 0.05$, FWER-corrected). We

found that task demands selectively gate and route facial information. During identity judgements, task-relevant shape features propagated through the ventral occipito-temporal pathway; during emotion judgements, dynamic movement features propagated through the lateral-social pathway. When these same features were task-irrelevant (e.g. movements during identity judgements), they were nonetheless represented but rapidly (< 140 ms) pruned in occipital cortex. The two pathways converged in temporal cortex, where MEG activity encoded synergistic interactions between identity and emotion features—an integrated, person-specific representation that emerged only for semantically known faces. Our findings reveal a dynamic computational mechanism through which the brain flexibly gates, routes and integrates perceptual features to construct social meaning, providing general principles for meaning formation in perceptual systems.

This work was funded by the Wellcome Trust (Senior Investigator Award, UK; 107802) and the Multidisciplinary University Research Initiative/Engineering and Physical Sciences Research Council (USA, UK; 172046-01) (P.G.S.); ERC [FACESYNTAX; 759796] (R.E.J.); the Wellcome Trust [214120/Z/18/Z] (R.A.A.).

TALK SESSION: SATURDAY, MAY 16, 2026, 2:30 – 4:15 PM, TALK ROOM 1

Functional Organization of Visual Pathways 1

Moderator: Martin Hebart, Justus Liebig University Giessen

TALK 1, 2:30 PM, 24.11

REORGANIZATION ACROSS CORTICAL HIERARCHIES AND MODALITY-SPECIFIC THALAMIC PLASTICITY IN CONGENITAL SENSORY LOSS.

Monami Nishio¹ (monami@sas.upenn.edu), Xingyu Liu¹, Maria Zimmermann², Marcin Szwid³, Allyson Mackey¹, Michael Arcaro¹; ¹University of Pennsylvania, ²Johns Hopkins University, ³Jagiellonian University

Loss of sensory input leads to reorganization of cortical sensory systems. Here, we asked whether plasticity varies systematically with cortical distance from primary sensory areas, extends to higher-order thalamus, and follows similar principles across sensory modalities. Using structural and functional MRI, we characterized cortical and thalamic reorganization in adults with congenital blindness (N=15; controls N=19) or deafness (N=25; controls N=30). In blind individuals, lateral geniculate nucleus (LGN) volume and V1 surface area were significantly reduced and correlated across participants, suggesting coordinated thalamocortical structural reorganization. Beyond V1, changes in extrastriate surface area and morphology diminished with geodesic distance from V1. Visual cortex in blind individuals showed

stronger resting-state correlations with the frontoparietal control network and weaker correlations with somatosensory cortex. This pattern was more pronounced in ventral than dorsal areas. Visual-control coupling was further enhanced during a decision-making task, suggesting recruitment of visual cortex for cognitive processing. Complementing these cortical changes, higher-order thalamic nuclei (dorsal pulvinar, central medial, mediodorsal) showed increased functional correlations with V1, indicating reorganization of thalamocortical circuits. Deaf individuals exhibited hierarchical plasticity in auditory cortex, with structural changes that diminished with distance from primary auditory cortex and increased cross-modal connectivity. However, unlike in blindness where deprived cortex coupled preferentially with frontal control regions, auditory cortex in deafness showed stronger coupling with visual cortex. Further, thalamic structures in deaf individuals showed no significant structural or functional alterations, indicating modality-specificity of thalamic plasticity from sensory loss. These findings reveal hierarchical principles governing cortical reorganization following early sensory loss and the modality-specific contribution of thalamic plasticity.

TALK 2, 2:45 PM, 24.12

CONTRIBUTIONS OF THE PRIMATE SUPERIOR COLLICULUS TO PERCEPTUAL DOMINANCE DURING FLASH SUPPRESSION

Divya Subramanian¹, Richard Krauzlis¹; ¹National Eye Institute, National Institutes of Health

We only consciously perceive a subset of our sensory inputs. How is this subset selected? We hypothesized that the primate superior colliculus (SC), important for selecting visual inputs to guide orienting, may play a role. To test this, we used binocular flash suppression to dissociate sensory inputs from the conscious percept. When an image presented to one eye is followed by an image added to the second eye, the second image becomes perceptually dominant even though both images are physically present. We asked whether visual activity in the macaque SC reflects the dominant image, the suppressed image, or the physically present mixed image. We presented oriented textures at two spatial frequencies that SC neurons distinguish at short latencies. We compared activity ($n = 106$ neurons from 2 monkeys) in the dichoptic, flash suppression condition to control conditions in which each image was unambiguously presented. We found that activity in the flash suppression condition matched the perceptually dominant image, regardless of whether the low or high spatial frequency image was dominant. However, SC activity only transiently reflected the dominant image. We trained linear classifiers to distinguish the two images based on population activity in the unambiguous control conditions and tested them either in the flash suppression condition or on held-out control trials. In the flash suppression condition, the classifier reliably labeled activity as belonging to the dominant image but only at short latencies (40 - 130ms from second image onset). For the held-out control trials, however, performance remained well above chance for the duration of the stimulus presentation (700 - 1000ms).

Thus, SC activity transiently reflects the dominant image, suggesting that it reflects the perceptual switch rather than the sustained visual percept.

TALK 3, 3:00 PM, 24.13

PRECISE FUNCTIONAL LOCALIZATION OF THE FOVEOLAR REPRESENTATIONS IN ANESTHETIZED MACAQUE VISUAL CORTEX

Meizhen Qian^{1,2}, Meixuan Chen², Coco Laska^{1,3}, Anna Wang Roe^{1,2}; ¹Translational Neuroscience Laboratory Division, Center for Biomedical Imaging and Neuromodulation (C-BIN), Nathan S. Kline Institute for Psychiatric Research, ²Interdisciplinary Institute of Neuroscience and Technology, Zhejiang University School of Medicine, Hangzhou, China, ³Department of Medicine, Health, and Society, Vanderbilt University

Background: The primate foveola, which represents the central $\sim 1^\circ$ of the visual field, is critical for high-acuity vision and gaze-directed attention. In a cortical region previously called the foveal confluence, Qian (2025 Nature Neurosci) using ultrahigh field 7T MRI in awake fixating macaques discovered multiple foveolar loci encircling a large area termed the “foveolar core”, suggesting novel hierarchy for central vision. To study this foveolar area in anesthetized monkeys to determine exactly where the foveola is directed on a monitor, we introduce a systematic approach for foveolar localization. Methods: Functional EPI data were acquired at 0.8 mm³ isotropic resolution in a 7T MRI. Data acquisition was monocular. We used a progressive method of imaging with large to narrow stimuli to triangulate the center of gaze. Initially, wide (4.5°) vertical and horizontal bars containing drifting gratings were presented, at each of 3 different positions; the bar that induced the strongest response was chosen as the most central. Successively narrower bars (1.5° wide, then 0.5° wide) were presented at 3 positions within the central-most location. 0.2° spots were presented at the estimated foveal coordinate to further refine the central-most location. Results: Retinotopic shifts in activation corresponded well with stimulus position. The most central vertical bar stimulus activated the vertical meridian at the V1/V2 border; off-center bars produced parallel activations on either side of the border. When stimuli crossed into the ipsilateral hemifield, activation shifted to the other hemisphere as predicted. For horizontal meridian, the central-most stimulus produced the largest BOLD responses and activated dorsal and ventral V2/V3 borders. Foveolar spot stimuli exactly at the center produced bilateral foveolar activations. Significance: These results confirm the feasibility of precise foveolar localization in anesthetized macaques and provide a foundation for further studies of foveolar cortical organization using fMRI coupled with other recording and stimulation methodologies.

TALK 4, 3:15 PM, 24.14

RETINOTOPIC ORGANIZATION OF INTERHEMISPHERIC STRUCTURAL CONNECTIVITY IN EARLY VISUAL CORTEX FROM INFANCY TO ADULTHOOD

Seda Karakose-Akbiyik¹ (sakbiyik@stanford.edu), Emily Kubota¹, Xiaojian Yan², Sarah Tung¹, Christina Grace Tyagi¹, Danya Ortiz¹, Vaidehi Natu¹, Kalanit Grill-Spector¹; ¹Stanford University, ²Fudan University

Early visual cortex (EVC: V1, V2, V3) exhibits a retinotopic organization defined by polar angle and eccentricity, with each hemisphere representing the contralateral hemifield. Previous work suggests that the corpus callosum tracts linking the EVC in the two hemispheres are also retinotopically organized, with the densest connectivity for vertical meridian representations. Yet, in vivo evidence for these principles in humans remains limited, partly because early tractography methods lacked the precision needed to map these fine-grained connections. Recent advances in diffusion MRI (dMRI) now allow a more precise characterization of these connections. Here, we used dMRI to test whether the interhemispheric structural connectivity of V1–V3 is organized by retinotopy (i.e., polar angle and eccentricity). By examining both human adults (N=26) and infants at 0, 3, and 6 months (N=23, 23, and 21), we also tested how this organization emerges during infancy. We parcellated V1–V3 into polar-angle segments (V1-horizontal; V1–V2 lower/upper vertical meridians; V2–V3 dorsal/ventral horizontal meridians) and eccentricity bands (0–3°, 3–6°, 6–12°, 12–24°, 24–48°) using the Benson atlas, and identified the projection zones of each segment by mapping where streamlines from one hemisphere terminated in the other. In adults, contralateral projections of V1–V3 followed a clear retinotopic organization, consistent with the animal literature. Contralateral connections were denser along the vertical than the horizontal meridian, and lower/upper vertical-meridian segments primarily terminated in homotopic zones in the opposite hemisphere. Eccentricity also mattered: connections were denser between homotopic or adjacent eccentricities and sparser between distant ones. An adult-like organization was evident in infants, with both polar angle and eccentricity-based patterns already present for 0-month-olds. Together, these findings suggest that human V1–V3 callosal projections follow a retinotopic organization from birth. This organization likely plays a key role in coordinating signals across the two hemispheres to form a unified representation of visual space.

NIHRO1EY033835, NIHRO1EY023915

TALK 5, 3:30 PM, 24.15

CANONICAL SPACES UNCOVER SHARED AND UNIQUE REPRESENTATIONAL GEOMETRY OF PRIMATE IT

Sander van Bree^{1,2,3}, Martin N. Hebart^{1,2,3}; ¹Justus Liebig University Giessen, ²Max Planck Institute for Human Cognitive and Brain Sciences, ³Center for Mind, Brain and Behavior

The inferotemporal cortex (IT) is central to high-level vision in primates, yet the extent to which its functional organization is conserved across species remains debated. In this study, we approach this question by systematically comparing macaque multi-unit activity and human fMRI responses to a large set of naturalistic images. Specifically, we used canonical correlation analysis (CCA) to construct three neural spaces, each of which provides complementary views on primate IT: macaque and human spaces obtained by aligning responses within species, and a universal cross-species space derived by combining all subjects. We demonstrate how these models can be used to uncover neural representations that are selective to species or generalize between primates, even with data obtained from different neuroimaging modalities. Analyzing these canonical spaces yields the following primary results. First, primate IT is high-dimensional and organized around shared axes, including but not limited to animacy and spiky-stubbiness. A non-negative, parts-based factorization of this universal space further uncovered a host of dimensions, ranging from grid-like patterns to biologically salient clusters related to faces and body silhouettes. Second, contrasting the species-specific spaces revealed a striking animacy asymmetry: the macaque space was driven largely by living categories, whereas the human space was better explained by non-living categories and human body parts. Furthermore, relative to the human space, the macaque space was more strongly captured by early visual representations, as evidenced by neural network activations and other predictors. Finally, the universal primate space held a shared representational structure for symbols, icons, and object ensembles, consistent with the hypothesis that human orthographic specialization builds on neural precursors already present in non-human primate IT. Together, this showcases how neural spaces derived from alignments within and across species provide a principled way to chart shared and species-specific representational geometry.

Supported by ERC StG COREDIM (101039712), LOEWE Start Professorship, and DFG: SFB/TRR 135 "Cardinal Mechanisms of Perception" (222641018- C11) & EXC 3066/1 "The Adaptive Mind" (533717223).

TALK 6, 3:45 PM, 24.16

CUT THE TREE: IMAGE-GROUNDED ENCODING MODELS REVEAL DISTINCT TEMPORAL PROFILES OF NATURALISTIC OBJECT AND SCENE PROCESSING

Niklas Mueller¹ (n.muller@uva.nl), H. Steven Scholte¹, Iris Groen¹; ¹University of Amsterdam

It is well established that the brain achieves efficient visual information processing using specialized regions, such as scene-, face-, and object-specific regions which have been localized in humans using functional magnetic resonance imaging (fMRI). The temporal dynamics of the processing of different types of visual information have been understood less well. Some experiments use carefully designed stimulus presentations, such as flicker-stimuli, to study temporal

information processing. However, these stimuli lack ecological validity to link temporal processing stages of visual perception to naturalistic, real-world visual information. Here, we use electroencephalography (EEG) recordings of human participants viewing large-field natural images and build brain-predictive image-grounded encoding models based on deep neural networks (DNNs) to study the temporal dynamics of object and scene processing. We separate object from scene information using manually created object bounding boxes and build separate encoding models for individual scene parts, such as different object classes (e.g., cars, tree, traffic signs) or scene elements (building, floor, sky). Our results show that object information is encoded delayed in time compared to early and rapid encoding of scene information. Moreover, with our approach, we can identify specific temporal encoding profiles for individual object and scene elements within a given naturalistic image. For example, we find that visual floor elements are maximally encoded around 150 ms, consistent with prior work showing the subsequent emergence of action affordance representations. Further, our encoding profiles show that while trees are commonly labeled as objects, they are processed more similarly to other scene elements than to objects. Overall, we establish a direct link between the pixel-wise information of ecologically valid stimuli and the temporal neural encoding profiles in EEG recordings. This opens new possibilities to study the neural processing of individual scene elements under more natural, real-world conditions.

TALK 7, 4:00 PM, 24.17

SPATIOTEMPORAL CONTINUITY SHAPES INFORMATION FLOW IN THE PRIMATE VENTRAL STREAM

Alec G. Sheffield¹, Yuxuan Xue¹, Monika P. Jädi¹, Anirvan S. Nandy¹;
¹Yale University

Recurrent interactions between cortical areas are thought to be critical for coherent perception, yet how these long-range pathways depend on the spatiotemporal structure of natural vision remains unclear. We tested whether disrupting spatiotemporal continuity selectively impairs inter-areal communication in the ventral visual stream. We recorded simultaneously across cortical layers in areas V2 and V4 of awake marmosets (n=2) using laminar electrode arrays during free viewing. Animals watched intact movies or temporally shuffled versions in which consecutive 120-ms segments were randomly reordered, preserving local visual statistics while disrupting long-timescale spatiotemporal relationships. Spike-triggered current source density confirmed functional connectivity between recording sites. We quantified inter-areal communication using spike-field phase consistency, probabilistic graphical models (Multi-Timelag weighted Dynamic Bayesian Networks), and reduced-rank regression to characterize communication subspace geometry. Temporal shuffling exerted minimal effects on oculomotor behavior, firing rates, and population dimensionality, ruling out gross state differences as explanations for neural effects. However, inter-areal coordination was profoundly disrupted. Spike-field coupling decreased across theta

frequencies in feedforward (V2→V4) and alpha/low-beta frequencies in feedback (V4→V2) directions. Critically, directed spiking dependencies revealed an asymmetric effect: feedback connections from deep-layer V4 to superficial V2 were selectively weakened, while feedforward dependencies remained intact. Subspace analyses showed that temporal discontinuity both rotated and destabilized the low-dimensional communication channels linking V2 and V4, reducing predictive accuracy and increasing geometric variability. In contrast, within-area laminar communication was largely preserved across conditions. These findings demonstrate a clear dissociation between local and long-range processing in the ventral stream. While local circuits can represent short-timescale visual information relatively independently, coherent coordination across hierarchical levels, particularly through feedback pathways, depends critically on the continuous spatiotemporal structure characteristic of natural vision.

Supported by NIH R01 EY034605, R00 EY025026, R21 MH126072, SFARI 875855 (MPJ); NARSAD YI, Ziegler Foundation, Yale Orthwein Funds; NIH R01 EY032555, R21 MH126072, SFARI 875855 (ASN); NSF GRFP (AGS); and NEI P30 EY026878 (Yale).

TALK SESSION: SATURDAY, MAY 16, 2026, 2:30 – 4:15 PM, TALK ROOM 2

Temporal Processing

Moderator: Nihan Alp, Sabanci University

TALK 1, 2:30 PM, 24.21

LATE, NON-OVERLAPPING MASKS DISRUPT FAST TARGET PROCESSING

Martina Morea¹ (martina.morea@epfl.ch), Roberta Cessa², Michael H Herzog¹, Marco Bertamini^{2,3}; ¹Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland., ²Department of General Psychology, University of Padova, Padova, Italy, ³Department of Psychology, University of Liverpool, Liverpool, UK.

Visual perception arises at remarkably fast speed: humans can react to visual stimuli within ~190 ms, or even distinguish animals from cars in around 300 ms. Perception is thought to arise from an initial feedforward sweep followed by reentrant processing that stabilizes and refines the percept. Backward masking has long served as a tool to disrupt this recurrent processing, typically using masks that spatially overlap with the target (pattern masking) or surround it (meta-contrast masking). Recent findings show that even foveally presented masks—far from the peripheral target—can impair perception, suggesting the existence of “foveal feedback,” whereby peripheral representations project to foveal cortex to exploit its higher resolution. In this study, we examined whether similar effects occur for a low-level stimulus and task such as peripheral vernier offset discrimination. Participants performed a vernier discrimination task in which the target was

followed by masks presented at various spatial locations and stimulus onset asynchronies (SOAs). We performed five experiments (with a total of 108 participants), in which the mask was either a dynamic noise patch or two vertical lines. For all mask types, locations, and SOAs (up to 250 ms), performance was impaired relative to no-mask baseline. Thus, perception is far less stable than it might appear, even for simple low-level tasks such as vernier acuity. Notably, the strongest masking occurred when parafoveal masks were presented in the target's direction roughly 100 ms after its onset. These findings expand traditional models that confine masking to strictly spatial, retinotopic interactions. We propose that perception depends on a dynamically modulated spatiotemporal buffer influenced by feedback, attention, and oculomotor mechanisms.

TALK 2, 2:45 PM, 24.22

ACTION PREPARATION MODULATES TEMPORAL CROWDING

Lilas Haddad¹ (lilas_haddad@brown.edu), Tri Nguyen¹, Yaffa Yeshurun^{2,3}, Joo-Hyun Song¹; ¹Brown University, ²School of Psychological Sciences, University of Haifa, ³Institute of Information Processing and Decision Making, University of Haifa

Temporal crowding is the impairment of object recognition when distractors appear before and after a target, and it usually reduces target encoding precision. Recent work suggests that perception and action interact in a bi-directional way, and that preparing an action can change how visual information is prioritized. Because prior studies of temporal crowding relied mainly on indirect mouse responses, it remains unclear whether preparing an action toward the target alters the impact of temporal crowding. We tested whether preparing an action, either relevant (Exp.1) or irrelevant (Exp.2) to the target orientation, changes temporal crowding. Participants judged the orientation (0 to 360°) of a target bar shown alone or with two temporally adjacent distractors, at short (200ms) or long (400ms) intervals. In Exp. 1 (N = 23), participants prepared an orientation relevant action by grasping the circle in the target orientation. In Exp.2 (N = 23), they prepared an orientation irrelevant action by pointing to the tip of the bar. Grasping requires specifying the orientation of the object to form a stable hand posture, whereas pointing requires locating the end point of the bar and does not depend on orientation. In both experiments, keypress responses in non-action trials served as baseline. Across both experiments, action preparation reduced guessing for crowded targets, suggesting increased access to target information. However, when the action required encoding orientation (grasping), temporal crowding decreased precision in orientation reports compared with the non-action baseline. This pattern was not observed when the action was orientation irrelevant (pointing). Results show that action preparation changes perception under temporal crowding. Preparing an action increases the availability of the target signal, but when action involves the crowded target feature, it increases susceptibility to distractor noise. Thus, action influences

temporal crowding in different ways depending on its relevance to the visual feature being judged.

TALK 3, 3:00 PM, 24.23

PREDICTION IN NATURALISTIC MOVIES INVERTS THE TEMPORAL ORDER OF THE VISUAL PROCESSING HIERARCHY

Tiziano Causin¹ (tiziano.causin@unitn.it), Ingmar de Vries^{1,2}, Christoph Huber-Huber¹, Eva Berlot², Floris de Lange², Moritz Wurm¹; ¹CIMeC, University of Trento, ²Donders Institute for Brain, Cognition, and Behaviour, Radboud University

In our ever-changing visual environment, the brain is continuously trying to predict what comes next at different temporal scales and levels of feature complexity. Up to now, visual perception has been mostly studied in static and artificial setups, thus hindering our ability to understand how representations unfold over time in the real world. While classical paradigms reveal neural activation to progress from low to high-level regions, we hypothesize that this order is flipped in realistic dynamic contexts because higher-level brain areas recurrently anticipate early processing stages. During MEG, 59 participants watched a movie ("1917"), which is characterized by a continuous stream of rich visual information without scene cuts. For each time point of the movie, we constructed gaze-dependent models from low-level (e.g. pixelwise, optical flow) to higher-level visual features (body posture, semantic saliency). We applied a dynamic extension to representational similarity analysis (dRSA) to measure the temporal alignment between timepoints in the movie and neural representations along the visual processing hierarchy. In occipito-parietal MEG sensors, we find that pixelwise luminance and optical flow (OF) magnitude are represented most strongly around ~40ms after stimulus onset, while OF direction was predicted ~-280ms in advance. In occipital sensors, body posture was represented after ~80ms from stimulus onset. Notably, this latency decreased markedly in temporal sensors (~0ms) and became predictive in frontal sensors (~-20ms), indicating that high-level areas represent body-posture earlier than occipital areas represent low-level visual information. Taken together, both regionwise and modelwise comparisons support the notion that an inverted hierarchy of predictive latencies governs the temporal order with which different representations appear over time. More generally, dRSA on naturalistic continuous stimuli paves the way for novel insights into the processing dynamics in the visual hierarchy.

TALK 4, 3:15 PM, 24.24

GRADUAL EMERGENCE OF TEMPORAL-ORDER JUDGMENTS IN LATE-SIGHTED CHILDREN

Marin Vogelsang¹ (ozaki@mit.edu), Lukas Vogelsang¹, Priti Gupta², Naviya Lall², Manvi Jain², Chetan Ralekar^{1,3}, Suma Ganesh⁴, Pawan Sinha¹; ¹Massachusetts Institute of Technology, ²Project Prakash, Dr Shroff's Charity Eye Hospital, ³IIT Roorkee, ⁴Pediatric Ophthalmology, Dr Shroff's Charity Eye Hospital

Determining the simultaneity or sequencing of visual events critically impacts perceptual inference. Cues signaling such temporal structure have been shown to significantly aid object discovery. This analysis constitutes a key developmental building block for transforming an infant's 'blooming buzzing confusion' into an organized sensorium. However, several factors underlying its genesis remain unclear. In particular, is early experience with spatially distinct visual entities necessary to learn their temporal arrangement, given that temporal relationships between different entities may be meaningful only if they are spatially resolvable? To address this question, we studied individuals with prolonged visual deprivation due to dense congenital cataracts who received sight surgeries late in childhood. We longitudinally tracked 13 patients from immediately before surgery through the first post-operative month and also examined 15 late-sighted individuals several years after surgery. Our control group comprised 22 normally-sighted individuals. In each trial, participants observed two transiently flashed visual stimuli with temporal lags between 17 and 500 ms and indicated which stimulus appeared first. We found that patients with several years of post-operative visual experience performed fully on par with controls. However, this ability was not evident immediately after surgery but followed a protracted progression, mirroring normal development. To examine the potential functional consequence of such gradual progression, we conducted simulations with video-based deep networks. Our results reveal that incorporating temporal degradations into early training phases yielded more temporally-extended model representations and improved downstream generalization, thereby conferring important advantages for robust visual development. Taken together, these findings reveal neural plasticity late into childhood for acquiring temporal-order judgments, in contrast to several other visual functions. They further suggest that time-based binding mechanisms may facilitate visual learning in the late-sighted, and that gradual temporal maturation may confer adaptive advantages, with implications for typical, atypical, and computational visual development.

NIH R01EY020517

TALK 5, 3:30 PM, 24.25

DISSOCIATING BEHAVIORAL AND NEURAL CORRELATES OF PERCEPT DURATION ACROSS THE VISUAL FIELD

Julia Papiernik-Kłodzińska^{1,2,3} (j.papiernik@doctoral.uj.edu.pl), Marek Binder¹, Renate Rutiku¹; ¹C-lab, Institute of Psychology, Jagiellonian University, ²Doctoral School in the Social Sciences, Jagiellonian University, Krakow, Poland, ³Centre for Brain Research, Jagiellonian University, Krakow, Poland

Despite the impression of uniform visual perception, it is evident that visual perception varies systematically across the visual field. It is better along the horizontal than the vertical meridian (horizontal-vertical asymmetry, HVA) and in the lower than the upper part of the vertical meridian (vertical meridian anisotropy, VMA). These

phenomena have been reported for various perceptual tasks, yet their impact on the temporal processing of visual percepts remains unclear. Therefore, we sought to determine whether the time needed to create new percepts is influenced by the visual field inhomogeneities. A custom procedure using figure-ground modulation and steady-state visual evoked potential analysis (SSVEP) was implemented to assess disparities between objective and subjective temporal characteristics of stimuli. During trials, 6 Hz flickering pop-out figures emerged in different areas of the visual field from a randomly changing background consisting of line segments flickering and rotating at 6 and 15 Hz. Participants counted how many times each figure flickered, with flickers objectively consisting of either 5, 6, or 7 cycles. The study involved 25 healthy participants and used a 64-electrode Biosemi EEG system. The results show that participants consistently underestimated the number of flickering stimuli, averaging 4.56 counts per trial. Spatial location significantly affected behavioral responses, revealing HVA without evidence of VMA. Neural measures (ERPs, SSVEPs) did not mirror these spatial effects. Significant ERPs clusters varied across objective stimulus durations, regardless of the low accuracy in their subjective perception (0.16). The SSVEPs similarly remained unaffected by location yet showcased an increase of 6 Hz activity in the figure versus baseline time window. These findings highlight a disconnect between subjective perceptual asymmetries observed in behavior and the lack of similarly robust neural markers in EEG, which invites further refinement of the methodology used to study the spatial aspects of subjective visual processing.

This work was supported by the National Science Centre in Poland (Grant nr. 2021/42/E/HS6/00425).

TALK 6, 3:45 PM, 24.26

DECOUPLING OF ILLUSORY TIME PERCEPTION AND AWARENESS: ELECTROPHYSIOLOGICAL EVIDENCE FOR ILLUSION-BLIND METACOGNITION

Tutku Öztel¹ (toztel17@ku.edu.tr), Martin Wiener¹; ¹George Mason University

Human time perception is subjective. This subjectivity renders timed intervals prone to illusory percepts induced by stimulus psychophysical properties. One robust example for stimulus-induced temporal illusions is the leftward shift in psychometric functions associated with p(long) responses for timed durations as a function of stimulus velocity (Karsilar, Kisa & Balci, 2019). Recently, we have shown that subjects are unaware when illusory distortions in time occur, as shown by a lack of metacognitive insight (Öztel & Balci, 2020). What we still do not know about this metacognitive inability is how (1) stimulus-induced temporal illusions are represented and (2) if at all, metacognitive processes associated with these illusions are encoded at neural level. To address these questions, we recorded 64-channel electroencephalogram (EEG) activity in human adults (n = 26) in a temporal bisection task where participants classified the duration (1-3.5s) of an animated walking stickman figure with different velocity

(i.e., 25, 50, 100 fps*) as being short or long and reported their confidence. Behaviorally we replicated our previous findings wherein faster velocities led to longer reported intervals, as well as lower precision yet faster reaction time; in contrast, confidence judgments did not vary with psychophysical shifts, only tracking reaction time. To our surprise, common EEG signatures of time perception, such as the frontocentral contingent negative variation (CNV) or late positive component of timing (LPCT) (Wiener & Thomspon, 2015; Ofir & Landau, 2022) did not vary with walking speed. Instead, onset-locked responses over occipital electrodes covaried with walking speed and correlated with subject-level shifts in psychometric functions. We further observed that these effects were driven by entrainment to walking speed. Altogether, our findings suggest that temporal illusions are engaged by bottom-up sensory processes, but fail to reach the metacognitive level; in contrast, metacognitive insight for time appears largely related to top-down sampling of motor output.

TALK 7, 4:00 PM, 24.27

IT DURATION CODING REFLECTS STIMULUS HISTORY, AND VIDEO ANNS QUALITATIVELY APPROXIMATE THIS COMPUTATION

Dominique Chuaqui¹, Matteo Dunnhofer^{1,2}, Kohitij Kar¹; ¹York University, ²University of Udine

The ventral stream is frequently modeled as a static object-recognition system, yet accumulating evidence points to meaningful history-dependent computation. Recent evidence shows that the inferior temporal (IT) cortex, long considered a shape-selective endpoint of the ventral stream, responds differently to static images based on preceding stimulus history. Dynamic tasks provide a powerful tool for studying these temporal influences. Therefore, here, we used duration estimation as a tractable test case to ask: Does IT compute how long an object persists by integrating temporal history, or would a purely frame-based system, such as a feedforward ANN, yield the same outcome? We further asked whether modern video ANNs provide a better mechanistic hypothesis for IT's dynamics. To diagnose history-dependence, we manipulated temporal coherence without altering the underlying images. Objects were shown either in their natural temporal sequence (coherent) or as temporally scrambled versions containing the exact same frames (incoherent). If IT duration estimates rely only on instantaneous object-bearing frames, coherence should not matter. If IT integrates stimulus history, coherent and incoherent videos should yield different duration predictions. We recorded activity from 119 reliable IT neurons (split-half > 0.5) while macaques viewed 200 videos. Linear models trained on IT activity showed significantly higher duration-prediction accuracy for coherent than incoherent stimuli ($r=0.68$ vs. 0.60 , $p<0.001$), demonstrating robust history-dependence in IT dynamics. We then evaluated two ANN hypothesis classes. Feedforward image models showed no coherence effect ($\Delta=0.0067$, $p=0.334$) and performed below IT for both stimulus types. In contrast, video ANNs, which explicitly integrate temporal context, exhibited a significant coherence advantage ($\Delta = 0.0677$, $p = 0.0073$, statistically

indistinguishable from IT: $\Delta = -0.0087$, $p=0.687$), though their absolute correlations remained lower than IT. Together, these findings show that IT's duration coding is shaped by stimulus history, and temporally integrated video ANNs partially capture this dependence, highlighting key constraints for future ANNs.

KK is supported by the Canada Research Chair Program (CRC-2021-00326), SFARI (967073), Brain-Canada Foundation (2023-0259), and NSERC (RGPIN-2024-06223).

TALK SESSION: SATURDAY, MAY 16, 2026, 5:15 – 7:00 PM, TALK ROOM 1

Scene Perception

Moderator: Miguel Eckstein, UC Santa Barbara

TALK 1, 5:15 PM, 25.11

LAION-FMRI: A DENSELY SAMPLED 7T-FMRI DATASET PROVIDING BROAD COVERAGE OF NATURAL IMAGE DIVERSITY

Josefine Zerbe^{1,2}, Johannes Roth^{1,2}, Maggie Mae Mell^{1,2}, Peer Herholz^{1,2}, Tomas Knapen^{3,4,5}, Martin N. Hebart^{1,2,6}; ¹Justus Liebig University, Giessen, Germany, ²Max-Planck Institute CBS, Leipzig, Germany, ³Spinoza Centre for Neuroimaging, Amsterdam, Netherlands, ⁴Netherlands Institute for Neuroscience, Amsterdam, Netherlands, ⁵Vrije Universiteit, Amsterdam, Netherlands, ⁶Center for Mind, Brain, and Behavior, Universities of Marburg, Giessen, and Darmstadt, Marburg, Germany

Research in visual neuroscience aims to arrive at generalizable conclusions, yet the complexity and breadth of our everyday visual experience make exhaustive stimulus sampling impossible. Massive fMRI datasets, such as BOLD5000, NSD or THINGS, have attempted broad coverage of the visual space by densely sampling a large-scale corpus of natural images from a small-scale cohort of human brains. Despite tremendous efforts involved in their collection, these datasets have challenges: First, recent work has questioned whether their stimulus sampling is sufficiently broad to allow for true generalization, given strongly overlapping training and testing distributions. And while some datasets provide out-of-distribution images, their scope remains limited. Further, existing datasets favor broad over deep sampling, making purely data-driven analyses of voxel selectivity challenging due to the noise inherent in the data. Addressing these challenges, we acquired a new multi-echo 7T functional MRI dataset (1.8mm isotropic, 1.9s TR) in which 5 individuals viewed 25,844 newly sampled unique images (2,284 shared) across 165 sessions. For broad representational sampling, most images were natural photographs derived from LAION-natural (120 million image-text pairs; Roth & Hebart, 2025), with an out-of-distribution subset including abstract shapes, visual illusions, and self-photographed images. Participants viewed centrally-presented stimuli (9×9 DVA; 2.5s on, 0.5s off;

repeated 4-12x) while performing a memory task. Preprocessing revealed single-trial betas with consistently excellent noise ceilings and minimal head motion estimates across all participants. Extending the scope, LAION-fMRI was collected in individuals for whom we previously acquired extensive retinotopic mapping, diverse functional localizers, and precision diffusion data for u-fiber mapping, providing the basis for a detailed understanding of visual functional neuroanatomy. We believe LAION-fMRI will be of broad utility for testing and evaluating the generalizability of models of vision and language, positioning itself uniquely for fine-grained theory-driven and data-driven analyses of visual patterns of brain activity.

This work was supported by a Max Planck Research Group Grant (M.TN.A.NEPF0009) and the ERC Starting Grant COREDIM (ERC-StG-2021-101039712) both awarded to MNH.

TALK 2, 5:30 PM, 25.12

BIFURCATION OF SCENE PERCEPTION AND MEMORY-RELATED PROCESSES IN SUPERIOR PARIETAL CORTEX

Adam Steel¹, Nicole Tang¹, Dominika Panek¹, Catriona Scrivener², Edward Silson²; ¹University of Illinois, ²University of Edinburgh

Scene perception and memory are closely linked in cortex: each of the three scene-selective brain areas in posterior cerebral cortex (parahippocampal place area, PPA), (occipital place area, OPA), and medial place area, MPA) has a paired area that activates when participants recall familiar places. Recently, a fourth scene-selective brain area located on the posterior intraparietal gyrus, referred to as PIGS (Kennedy et al., 2024, Yoon et al., 2025). Here, in two independent datasets, we investigated the topography of perception and memory responses for visual scenes in posterior intraparietal gyrus. In Dataset 1, we localized PIGS in 12 participants by comparing fMRI activity while participants viewed static images of scenes versus faces. In Dataset 2, we localized PIGS by comparing fMRI activity while 12 participants viewed 3-second videos of scenes, faces, objects, and bodies. In both datasets, we compared the locations of PIGS with place memory selective activity, localized by comparing activation when participants visually recalled personally familiar places versus people. In both datasets, we found a swath of memory-selective activity immediately anterior and adjacent to PIGS, consistent with the posterior-anterior perception-memory arrangement observed for the other scene selective areas. We then characterized these areas' position in the larger cortical hierarchy using resting-state functional connectivity. Intriguingly, PIGS sits at the confluence of two distinct cortical pathways: i) a posterior-medial pathway spans medial parietal cortex, including Area 23a, the parietal-occipital sulcus, and to the hippocampus, and ii) a lateral prefrontal pathway comprised of caudal and rostral intraparietal cortex, frontal eye fields, and inferior frontal junction. Together, these results show that all brain areas that process visual scenes have a paired memory-responsive area. Moreover, PIGS may be uniquely positioned at an intersection of two distinct processing streams that facilitate visually-guided navigation.

Brain and Behavior Research Foundation

TALK 3, 5:45 PM, 25.13

INTRACRANIAL RECORDINGS IN HUMAN POSTERIOR PARIETAL CORTEX REVEAL A NODE IN THE INTUITIVE PHYSICS NETWORK

Vasiliki Bougou¹ (cbougou@caltech.edu), RT Pramod², Jorge Gamez¹, Emily Rosario³, Charles Liu⁴, Kelsie Pejsa¹, Ausaf Bari⁵, Richard Andersen¹, Nancy Kanwisher²; ¹California Institute of Technology, ²Massachusetts Institute of Technology, ³Casa Colina Hospital and Center for Healthcare, ⁴University of Southern California, ⁵University of California Los Angeles

fMRI studies have identified parietal and frontal regions in humans that respond preferentially during intuitive physics judgments compared to difficulty matched control tasks (Fischer et al., 2016). Further, analysis of fMRI patterns has shown that these regions encode object mass, stability, contact, and imminent collisions (Pramod et al., 2021, 2025). However, clarifying the underlying neural computations requires finer-grained sampling of neural responses in space and time. Here, we recorded single unit activity (SUA) and local field potentials (LFPs) from two individuals with spinal cord injury. Participant RD had one Utah array in superior parietal lobule (SPL) and one in supramarginal gyrus (SMG), whereas participant JJ had one array in SPL and contributed only LFPs due to the long implantation duration. Across sessions in RD, we recorded from more than 500 units per array and directly replicated the fMRI intuitive physics experiments. In the tower falling task, firing rates were markedly higher during physical than color judgments (52 physics selective units in SPL, 37 in SMG, and fewer than 10 selective for the color task in either region). The same SPL array responded strongly to short videos showing physical events but minimally to videos showing social events (95 physical selective units versus 2 social). Yet, neither region showed increased responses or above chance decoding for difficult vs easy spatial working memory tasks, indicating that our findings cannot be attributed to these arrays being part of the multiple demand network. High gamma activity in both participants mirrored the spiking results from RD. Our results provide direct electrophysiological evidence that the posterior parietal cortex is engaged in intuitive physical reasoning, consistent with prior work showing that these areas encode internal models of the body and environment and form movement intentions, making intuitive physics a natural component of their function (Chivukula et al. 2025).

National Eye Institute grant UG1EY032039, Tianqiao and Chrissy Chen Brain-Machine Interface Center at Caltech, Swartz Foundation, Boswell Foundation, Fonds Wetenschappelijk Onderzoek (FWO) grant 1264926N

TALK 4, 6:00 PM, 25.14

CONTRIBUTIONS OF FOVEATION, SCENE AMBIGUITY, AND VISUAL COMPLEXITY TO HUMAN TIME TO COMPREHEND SCENES

Ziqi Wen¹ (ziqiwen@ucsb.edu), Jonathan Skaza², Sharvan Murlidaran³, Miguel P. Eckstein^{1,2,3}; ¹Department of Computer Science, UC Santa Barbara, ²Graduate Program in Dynamical Neuroscience, UC Santa Barbara, ³Department of Psychological and Brain Sciences, UC Santa Barbara

Introduction: Scene comprehension is central to everyday visual cognition. We understand the factors (e.g., target detectability/discriminability, crowding, expectations) influencing response times (RTs) for basic perceptual tasks such as pattern discrimination, identification, search, and scene classification. Furthermore, there are image-computable models to predict RTs for such tasks (Spoerer et al., 2020; Goetschalckx et al., 2023; Rafiei et al., 2024). However, less is known about the factors that contribute to RTs for scene comprehension, nor is there an image-computable model. Here, we investigate how the interactions between image content and foveation, visual complexity, and scene ambiguity contribute to human scene comprehension RTs. We also propose a multi-factor image-computable model to predict RTs and scene description accuracy under limited viewing time. **Methods:** We used image-computable scores to quantify the costs of foveation (foveated scene understanding, F-SUM; Wen et al., 2025), scene ambiguity (language entropy of Multi-Modal Large Language Models; LE, Malinin & Gales, 2020), and perceived image complexity (IC, using trained deep learning models; Skaza et al., 2025; Feng et al., 2023) for 277 scenes. We measured human (N = 100) RTs for comprehending the scenes, and the description accuracy for another observer group (N = 20) allowed only two saccades. **Results:** Foveation and scene ambiguity correlated the most with RTs (F-SUM: 0.561; LE: 0.528; IC: 0.422, $p < .05$). A linear combination of F-SUM, LE, and IC scores resulted in higher correlations (F-SUM+LE+IC: 0.676, $p < .0001$) than either metric alone. When observers were limited to two saccades, foveation (F-SUM) became the better predictor of description accuracy ($p < .0001$: F-SUM, $r = -0.54$; LE, $r = -0.405$; and IC, $r = -0.195$). **Conclusion:** Foveation and scene ambiguity contribute more than image complexity to scene comprehension RTs, but all three are complementary image-computable factors to best predict human RT prediction.

TALK 5, 6:15 PM, 25.15

SEEING WITHOUT SEEING: IMPLICIT PERIPHERAL COLOR PROCESSING

Shao-Min (Sean) Hung^{1,2} (konaes@gmail.com), Junhao Jiang², Sotaro Taniguchi², Katsumi Watanabe²; ¹Tohoku University, Sendai, Japan, ²Waseda University, Tokyo, Japan

Human visual experience often appears as subjectively rich and coherent across the visual field. However, recent work using virtual reality demonstrated that observers frequently fail to notice peripheral color removal even when explicitly instructed to detect it (Cohen et al., 2020), suggesting that peripheral color phenomenology may be impoverished. Critically, this does not preclude the possibility that unseen peripheral color is still processed and influences behavior. Here, we examined whether colors in the “undetectable periphery” contribute to visual performance. In Experiment 1 ($n = 30$), participants navigated a panoramic VR scene in which peripheral colors were gradually desaturated. At random intervals, a fixation cross appeared, followed by a target dot presented in one of three regions: saturated/near, boundary, or desaturated/peripheral. Participants responded faster and localized targets more accurately in saturated/near regions. In Experiment 2 ($n = 20$), we controlled for eccentricity by including both saturated and desaturated conditions in the periphery. To ensure that color differences remained undetectable, each participant completed a pre-experiment calibration to estimate their individual chance-level color detection radius. The results replicated: even when participants could not phenomenally distinguish saturated from desaturated periphery, targets appearing in physically saturated regions yielded better performance. These findings show that peripheral colors—despite our limited capacity to reliably detect them—can facilitate visual detection and localization. More broadly, they reveal a dissociation between peripheral color phenomenology and the functional use of peripheral color information in guiding behavior.

SMH was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI “Early-Career Scientists” (24K16877).

TALK 6, 6:30 PM, 25.16

HOW GAZE BEHAVIOR EXPLAINS GENDER DIFFERENCES IN SPATIAL LEARNING IN MIXED REALITY NAVIGATION

Yu Zhao¹ (yzhao20@kennesaw.edu), Holly Gagnon, Jeanine Stefanucci², Sarah Creem-Regehr², Bobby Bodenheimer³; ¹Kennesaw State University, ²University of Utah, ³Vanderbilt University

Gender differences are commonly observed in the adoption of strategies for spatial cognition as well as for certain immersive technologies. Understanding such gender-specific behaviors is crucial for designing inclusive technologies to support spatial tasks. We conducted a gender-balanced experiment with 90 participants who performed a navigation task with three mixed reality navigation interfaces to investigate how gaze behaviors relate to spatial learning performance, differ by gender, and potentially mediate the gap in spatial learning that is often observed between genders. The study consisted of a spatial learning phase (memorizing landmarks while following route) and a retrieval phase (estimating directions with a pointing task). Gaze metrics, including gaze dispersion, saccadic

amplitude, fixation distance, and gaze entropy, were calculated. Using multilevel mediation analysis, we assessed (1) if gender predicted gaze behavior; (2) if gaze predicted pointing error; and (3) the mediation effect of gaze behavior using the Aroian Test, controlling for interface condition and trial order. Our results reveal a phase-dependent effect. During the learning phase, female participants adopted distinct visual strategies, including significantly wider horizontal scanning and higher gaze entropy compared to males. However, these strategies did not mediate spatial learning performance. In contrast, during the pointing phase, horizontal gaze dispersion emerged as a significant mediator. Females continued to exhibit wide horizontal scanning during retrieval, a behavior that strongly predicted higher pointing errors. We also distinguish this gender-linked strategy from general states of confusion: while gaze entropy predicted errors for all users, it did not differ by gender. These findings suggest that the gender gap in spatial learning is not driven by inefficient gaze behavior of information gathering during learning, but a horizon-scanning behavior during retrieval that is specific to females may lead to a performance deficit compared to males.

This material is based upon work supported by the Office of Naval Research under grant N0014-21-1-2583

TALK 7, 6:45 PM, 25.17

IDIOSYNCRASIES IN INTERNAL MODELS PREDICT INDIVIDUAL DIFFERENCES IN SPATIOTEMPORAL NEURAL PROCESSING OF NATURAL SCENES

Micha Engeser^{1,2} (micha.engeser@math.uni-giessen.de), Thea Schmitt¹, Daniel Kaiser^{1,2,3}; ¹Neural Computation Group, Department of Mathematics and Computer Science, Physics, Geography, Justus Liebig University Giessen, 35392 Giessen, Germany, ²Center for Mind, Brain and Behavior (CMBB), Philipps University Marburg, Justus Liebig University Giessen, and Technical University Darmstadt, 35032 Marburg, Germany, ³Cluster of Excellence "The Adaptive Mind", Justus Liebig University Giessen, Philipps University Marburg, and Technical University Darmstadt, 35392 Giessen, Germany

Why do humans differ in how they perceive the world around them? Traditionally, this question has received limited attention, with variability between participants often dismissed as noise. Building on predictive processing theories, we propose that idiosyncrasies in internal models—expectations about what the world should look like—are a key source of such perceptual variability. Using an inter-subject representational similarity analysis (IS-RSA), we tested whether inter-individual similarities in internal models for natural scene categories predict similarities in neural fMRI and EEG responses when viewing scenes from these categories. To characterize internal models, participants drew what they considered the most typical version of specific scene categories. We then used deep-learning tools to transform the drawings into photorealistic images and, in turn, quantify inter-individual similarities in the resulting images. Relating the

resulting inter-individual similarities in internal models to inter-individual similarities in neural responses yielded two key insights: First, participants with more similar internal models showed greater alignment in fMRI BOLD time courses within lateral occipital and lateral prefrontal cortices. Second, participants with more similar internal models exhibited more similar scene representations in EEG signals, emerging around 400 ms after stimulus onset. Together, these findings demonstrate that individual priors regarding the structure of the world offer a parsimonious explanation for why spatiotemporal processing in the visual system varies across individuals.

TALK SESSION: SATURDAY, MAY 16, 2026, 5:15 – 7:00 PM, TALK ROOM 2

Visual Working Memory: Influences, models

Moderator: Brad Wyble, Penn State University

TALK 1, 5:15 PM, 25.21

UNFORGETTABLE LESSONS FROM FORGETTABLE IMAGES: INTRA-CLASS MEMORABILITY MATTERS IN COMPUTER VISION

Serena Wang¹ (serena.wang.wjw@gmail.com), Jie Jing^{1,2}, Yongjian Huang¹, Shuangpeng Han¹, Lucia Schiatti^{3,4}, Yen-Ling Kuo⁵, Qing Lin¹, Mengmi Zhang¹; ¹Nanyang Technological University, ²Sichuan University, China, ³Massachusetts Institute of Technology, USA, ⁴Istituto Italiano di Tecnologia, Italy, ⁵University of Virginia, USA

We introduce intra-class memorability, where certain images within the same class are more memorable than others despite shared category characteristics. To investigate what features make one object instance more memorable than others, we design and conduct human behavior experiments, where participants are shown a series of images, and they must identify when the current image matches the image presented a few steps back in the sequence. To quantify memorability, we propose the Intra-Class Memorability score (ICMscore), a novel metric that incorporates the temporal intervals between repeated image presentations into its calculation. Furthermore, we curate the Intra-Class Memorability Dataset (ICMD), comprising over 5,000 images across ten object classes with their ICMscores derived from 2,000 participants' responses. Subsequently, we demonstrate the usefulness of ICMD by training AI models on this dataset for various downstream tasks: memorability prediction, image recognition, continual learning, and memorability-controlled image editing. Surprisingly, high-ICMscore images impair AI performance in image recognition and continual learning tasks, while low-ICMscore images improve outcomes in these tasks. Additionally, we fine-tune a state-of-the-art image diffusion model on ICMD image pairs with and without masked semantic objects. The diffusion model can successfully manipulate image elements to enhance or reduce memorability. Our

contributions open new pathways in understanding intra-class memorability by scrutinizing fine-grained visual features behind the most and least memorable images and laying the groundwork for real-world applications in computer vision. We will release all code, data, and models publicly.

TALK 2, 5:30 PM, 25.22

HOW (NOT) TO MEASURE MEMORABILITY: UNDERLYING "MEMORY AXES" REVEAL THAT MEMORABILITY STUDIES OFTEN REFLECT RESPONSE BIAS

Niels Verosky¹, Brian Scholl¹; ¹Yale University

Memory for a visual stimulus depends not only on the observer, but also on the stimulus itself: some objects or events seem to stick in our minds much more readily than others. But just how can — and should — this type of visual *memorability* be measured? A tidal wave of recent work has almost always employed either 'corrected recognition' (hit rate minus false-alarm rate) or d' (as a measure of signal-detection sensitivity). Here we show how these measures actually capture a mixture of memorability *and response bias* — and in the worst cases may degenerate into pure measures of bias. We demonstrate this in three steps. First, we show how this problem can occur in principle, in constructed examples. Second, we show how it operates in practice, in a set of actual case studies from the published empirical literature — highlighting examples both where this problem is tempered, and where it is catastrophic. Third, we survey the extent of this problem across the literature, in an analysis of 34 previous memorability experiments. We find that the clear majority of experiments show an unexpected positive correlation between hits and false alarms (indicative of response bias), and that inadvertently measuring response bias is ubiquitous. To address the serious limitations of existing measures, we introduce a new way of measuring memorability based on estimating "memory axes". This approach makes explicit the underlying covariance structure of the data, using principal-components analysis to directly orthogonalize memorability from response bias. We suggest that future work on memorability should aim to separate response bias and memorability as distinct dimensions of stimulus-driven memory performance. Doing so may profoundly change the theoretical interpretation of visual memorability studies.

TALK 3, 5:45 PM, 25.23

SEMANTICALLY-RICH OBJECTS INCREASE THE DISTINCTIVENESS OF LOW-LEVEL VISUAL FEATURES

Joseph M. Saito¹ (josaito@ucsd.edu), Yong Hoon Chung², Viola S. Störmer², Timothy F. Brady¹; ¹University of California San Diego, ²Dartmouth College

A longstanding question in the field of visual memory concerns the nature of stimulus representations that are stored in the absence of sensory input. Models inspired by the architecture of the visual system posit that individual stimuli are represented at multiple levels of

complexity along the visual hierarchy, ranging from bundles of low-level features (e.g., colors, orientations) up to higher-level objects (e.g., fruits, animals). Critically, these models predict that binding similar low-level features to different object-level representations can reduce feature confusability when those objects activate additional semantic knowledge that increases the dimensionality of each memory. To test this prediction, we used a continuous estimation working memory paradigm where observers were asked to remember pairs of similar colors that were presented as part of semantically-rich real-world objects (e.g., backpack, camera) or their unrecognizable, scrambled counterparts (e.g., scrambled backpack, camera). To minimize any incidental long-term associations between specific objects and colors (e.g., red on an apple), we chose real-world objects that were color-neutral and assigned the colors to them randomly. Furthermore, to isolate the distinctiveness conferred specifically by the semantic richness of the objects, we used the same color pairs in each condition and closely matched the visual similarity between the shapes of the intact and scrambled objects using feature values extracted from the average pool layer of the convolutional neural network model, VGG16. In doing so, we found that observers committed fewer swap errors and exhibited weaker repulsion biases when the same color pairs were encoded as part of real-world objects than scrambled objects, consistent with hierarchical memory models. These findings suggest that semantically-rich object representations reduce the confusability between low-level features by enhancing feature binding. This enhanced binding, in turn, reduces the reliance on compensatory processes, like adaptive distortions, to preserve the distinction between similar features in memory.

This work was supported by NSF BCS-2146988 awarded to TFB

TALK 4, 6:00 PM, 25.24

ADAPTIVE COMPUTATION IN WORKING MEMORY: GOAL-CONDITIONED SPARSE VARIATIONAL GAUSSIAN PROCESS EXPLAINS RETRO-CUE BENEFITS AND INTERFERENCE DYNAMICS

Dongyu Gong¹, Mario Belledonne¹, Ilker Yildirim¹; ¹Yale University

Despite extensive research on working memory (WM) limits, an integrative algorithmic account that explains how finite neural resources are dynamically reconfigured to support goal-directed behavior remains elusive. We propose a novel algorithmic model of WM based on Sparse Variational Gaussian Processes (SVGP), where memory is conceptualized as a continuous density estimation problem and capacity limits arise from a finite set of "inducing points" (computational resources) used to approximate this function. In this framework, we model the maintenance of visual information as an online, goal-driven optimization process. We tested the model against human performance in a preregistered behavioral experiment (N=30) using a continuous color reproduction task with retro-cues. Participants memorized the colors of 2 or 4 items; after a delay, a spatial cue either indicated the target item (retro-cue) or no cue was

provided. The SVGP model implements a novel "adaptive computation" mechanism: during the delay, top-down goals (cues) dynamically re-weight the model's variational objective function. This forces the limited inducing points to migrate toward the cued location in representational space, sharpening the fidelity of the target while degrading non-targets. The model quantitatively predicts three key empirical signatures observed in our human data: 1) A robust retro-cue benefit (faster RT and reduced error) driven by the reallocation of inducing points; 2) A set-size effect where precision degrades as inducing points are stretched across multiple items; and 3) complex interference patterns where spatial proximity interacts with feature similarity. Moreover, the model reproduces the WM distortion phenomena reported by Chunharas et al. (2022), capturing the transition from attraction, which dominates when capacity is overloaded (e.g., high set sizes), to repulsion, which emerges when capacity is sufficient (e.g., low set sizes). By formalizing WM as an online, goal-driven density approximation process, this work offers an algorithmic bridge between neural dynamics and flexible cognitive control.

TALK 5, 6:15 PM, 25.25

MODELLING THE COMPOSITIONAL AND GENERATIVE ASPECTS OF VISUAL WORKING MEMORY

Brad Wyble¹, Ian Deal¹; ¹Penn State University

Visual working memory (VWM) is conceived as holding multiple objects in service of other cognitive tasks. We develop neurocomputational mechanisms that allow compositional and generative processes to expand on this traditional view. Compositionality allows objects or scenes to be decomposed into constituents that can be manipulated or recombined to form new representations. Generative processing allows conceptual information to be reconstructed in a format akin to visual sensory representations that can be re-processed by perception. Together, compositional and generative mechanisms enable the traditional memory comparison functions of VWM as well as forming a basis for visual imagery and some aspects of creativity. Our model uses a modified variational autoencoder to simulate visual processing of distinct representations of location, size, color and shape features, that are extracted from a visual canvas. These representations are then selectively stored in a binding pool that groups features into units akin to pointers or tokens. Once stored, memories can be selectively pushed back into the processing hierarchy to regenerate visual forms. This model exhibits typical effects of working memory storage including set size effects, loss of precision with load, and feature swaps. It can also store individuated copies of repeated items and novel shapes, but exhibits better memory for familiar shapes. The model also connects visual patterns with categorical abstractions of shape and color thereby allowing it to classify visual forms and even to generate new visual forms based on top-down instructions. Our model suggests that generative and compositional properties provide the functionality of both visual memory and visual imagery. The model can store multi-

feature objects, selectively modify one feature without modifying others, compose simple visual scenes piecemeal and then recognize patterns in those created scenes. These results provide concrete computational instantiations of basic cognitive mechanisms that allow a much wider scope of functionality.

This work was supported by NSF grant BCS-2216127

TALK 6, 6:30 PM, 25.26

VISUALIZING SCENE UNDERSTANDING IN SHORT AND LONG-TERM MEMORY

Gregory Zelinsky¹, Ritik Raina¹, Abe Leite¹, Seoyoung Ahn²; ¹Stony Brook University, ²University of California, Berkeley

At VSS25, we introduced a real-time method for identifying AI-generated scenes that viewers cannot discriminate from originals—scene metamers. Here we show how our method enables the testing of previously untestable hypotheses about scene representation. We evaluated the hypothesis that long-term memory (LTM) representations may be simply language-based generations, whereas short-term memory (STM) representations are primarily visual. We tested this hypothesis by generating images under these conditions and comparing metamerism rates. To obtain scene generations, we introduce Seen2Scene, a generative AI model that combines fixation tokens (DINOv3 patches) with a gist-like scene representation (from blurred peripheral pixels) to generate plausible proxies for scene understanding. Experiment 1 obtained STM metamers by showing participants (n=47) each scene (n=300) for 1-10 fixations, followed by an 8-second delay during which Seen2Scene generated another scene from the participant's viewing behavior. We then briefly (200 msec) presented either the originally viewed scene or the generation and asked the participant to make a same/different judgment. Metameric scenes were generations incorrectly eliciting "same" responses. Experiment 2 obtained LTM metamers for the same scenes using the same Seen2Scene model. Participants first viewed a scene (1-10 fixations) and then immediately reported (verbally, via microphone) a detailed description of the just-viewed scene. After a minimum 1-day delay, participants returned to the lab where they were also shown scenes (200 msec) in a same/different task. Different images were either actually different or were generations from the participant's verbal description, plus baseline variants. We found high metamerism rates in Experiment 1, and much lower metamerism rates in Experiment 2 when scenes were generated using language alone. Adding vision to language, however, resulted in metamerism rates jumping to Experiment 1 levels, suggesting that LTM scene representations are not generated from language alone and require visual representation.

RR and GJZ are supported by NSF-CompCog #2444540 to GJZ. AL is supported by NSF-GRFP #2234683 and NIH-NEI R01EY030669 to GJZ.

TALK 7, 6:45 PM, 25.27

FROM BINDING TO BIAS: LEARNED STIMULUS–ACTION ASSOCIATIONS SHAPE VISUAL MEMORY

Cate Trentin^{1,2} (c.trentin@vu.nl), Chris N.L. Olivers^{1,2}, Heleen A. Slagter^{1,2}; ¹Vrije Universiteit Amsterdam, ²Institute for Brain and Behaviour Amsterdam (iBBA)

Theories propose that visual representations, both perceptual and mnemonic, are organized not merely by their visual properties, but also by their relevance for action. Consistent with this view, we recently found that two items held in visual working memory are remembered as more dissimilar when mapped to two different actions than when mapped to the same action. This action-dependent distortion arose very quickly, when mappings varied on a trial-by-trial basis. Here, in two experiments (N=48 each), we examined whether stable stimulus-action mappings further amplify this representational bias. On every trial, participants memorized the orientations of two bars and later reported them on a touchscreen by performing a cued action (i.e., grip or slide). For half of the stimulus pairs, action mappings were fixed throughout the experiment. For the remaining pairs, action mappings varied randomly across trials. Learning of the fixed contingencies was assessed afterwards. In Experiment 1, we replicated our earlier finding: For the randomly varied mappings, orientations linked to different actions repelled each other more than those linked to the same action. Critically, this repulsion effect was enhanced for the fixed mappings, but only for participants who had learned the contingencies. To increase learning, in Experiment 2 we reduced the number of contingencies. We again observed stronger action-based repulsion for learned (fixed) mappings than for random mappings, now for the full group. These findings indicate that as stimulus-action associations become consolidated through learning, they increasingly shape visual memory representations.

TALK SESSION: SUNDAY, MAY 17, 2026, 8:15 – 9:45 AM, TALK ROOM 1

Attention 2

Moderator: Thomas Sprague, UC Santa Barbara

TALK 1, 8:15 AM, 31.11

CORTICAL DOUBLE DISSOCIATION FOR ENDOGENOUS AND EXOGENOUS COVERT SPATIAL ATTENTION

Marisa Carrasco¹, Nina M. Hanning^{1,2}, Antonio Fernandez¹, Qingyuan (Rachel) Chen¹, Hsing-Hao Lee¹; ¹NYU, ²Humboldt-Universität zu Berlin

[Background] Psychophysical studies have shown that both endogenous (voluntary) and exogenous (involuntary) covert spatial attention enhance visual sensitivity, but in distinct ways. fMRI research

indicates that these attention types recruit partially overlapping brain regions. However, because fMRI is correlational, neurostimulation is required to establish critical and infer causal involvement. Here, we used transcranial magnetic stimulation (TMS) to investigate whether early visual cortex (V1/V2) and the human homologue of the right frontal eye field (rFEF+) are equally critical for both attention types. [Method] We used double-pulse TMS during an orientation-discrimination task in which attention was manipulated with valid or invalid cues preceding one of two oriented Gabor patches or neutral cues preceding both. In the endogenous-attention experiments, an informative central cue preceded the Gabor stimuli by 500 ms; in the exogenous-attention experiments, a non-informative peripheral cue preceded the display by 100 ms. A response cue indicated the target location—the stimulus on which the orientation discrimination was to be performed. MRI-guided TMS was applied to each observer's V1/V2 or rFEF+ during stimulus presentation, targeting the cortical representation of either to the target or the distractor. Psychometric contrast-response functions were obtained, and sensitivity (d') was analyzed separately for the target-stimulated and distractor-stimulated hemifields. [Results] For both attention types and both areas, when the distractor was stimulated, we observed response-gain benefits at attended locations and costs at unattended locations. Critically, however, when the target was stimulated, TMS to V1/V2 eliminated exogenous but not endogenous attention effects. In contrast, TMS to rFEF+ significantly reduced endogenous attention, but not exogenous attention. [Conclusion] These findings reveal a double dissociation: early visual areas V1/V2 are critical for exogenous but not endogenous attention, whereas rFEF+ is essential for endogenous but not exogenous attention. This dissociation highlights that distinct areas support voluntary and involuntary covert selection.

TALK 2, 8:30 AM, 31.12

NEURAL MARKERS OF DISTRACTOR SUPPRESSION IN THE EARLY VISUAL CORTEX

Sojung Youn¹, Brian Anderson¹; ¹Texas A&M University

Although physically salient stimuli generate a strong priority signal in the attentional system, it is possible to suppress attentional processing of salient stimuli under certain conditions. Distractor suppression has been extensively examined using behavioral and electroencephalographic measures; however, its underlying neural mechanisms, particularly at the early stages of visual information processing, remain underexplored. Therefore, in the present study, we investigated whether we are able to detect evidence of distractor suppression at a neural level in the early visual cortex using functional magnetic resonance imaging (fMRI), in an attempt to better understand the core mechanism of distractor suppression. Participants performed a four-item feature-based visual search task for a shape-defined target amongst heterogeneously-shaped non-targets, one of which was sometimes rendered in a unique color (the physically salient distractor). We observed behavioral evidence of suppression

where participants showed a significant distractor-presence benefit in response time during both the laboratory practice session (48ms) and while performing the task in the scanner (42ms). This behavioral indicator of distractor suppression was positively correlated with distractor-evoked blood oxygen level dependent (BOLD) response in the early visual cortex, suggesting that the initial strength of the priority signal evoked by the distractor in the early stages of visual information processing could serve as a marker triggering suppression at later stages of visual information processing. In addition to our current findings, all stimuli were presented with superimposed oriented gratings and multivoxel pattern analysis (MVPA) will be performed on the gratings to examine how distractor suppression is related to the fidelity of stimulus representation in early visual areas.

TALK 3, 8:45 AM, 31.13

WHOLE-BRAIN REPRESENTATION OF OBJECT, ATTENTION, AND LOAD

Shijie Qu¹ (shijie.qu@yale.edu), Marvin Chun¹, Yaoda Xu¹; ¹Yale University

When exploring the world around us, we often direct attention to specific features of a target object under varying levels of task demands. Thus, to support visual processing, the brain needs to represent object information, attentional modulation, and the overall task load. Although prior studies have examined these components in isolation and have linked object representations primarily to occipitotemporal cortex (OTC) and task representations to frontoparietal cortex, it remains largely unknown how the representations extend beyond these localized regions and interact across the whole brain. In this fMRI project, we tested this idea by placing colored dots on top of objects and orthogonally manipulating object category, selective attention (attending to color vs shape), and task load (1back vs 2back repetition detection). Multivoxel pattern analysis revealed that all three forms of information were reliably decodable across occipitotemporal cortex (OTC), posterior parietal cortex (PPC), and prefrontal cortex (PFC). While object category was decoded most strongly in OTC, its representation became increasingly distributed over PPC and PFC, encompassing most of the cortical surface when shape was attended at the higher load. In contrast, attention and load were consistently decodable across much of the cortical surface, including and beyond OTC, PPC, and PFC. Cross decoding analyses further showed that object and attention representations were highly interdependent, whereas task load was encoded more independently across the cortical surface. Together, in addition to confirming the role of canonical regions such as OTC, PPC, and PFC in object, attention, and load processing, respectively, our findings provided fresh evidence that these representations are not uniquely localized to these regions, but are instead encoded simultaneously within and beyond them across the broader cortical surface.

This research was supported by NIH grant R01EY030854 to YX.

TALK 4, 9:00 AM, 31.14

A CENTRAL-FIELD FOCUS IN VENTRAL-STREAM FEEDBACK TO THE PRIMARY VISUAL CORTEX (V1) IN PRIMATES: THEORETICAL PREDICTION CONFIRMED

Piotr Majka¹, Li Zhaoping², Marcello Rosa³; ¹Laboratory of Neuroinformatics, Nencki Institute of Experimental Biology of the Polish Academy of Sciences, ²University of Tübingen, Max Planck Institute for Biological Cybernetics, Tübingen, Germany, ³Department of Physiology and Neuroscience Program, Biomedicine Discovery Institute, Monash University, Clayton, Australia

The central-peripheral dichotomy (CPD) theory is motivated by the fact that an attentional bottleneck allows only a tiny fraction of visual inputs to reach recognition. It proposes that the peripheral visual field primarily serves to select this fraction --- by directing gaze or attention --- whereas the central field is specialized for object recognition within the selected inputs. Because of this bottleneck, the feedforward information from the primary visual cortex (V1) to downstream visual areas is limited, constraining ongoing recognition. To compensate, downstream areas send feedback to upstream regions such as V1 to query for task-relevant information. The CPD theory therefore proposes that this feedback should target mainly the central visual field. Since the ventral stream is specialized for recognition, the theory predicts that ventral stream feedback to V1 is concentrated in the central field representation. We tested this prediction by injecting cellular-resolution retrograde anatomical tracers in locations representing various visual eccentricities (2° - 18°) in V1 of marmoset monkeys. Given V1's position in the visual processing hierarchy, feedforward connections to V1 originate primarily (if not exclusively) in the lateral geniculate nucleus, whereas feedback connections arise from multiple cortical areas. As predicted, the proportion of V1-projecting connections originating from feedback sources varied systematically with eccentricity. In the ventral stream as well as V2, this proportion was approximately ten times higher for the central field representation than for the peripheral representation. Moreover, with V2 excluded, ventral stream dominated feedback to the central visual field (<5°), whereas dorsal stream dominated feedback to peripheral V1 (up to 18°). These findings suggest that the brain's architecture for vision should be understood not only in terms of distinct cortical areas, but also in terms of the qualitative differences in computation applied to the central versus peripheral visual fields in each area.

National Science Centre grant (2019/35/D/NZ4/03031) to PM; European Commission EBRAINS 2.0 (101147319 FSTP) to PM; University of Tübingen and the Max Planck Society to LZ;

TALK 5, 9:15 AM, 31.15

HISTORY-DRIVEN SUPPRESSION RESHAPES VISUOSPATIAL TUNING ACROSS THE VISUAL HIERARCHY

Dock Duncan¹, Gilles de Hollander², Ningkai Wang¹, David Richter³, Tomas Knapen¹, Jan Theeuwes^{1,3,4,5}; ¹Vrije Universiteit Amsterdam, the Netherlands, ²University of Zurich, Switzerland, ³University of Granada, Spain, ⁴William James Center for Research, ISPA-Instituto Universitario, Lisbon, Portugal, ⁵Zhejiang University, Hangzhou, China

A large body of behavioural work has provided strong evidence that visual information processing dynamically adapts to recent experiences, already at the earliest levels of the visual system. However, the precise neural mechanisms by which our visual system is biased towards previous target locations and away from previous distractor locations remains largely speculative. Here, in a high powered 7-tesla fMRI project, we investigated how history effects alter visuospatial tuning during visual search using a novel population receptive field (pRF) modelling approach. Thirty participants across two sessions each performed a version of the additional singleton task with a high-probability distractor location. While participants searched for the target while ignoring the distractor, we continuously presented moving, flickering checkerboard bars in the centre of the screen, allowing us to model the receptive field responses across the visual processing hierarchy concurrently while they performed the task. Estimated population receptive field parameters show that visuospatial receptive fields shifted away from the current high-probability distractor location in each block. The magnitude of receptive field shifts increased up the visual processing hierarchy, reflecting growing field sizes and implicating higher visual processing areas as more influential in the familiar learned suppression effect. Together, our results provide novel insights into how experience alter the flow of salient information through the visual system by biasing competitive interactions away from suppressed regions. Crucially, we demonstrate that our early visual system is not static but dynamically adapts to recent experiences, with implications for how we should think of early visual processing in the brain.

Jan Theeuwes was supported by an NWO grant (406.21.GO.034)

TALK 6, 9:30 AM, 31.16

THE BRAIN MECHANISMS OF ATTENTIONAL LAPSES DURING GOAL COMPETITION: AN EEG STUDY

Matthieu Chidharom¹ (matthieuchidharom@gmail.com), Monica Rosenberg¹, Edward Vogel¹; ¹The University of Chicago

Sustained attention is the ability to maintain focus on a specific goal over time, but attentional lapses are frequent. Theories have attributed these lapses to a transient failure of cognitive control in maintaining

the goal in mind. However, this proposal has been challenged by recent findings showing greater engagement of cognitive control during states more prone to lapses. To explain this result, we proposed that lapses occur when goals compete, requiring stronger cognitive control to sustain performance. We recently tested this goal-competition hypothesis with a Switch-Continuous Performance Task (CPT) in which subjects alternated task goals between blocks—either switching or holding the same goal—in an effort to manipulate periods of higher and lower competition between goals. Participants viewed a bilateral display showing a letter (vowel/consonant) and a number (even/odd) on each trial. After every 20 trials, a cue instructed participants to perform either the letter task (e.g., press for frequent vowels, not infrequent consonants) or the number task (e.g., press for frequent even, not infrequent odd, numbers). Our results showed that lapses were more frequent when competition between goals was higher during switch compared to repeat trials. Although this result has been replicated several times, the brain mechanisms by which competing goals favor lapses remain unclear. In the current study, we recorded the EEG of 30 participants during the Switch-CPT. Our ERP analysis revealed that competition between goals reduces the attention allocated to the relevant stimulus, as revealed by the lower N2pc amplitude during switch trials. Follow-up analyses showed that this effect is driven by a failure to suppress the distracting goal. A decoding analysis also revealed lower classification accuracy of task goals during switch compared to repeat trials, providing additional evidence that the relevant goal is less activated in mind during competition.

TALK SESSION: SUNDAY, MAY 17, 2026, 8:15 – 9:45 AM, TALK ROOM 2

Object Recognition

Moderator: Zeynep Saygin, The Ohio State University

TALK 1, 8:15 AM, 31.21

CHARACTERIZING VISUAL FEATURE ENCODING IN DROSOPHILA VISUAL PROJECTION NEURONS LPLC1 AND LPLC2

Bryce Hina¹ (bwh44@drexel.edu), Haley Croke¹, Anthony Moreno-Sanchez², Omika Wadhwa², Natalie Smolin¹, Jessica Ausborn², Catherine von Reyn^{1,2}; ¹Drexel University, School of Biomedical Engineering, Science & Health Systems, ²Drexel University College of Medicine, Neurobiology & Anatomy

Visual systems play a crucial role in quickly detecting important features in our surroundings, such as potential threats. However, understanding how visual systems encode visual features has been challenging, as many visual feature encoding cell types have yet to be fully identified. We here benefit from the fruit fly, *Drosophila*

melanogaster, where all neurons in the visual system have recently been identified and mapped at EM resolution. We focus on a class of neurons called visual projection neurons (VPNs), which have been hypothesized to selectively encode visual features. Here, we provide the first evidence that VPns encode visual information through spiking activity, with a well-defined axonal initial segment. By recording single cell activity while presenting visual stimuli, we reveal two VPN cell types (LPLC1 and LPLC2) exhibit robust spiking responses to both expanding (looming) objects and small, translating objects, suggesting VPns are less selective to visual features than previously hypothesized. Next, we mapped the receptive fields of VPN using expanding disks and translating stimuli, and found receptive fields based on spiking responses are smaller than those reported with Ca²⁺ imaging. Finally, we developed a pipeline for rapid receptive field mapping to further investigate differences in VPN tuning across a stimulus parameter space and examine habituation in visual responses. Our work suggests the encoding of visual features enlists spiking activity across multiple VPN cell types, generating a population code more similar to what is seen in vertebrate visual processing. Our work therefore has potential to uncover these population codes, and to also establish how they are generated through cellular and circuit mechanisms.

Funding was provided by the National Science Foundation Grant No. IOS-1921065 (Catherine R. von Reyn) and the National Institutes of Health NINDS R01NS118562 (Catherine R. von Reyn, Jessica Ausborn) and NEI F31EY037197 (Bryce W. Hina).

TALK 2, 8:30 AM, 31.22

CHARACTERIZING THE INPUTS TO INFANTS' OBJECT CATEGORY REPRESENTATIONS

Jane Yang¹ (jzyang@ucsd.edu), Tarun Sepuri¹, Alvin W.M. Tan², Khai Loong Aw², Michael C. Frank², Bria Long¹; ¹University of California San Diego, ²Stanford University

Infants acquire object category representations from their everyday experiences in the first few years of life. What do the inputs to this learning process look like? Here, we capitalize on innovations in data and in machine learning to characterize the structure of naturalistic infant experience. We analyzed egocentric videos from the BabyView dataset (N=31, 5–36 months, 868 hours), using an object detection model (YOLOE) to detect 205 categories from a vocabulary checklist in 3.68M frames (1 fps). Manual annotations of 480 frames suggested reasonable detection performance for 163 categories (average precision = 0.6, SD = 0.26); false alarms were filtered by comparing vision embeddings of cropped objects to text embeddings of their predicted labels in a multimodal CLIP model, retaining detections with cosine similarity > 0.26. Consistent with prior work, infants' object exposure was highly skewed: a few categories (chair, toy, book) dominated infants' visual experiences while most categories appeared rarely. We compared infants' object exposures to exemplars of

categories in a curated image dataset (THINGS) by examining average category embeddings across all exemplars in both datasets in a multimodal (CLIP) and a self-supervised visual encoder (DINOv3). The category-level similarity between BabyView and THINGS varied widely (e.g., zebra, $r = 0.75$, slipper, $r = 0.15$), as infants encountered object categories in non-canonical forms—occluded, cluttered, as depictions, and from unusual angles. Nonetheless, we found that representational dissimilarity matrices of the between-category similarity in BabyView vs. THINGS were moderately correlated (CLIP Spearman's $\rho = 0.37$, $p < .01$; DINOv3 Spearman's $\rho = 0.31$, $p < .01$). These results suggest that infants experience infrequent exemplars of many categories with notably different types of diversity than is captured in modern curated datasets. Models seeking to emulate human-like category learning must grapple with how children learn so efficiently from these radically different experiences with object categories.

We thank the families who participated in the BabyView Dataset. This work was supported by NIH R00HD108386 (to B.L.), Schmidt Futures, Meta, the Stanford Center for the Study of Language and Information, and the Stanford HAI Hoffman-Yee program.

TALK 3, 8:45 AM, 31.23

VOT SPECIALIZATION FOR READING IN CONGENITALLY BLIND INDIVIDUALS

MARIA CZARNECKA¹, Florencia Martinez Addiego², Marcin Szwed³; ¹Jagiellonian University, ²Georgia Institute of Technology, ³Jagiellonian University

A central component of reading is linking written symbols to their meanings. The left ventral occipitotemporal cortex (vOT) is considered a key interface for this function. This study examines what shapes the development of the neural correlates of reading—whether they stem from an innate sensitivity to shapes refined through visual experience, or instead from mechanisms that emerge independently of vision. To investigate this, we conducted an fMRI experiment with 21 congenitally blind and 21 sighted participants who read words through touch (Braille) or vision (print), respectively. The stimuli were designed to tap into multiple processing levels: low-level (visual for print, spatial for Braille), orthographic, and semantic. Using representational similarity analysis (RSA), we observed that both orthographic and semantic representations are present in the left vOT of blind readers, suggesting that this region adopts a reading-related role even without visual input. The primary group difference appeared at a low-level stage. In sighted participants, processing was localized to the early visual cortex, whereas in blind participants it recruited sensorimotor regions. Overall, these findings demonstrate that despite differences in low-level sensory pathways, the left vOT assumes a similar functional role in reading regardless of visual experience.

TALK 4, 9:00 AM, 31.24

LINKING FUNCTIONAL ARCHITECTURE TO PATHOLOGY: NEURAL OVERLAP AND INTER-CATEGORY ASSOCIATIONS DICTATE STROKE-INDUCED VISUAL RECOGNITION DEFICITS

Laura Soen¹, Roos Malpart¹, Céline Gillebert¹, Hans Op de Beeck¹;
¹KU Leuven

Visual recognition relies on specialized, category-selective functional architecture within the occipitotemporal cortex (OTC). While focal stroke damage can induce classic category-specific deficits (e.g., prosopagnosia, pure alexia, object agnosia), lesions rarely affect a single category. The resulting patterns of co-occurring impairments remain theoretically and clinically unclear. We provide a comprehensive, multi-modal framework to systematically study the OTC organization and demonstrate how this structure dictates stroke-induced deficits. We employed a three-way validation approach integrating behavioral, neural, and patient data. We developed and validated the Word, Object, and Face Categorization Test (WOF-CT) to assess recognition across 10 neuropsychologically relevant categories in three cohorts: young adults (age<30, N = 250), healthy older adults (age>50, N = 95), and stroke patients with damage limited to the posterior cerebral artery territory (N = 22). In healthy participants (N = 28), we used fMRI with Multivoxel Pattern Analysis (MVPA) to map the fine-grained neural overlap and associations within the OTC. Behavioral analyses established systematic associations stronger within animated and inanimate domains than across them. These behavioral associations were validated by the fMRI findings, which confirmed distinct neural overlap within these domains. The stroke patient data revealed deficit patterns that can be explained by the behavioral data and neural organization of the healthy brain. Furthermore, our data showed a differential age-related decline in visual recognition for word and face recognition, but not for house and animal recognition, demonstrating category-specific vulnerability. Our findings demonstrate that stroke-induced visual deficits follow structured patterns dictated by the fundamental organization of the OTC. The observed neural overlap between related categories explains why OTC lesions produce co-occurring deficits in associated categories. This research establishes a neuro-behavioral framework to interpret and anticipate visual recognition impairments. These empirically-derived principles can inform and guide the development of more effective and personalized neurorehabilitation protocols in the future.

Fonds Wetenschappelijk Onderzoek

TALK 5, 9:15 AM, 31.25

TEMPORAL DYNAMICS OF IT REPRESENTATIONS DEPEND ON IMAGE-MANIFOLD SCALE

Ammar I Marvi¹, Jacob S Prince¹, George A Alvarez^{1,2}, Talia Konkle^{1,2};
¹Harvard University, ²Kempner Institute

The time course of neural activity offers a window into the mechanisms of high-level visual processing. However, challenges posed by high-resolution brain measurement across space, time, and natural images have left many questions regarding the nature and time-scale of representational dynamics unanswered. We analyzed single-unit macaque electrophysiology (Triple-N dataset; Li et al, 2025) in 33 patches of inferotemporal (IT) cortex to explore how their representational geometry changed over time. Using time-resolved analyses, we measured representational dynamics at various image 'scales,' sampling image sets of increasing size based on overall response magnitude such that smaller sets contained the most preferred images. We reasoned that a larger, less selective set of images would span a larger portion of the natural image manifold to reveal global tuning dynamics, while smaller scales—using a subset of highly activating images—would reveal more granular, local changes to geometry. We find scale-dependent representational dynamics across IT cortex. With global image sets, geometry evolved smoothly and generalization via linear decoding remained relatively stable over time. In contrast, local sets yielded clearer temporal reconfiguration: representational geometry changed over ~100–250 msec post-image onset, e.g. one medial face patch showed a sharp transition between two distinct geometric configurations. More generally, the regional time-time geometry similarity matrices were more structured over local image sets compared to the full image set, with effective dimensionality (ED) compressed by up to 50%. Critically, these scale-linked dynamics were greatly reduced when we disrupted structure with scale-matched, rank-randomized controls (median ED change 0.6%). Together, our results suggest that IT geometry appears globally stable when measured over a diverse set of images yet locally dynamic for preferred inputs. This scale dependent view reconciles findings of temporal stability with recent findings of rapid tuning shifts and points toward dynamic coding mechanisms that locally refine visual representations over time.

This work was supported by funding from the Kempner Institute (to T.K.) and the Harvard Department of Psychology

TALK 6, 9:30 AM, 31.26

PROBING THE GRANULARITY OF HUMAN-MACHINE ALIGNMENT

Yash Mehta¹ (yashmehta3@jhu.edu), Raj Gauthaman, Michael Bonner;
¹Johns Hopkins University

Deep neural networks trained on fine-grained object classification (e.g., ImageNet 1000-way) are currently the leading models of the primate ventral visual stream. However, it remains unclear whether dense semantic supervision is a prerequisite for brain-model alignment, or if coarser categorical distinctions suffice. To address this,

we derived category boundaries directly from visual features, avoiding external semantic hierarchies. Using recursive PCA on pre-trained AlexNet and ViT representations, we generated hierarchical label sets ranging from 2 to 64 coarse categories while keeping the training images constant. We then trained networks from scratch on these coarse tasks and evaluated their representations against the Natural Scenes Dataset (fMRI) and the THINGS database (behavioral similarity). We found that fine-grained supervision is not necessary for brain-like representations. Strikingly, models trained to make only broad distinctions (e.g., just a handful of coarse categories) achieved the highest alignment with human behavioral similarity judgments, significantly outperforming standard 1000-class models. For neural alignment, models trained on intermediate granularity (32–64 categories) matched or surpassed fine-grained models in the early visual cortex and maintained comparable performance in the ventral stream. Additionally, coarse-trained models demonstrated superior stability when tested on out-of-distribution synthetic stimuli, suggesting that their alignment with human vision is more robust than conventional networks. Control analyses confirm these representations are structurally distinct from low-dimensional projections of fine-grained models. These findings suggest that networks trained on broader categorical distinctions are better at capturing the most cognitively salient organizing principles of human vision.

TALK SESSION: SUNDAY, MAY 17, 2026, 10:45 AM – 12:30 PM, TALK ROOM 1

Multisensory Processing

Moderator: Stephanie Badde, Tufts University

TALK 1, 10:45 AM, 32.11

DISRUPTED SENSORY REWEIGHTING FOR POSTURAL CONTROL IN GLAUCOMA

Rakie Cham^{1,2}, Galen Holland², Michelle Harter², Ian Conner², Mark Redfern¹; ¹Department of Bioengineering, University of Pittsburgh, ²Department of Ophthalmology, University of Pittsburgh

Glaucoma is associated with increased falls and loss of independence, but vision loss alone does not fully explain the elevated fall risk. Mobility impairments, particularly balance and gait deficits, likely reflect altered central sensory integration. A key candidate mechanism is abnormal sensory reweighting for postural control—how the nervous system adjusts its reliance on visual, vestibular, and somatosensory cues when some inputs become unreliable. Prior work shows that people with glaucoma sway more than controls during standing, especially when somatosensory input is degraded (e.g., standing on foam), but whether they appropriately upweight vision under these conditions is unknown. In this study, we directly tested a sensory-

reweighting mechanism in glaucoma. Twenty participants were enrolled in three groups: adults with glaucoma, older controls, and young controls. Participants completed an adapted Sensory Organization Test (SOT) during quiet standing on firm and foam surfaces with eyes open and eyes closed. The primary outcome was a normalized visual reliance metric: the ratio of center-of-pressure path length in eyes-closed versus eyes-open conditions, computed separately for firm and foam surfaces. All participants successfully completed the balance assessments. Consistent with classic sensory-reweighting theory, young and old controls increased their reliance on vision when somatosensory information was unreliable (foam). In contrast, adults with glaucoma did not increasingly rely on visual information to maintain balance on foam. These findings provide mechanistic evidence that glaucoma disrupts sensory reweighting for postural control, rather than simply reducing the amount of visual input. Identifying a central sensory integration deficit has direct clinical implications: it motivates development of targeted sensory retraining in occupational/physical therapy interventions. Such interventions should explicitly train more effective use of remaining visual and nonvisual cues to reduce falls and preserve independence in glaucoma.

The Henry L. Hillman Foundation

TALK 2, 11:00 AM, 32.12

MULTISENSORY AUDIOVISUAL CODING IN THE HUMAN HIPPOCAMPUS

Omri Raccach¹ (omri.raccach@yale.edu), Aryan Agarwal¹, Yannan Zhu¹, Nicholas Turk-Browne¹; ¹Yale University

The human hippocampus is traditionally considered a brain region dedicated to long-term memory. An emerging literature suggests that it may also contribute to several aspects of vision, including eye movements, mental imagery, scene discrimination, and visual search. However, the hippocampus receives anatomical inputs from virtually all sensory systems, raising the question of how it contributes to the processing of sensory modalities beyond vision and to the integration of visual information with other modalities. Using high-resolution fMRI (N = 30), we examined unisensory and multisensory representations of audiovisual movies across subfields and the longitudinal axis of the human hippocampus. Participants were exposed to multiple repetitions of short naturalistic movie clips, each presented in four formats: audio only, visual only, congruent audiovisual, and incongruent audiovisual (audio and video from different movies). Although univariate analyses detected only visual responses across hippocampal subfields, with no activation for auditory stimuli and no benefit of congruent multisensory stimulation, multivariate analyses revealed representations of both auditory and visual scenes. The posterior hippocampus showed enhanced pattern similarity for congruent stimuli relative to unisensory stimuli, demonstrating multisensory facilitation. Incongruent stimuli similarly showed reduced

pattern similarity relative to congruent stimuli, providing additional evidence of multisensory integration. The anterior hippocampus supported cross-modal decoding, suggesting a common representation of auditory and visual information. Finally, whole-brain searchlight analyses revealed parallel effects in cortical regions known to support multisensory integration. These findings advance our understanding of multisensory coding in the human hippocampus, including the discovery of a functional dissociation along its longitudinal axis, from distinct but interactive representations in posterior, to a shared amodal representation in anterior.

NIH F32 EY035941; NIH R01 MH069456

TALK 3, 11:15 AM, 32.13

TWO CORTICAL MECHANISMS FOR NATURAL AUDIOVISUAL PROCESSING

Subha Nawa Pushpita¹ (spushpit@andrew.cmu.edu), Leila Wehbe¹; ¹Carnegie Mellon University

Understanding how human brains process naturalistic audiovisual information remains a central challenge in cognitive neuroscience. Progress has been limited by difficulties of modeling complex audiovisual features - most prior work has relied on short, controlled stimuli or a single modality, leaving real-world comprehension mechanisms poorly characterized. Although recent advances in AI have enabled the extraction of high-quality features, how cortical regions dynamically process auditory and visual information as time unfolds remains largely unexplored. Using large-scale fMRI data collected while participants watched movies, we developed two novel computational approaches relying on prediction performance to map moment-by-moment sensory dynamics across cortex: one detects sustained periods when one modality predicts a region substantially better than the other, identifying regions that switch the modality they encode for meaningful stretches of time; the other isolates periods when both modalities predict well, revealing regions maintaining balanced encoding of both modalities. These approaches reveal two mechanisms of audiovisual processing: a pair of “bows” that switch modality—one posterior bow encircling category-selective visual cortex and another anterior bow spanning dorso-lateral frontal areas—and an arrow-like axis of bimodally predicted regions extending from lateral occipital into temporal cortex. We validate these findings in two ways. First, human raters confirm that scenes identified as audio or video-dominated by our method are ones in which they rely more heavily on the corresponding modality to understand the content. Second, we develop an audio-visual transformer that predicts fMRI responses, and its learned time-dependent attention-distribution over audio and visual inputs across regions matches those identified through our statistical frameworks. The coexistence of these systems suggests a cortical architecture that adaptively reweights sensory inputs while maintaining balanced multimodal representations,

supporting robust comprehension of complex natural events. More broadly, this work shows how naturalistic neuroimaging paired with modern machine learning can reveal new principles of dynamic audiovisual processing.

NSF CAREER grant 2237064

TALK 4, 11:30 AM, 32.14

IMPROVED TIME PERCEPTION FROM THE INTEGRATION OF VISUAL AND AUDITORY CUES

Anthony Bruno¹ (anthony_bruno@brown.edu), Jovan Kemp², Fulvio Domini¹, Leslie Welch¹; ¹Brown University, ²NYU Abu Dhabi

Humans integrate auditory and visual information to form a coherent perception of objects and events. Improvement from integrating two sources of information is greatest when performance between cues is comparable. In time perception, discrete stimuli like flashes and tones are often used to define time intervals, but discrimination performance is dramatically better in audition compared to vision (e.g., Burr, Banks, & Morrone, 2009). Prior attempts to measure cue integration relied on a weakened auditory cue (Burr et al., 2009; Hartcher-O'Brien, Di Luca, & Ernst, 2014). Equating the strength of auditory and visual cues can be challenging, which begs the question of whether discrete stimuli are appropriate for evaluating cue integration. Previously, we showed that a continuous bouncing stimulus improved visual time perception relative to a discrete flashing stimulus (Bruno & Welch, 2022). Performance with a bouncing stimulus was comparable to a continuous auditory stimulus, a tone that modulated in intensity to define a time interval. In the present experiment, participants completed a time interval discrimination task with continuous visual, auditory, and audiovisual stimuli. Performance with the audiovisual stimulus was significantly better than with either of the single-cue stimuli. To investigate the mechanism behind the integration, we compared audiovisual performance to two cue integration models. The models consider single-cue performance as the sum of sensory and memory noise. One model predicts that auditory and visual duration estimates are represented separately and then integrated, while the other predicts that auditory and visual event-markers are integrated from which a single duration estimate is produced (Hartcher-O'Brien et al., 2014). Performance with the audiovisual stimulus more closely resembled the prediction from the model that integrates the single-cue duration estimates. This suggests that the integration mechanism brings together separate visual and auditory duration estimates to result in improved discrimination performance.

TALK 5, 11:45 AM, 32.15

(MICRO-)SACCADIC EYE MOVEMENTS ALTER AUDITORY AND TACTILE TEMPORAL ORDER PERCEPTION

Stephanie Badde¹, Vanalata Bulusu¹; ¹Tufts University

Inferences about the causality of two events fundamentally rely on their temporal order. Causally linked events might be registered through different senses, suggesting that this distinct area of temporal perception is governed by a supramodal mechanism. To scrutinize this hypothesis, we tested whether oculomotor actions alter the perceived order of auditory and tactile events, as is the case for visual events. Participants judged the spatiotemporal order of pairs of auditory and tactile stimuli presented sequentially in different locations. Simultaneously, they either made saccades following a target stimulus that moved from one side of a distant screen to the other every four seconds (Expts. 1-3) or they fixated a stationary, central target that was presented alone (Expts. 4-5) or with the moving target, which's position they indicated verbally (Expts. 6-7). Participants tended to perceive the spatiotemporal order of auditory and tactile stimulus pairs presented around the onset of a saccade as reversed. This perceptual reversal of temporal order was clearly linked to the saccades: it did not occur for stimulus pairs presented at other times or for those combined with a verbal task. The saccadic suppression effect also emerged for tactile stimulus pairs that elicited apparent radial rather than horizontal motion. Moreover, when participants fixated a stationary target, auditory, tactile, and visual (Expt. 8) spatiotemporal perception was distorted around the onset of microsaccades—tiny, involuntary saccades that observers are typically not aware of—further ruling out cognitive variables linked to the guided saccades, such as perceptual and processing load, as a source of the effect. These novel effects of (micro-)saccades on the perceived order of auditory and tactile events demonstrate that action and perception are linked across modalities and reveal supramodal encoding of temporal order.

TALK 6, 12:00 PM, 32.16

DISSOCIATING PREDICTIVE AND POSTDICTIVE AUDIOVISUAL INFERENCE

Manda Fischer^{1,2} (manda.fischer@utoronto.ca), Keisuke Fukuda^{1,2}; ¹University of Toronto, ²University of Toronto Mississauga

Our brains have the remarkable ability to use contextual information to resolve perceptual uncertainty. This reliance on context has been demonstrated when information is presented before a stimulus (supporting predictive inference) and after it (supporting postdictive inference). However, it remains unclear whether these two forms of inference rely on overlapping or distinct mechanisms. We addressed this question using an audiovisual working memory task designed to test how category-level auditory cues influence visual face perception in a predictive and postdictive manner. Each trial participants (N=84) briefly viewed a face morphed along a female–male continuum (100 ms) and later reconstructed it after a short retention interval (1900 ms) using a continuous morph slider. To manipulate auditory context, either a male or female voice was presented 1000 ms before (pre-cue) or after (post-cue) face onset. Mixed-effects modeling revealed that

the voice cue biased face reconstructions toward the gender of the voice in both pre- and post-cue conditions, but in distinct ways. Pre-cue presentation produced a uniform shift toward the cued gender across the continuum (cue main effect), reflecting a predictive bias. Post-cue presentation produced a gender-dependent modulation, strongest for ambiguous/moderately gendered faces (cue × face interaction), reflecting a postdictive reinterpretation of the visual input. Significant cue effects in both pre- and post-cue conditions clustered near ambiguous faces, highlighting the range where perception is most malleable. Cue effects were uncorrelated across individuals, suggesting that predictive and postdictive mechanisms are dissociable. Moreover, the influence of the auditory cue was largest when the visual face signal contributed less to the final percept, reflecting a trade-off between auditory and visual sources of information in shaping perception. Taken together, our results suggest that distinct mechanisms underlie predictive and postdictive inference, each dynamically leveraging auditory context to disambiguate visual input, especially when the fidelity of this input is low.

We thank Kayla Vasquez for their help with data collection.

TALK 7, 12:15 PM, 32.17 - REMOTE PRESENTATION

BIOLOGICAL MOTION AS A MULTISENSORY SIGNAL: PREDICTIVE INTEGRATION OF SPACE AND TIME

Melissa Nur Robinson¹, Ataol Burak Ozsu^{1,2}, Andreas Treske¹, Ufuk Onen¹, Burcu A. Urgan¹; ¹Bilkent University, ²University College London

The brain receives input from sensory organs that is often incomplete and noisy, making prioritization and integration of key features across modalities a non-trivial process supporting adaptive behavior. Biological motion is a particularly salient signal, robustly processed across the lifespan due to its survival relevance and social significance. While visual processing of biological motion has been extensively studied, its multimodal nature has received comparatively little attention. Prior multisensory studies have largely focused on temporal coupling between visual and auditory cues, yet in naturalistic contexts, audio-visual signals are also spatially aligned. Here, we investigated how temporal and spatial coupling interact to shape biological motion perception using a novel stimulus set with spatially dynamic auditory footstep sequences. We conducted two experiments to test whether biological motion selectively enhances temporal binding and whether this effect is modulated by spatial congruency between auditory stimuli and visual motion. In Experiment 1, participants viewed upright or scrambled point-light walkers paired with temporally synchronous or asynchronous footsteps and judged whether the trials were synchronous. In Experiment 2, we extended this design by additionally manipulating spatial congruency between the walker's direction and the auditory motion trajectory, where the auditory motion trajectory either matched the walker direction (congruent) or did not match (incongruent). Our results indicate that

both temporal and spatial alignment influence synchrony judgments between modalities, and critically, that spatial coupling enhances synchrony detection when auditory spatial information is misaligned. These findings highlight the importance of considering multiple dimensions of audiovisual integration in biological motion perception and provide new insights into how the brain flexibly integrates multisensory information.

TALK SESSION: SUNDAY, MAY 17, 2026, 10:45 AM – 12:30 PM, TALK ROOM 2

VISUAL WORKING MEMORY: SPACE, FEATURES
Moderator: Freek van Ede, Vrije Universiteit Amsterdam

TALK 1, 10:45 AM, 32.21

SUPPRESSION OF PASSIVE MEMORY REFLECTS PROACTIVE AND CONTEXT-DRIVEN CONTROL IN VISUAL WORKING MEMORY

Ziyuan Li¹ (zyli1213@163.com), Mengyao Sun¹; ¹Anyang Normal University

Hierarchical models of visual working memory (VWM) propose that information can be retained in multiple functional states, with currently relevant representations maintained in an active state and prospectively needed representations stored in a passive state. Recent work shows that passive memories undergo sustained suppression, a mechanism that stabilizes passive maintenance and prevents interference with concurrent active processing. However, it remains unclear whether such suppression is obligatorily triggered by the actual demands of a secondary task or reflects an expectation-driven cognitive strategy. To address this question, we designed two tasks performed in a counterbalanced order while recording EEG: (1) a baseline task where memory items (array 1) were always relevant and maintained in the active state until retrieval; and (2) a probability task where array 1 was followed either by a blank retention interval (30%) or by frequent presentation of a second array (array 2; 70%). In this probability task, array 1 became prospectively relevant and was expected to transition to a passive state when array 2 required active maintenance. When participants firstly habituated to active maintenance in the baseline block, suppression of array 1 emerged only when array 2 was actually presented, indicating that suppression of passive maintenance was driven by actual processing demands. In contrast, when the baseline mindset was removed, participants adopted a proactive strategy of switching array 1 into a passive state regardless of whether array 2 appeared; this suppression was flexibly withdrawn once participants has no anticipation of additional processing. Behaviorally, switching memory state incurred performance cost. Together, these findings demonstrate that the suppression of passive memory is not purely stimulus-driven, but

reflects an expectation-guided mindset preparatory for anticipated processing demands. This study reveals that state transitions in VWM are dynamically governed by learned task structure, underscoring the proactive and context-sensitive nature of human memory control.

National Natural Science Foundation of China

TALK 2, 11:00 AM, 32.22

THE CAUSAL ROLES OF VISUAL AND PREFRONTAL CORTEX IN VISUAL WORKING MEMORY PRIORITIZATION: A PHOSPHENE- AND MRI-GUIDED RTMS-EEG STUDY

Chenlingxi Xu¹ (cxu5@nd.edu), Nathan Rose²; ¹University of Notre Dame, ²Northeastern University

While the frontal cortex is traditionally viewed as the hub for working memory (WM) maintenance, recent theories predict that sensory cortex [early visual cortex (VC)] assists in the maintenance of unattended memory items via “activity-silent” short-term synaptic plasticity. Prior studies have implicated interregion communication between both dorsolateral prefrontal cortex (dlPFC) and VC during WM maintenance. Here we propose a functional division of labor: dlPFC controls attentional priority and interference resolution whereas VC maintains high-fidelity sensory traces. Causal evidence dissociating these contributions during dynamic prioritization is limited. Sixty-eight healthy younger adults performed a two-item, double-retrocue, double-recall orientation visual WM task with concurrent EEG recording before and after receiving offline active or sham rTMS (40-s) at either theta-burst (cTBS) or 10 Hz (alpha) stimulation to MRI-guided left dlPFC or phosphene-localized left VC. Behavioral performance was quantified using absolute recall errors and computational modeling (precision, swap errors, guessing). Our pre-registered predictions are that, relative to sham, 10 Hz-rTMS to dlPFC will increase swap error rates, while cTBS to VC will decrease the precision of orientation recall. Preliminary analyses of the active stimulation group reveal emerging site-specific patterns. Following dlPFC stimulation, participants showed general performance gains across trial types. In contrast, VC stimulation showed hemifield-dependent effects: improvements were attenuated when the retrocued item appeared in the TMS-targeted hemifield. Furthermore, a site-by-frequency interaction was observed for precision during attention switching: 10-Hz-stimulation decreased precision when applied to dlPFC but increased precision when applied to VC. While active stimulation drove overall gains rather than predicted impairment, the dissociation between general (dlPFC) and hemifield-specific (VC) patterns were consistent with the hypothesized division of labor. Planned EEG analyses will further evaluate whether dlPFC stimulation alters control-related signals (e.g., N2pc, P3, midfrontal theta) and VC stimulation alters sensory-maintenance signals (e.g., N1/P1, CDA,

posterior alpha), providing potentially converging mechanistic evidence.

TALK 3, 11:15 AM, 32.23

CONTRALATERAL DELAY ACTIVITY AND ALPHA LATERALISATION DURING WORKING MEMORY ARE CONTINGENT ON THE TRANSITION FROM SENSATION TO WORKING MEMORY

Baiwei Liu¹, Freek van Ede¹; ¹Vrije Universiteit Amsterdam

Delay-period EEG signals, such as the Contralateral Delay Activity (CDA) or the lateralisation of 8-12 Hz posterior alpha activity, are widely viewed and used as direct markers of selective working-memory maintenance. However, these EEG markers are conventionally studied in pre-cue paradigms where relevant memory content must first be selected at sensory encoding. This potentially conflates the selective maintenance of relevant information with the active translation of this information from sensation into memory. To disentangle selective sensory-to-memory transition from selective memory maintenance, we leveraged a novel task where we could compare situations that were otherwise identical, but where in one case relevant memory content had to first be selected from perception, while in the other case this content could be selected from working-memory directly. In both cases, the selected visual content and the ensuing working-memory task were identical, but only the former engaged a sensory-to-memory transition. Strikingly, our results show that sustained delay activity was contingent on the availability of sensory input: both the CDA and prolonged alpha lateralisation were robust in the 'select at encoding' condition but absent in the 'select from working memory' condition. Crucially, this was not due to a lack of spatial organisation in the latter 'select from working-memory' condition, as we observed robust and comparable N2pc and initial alpha lateralization in both conditions. This result was further supported with multi-variate pattern analysis. Together, our results show how prolonged "delay activity" EEG signatures in posterior electrodes are surprisingly contingent on sensory input, and suggest that such signatures may track the transition from sensory inputs to working memory traces, rather than the pure maintenance of relevant information per se.

TALK 4, 11:30 AM, 32.24

ELECTROENCEPHALOGRAPH DECODING SUGGESTS SEPARATE INDEXING MECHANISMS FOR ATTENTIONAL TRACKING OF FEATURELESS OBJECTS AND WORKING MEMORY STORAGE

Henry Jones¹ (henryjones@uchicago.edu), Dawei Bai², Brian Scholl², Edward Awh¹; ¹University of Chicago, ²Yale University

Recently, multivariate decoding of EEG data has identified a signal that scales with the number of items in working memory (WM), regardless of the specific content being maintained or the number of spatially attended locations. One possibility is that this signal reflects an abstract indexing process that binds the content of items to their context in space and time for maintenance and accessibility. To explore this possibility, we examined whether a similar load signal exists for "featureless objects", which can only be described by their coordinates in space and time. On each trial, participants viewed a dense grid of crosses of random orientations. A moving object was implemented by having a single cross change from one random orientation to another, with changes occurring between adjacent crosses across frames. These transients yielded a persisting trackable object, even though (a) there is no constant surface feature across frames, and (b) it is impossible to identify objects on any static frame. In 2 EEG datasets, participants completed a task in which they tracked either 1 or 2 cued featureless objects, and a WM task in which they remembered either colors or the shape and location of "dot cloud" stimuli while spatial attention was controlled for. In both experiments, EEG decoding found a stable signal that scaled with the number of tracked featureless objects. However, we found no consistent evidence that the tracking load signal and the WM load signal generalized to one another. Although planned studies will examine whether this conclusion generalizes to traditional tracking tasks that may rely on other mechanisms (Lu and Sperling, 1996), these results suggest that distinct mechanisms may guide WM storage and attentional tracking.

TALK 5, 11:45 AM, 32.25

SHIFTS OF SPATIAL ATTENTION WITHIN WORKING MEMORY OCCUR FASTER THAN FEATURE RE-PRIORITIZATION.

Caitlin V. Cunningham¹, Julie D. Golomb¹; ¹The Ohio State University

In our complex visual world, spatial attention is used to boost processing for objects at an attended location. In addition to operating on items in the external environment, spatial attention can be directed internally, toward items within working memory. Spatial attention is frequently reoriented to keep up with the dynamic environment, but these attentional shifts are imperfect: e.g., shifts of externally-directed spatial attention are known to result in errors in feature perception. Here, we seek to track the dynamics and consequences of shifting internal attention. In Experiment 1, participants encoded an array of memory items (colored squares), then were peripherally cued to orient internal attention toward one item's location. On some trials, a second postcue subsequently cued participants to shift spatial attention to a different item within working memory. After a short (50ms) or long (500ms) delay, participants reported the color and location of the most recently cued item (target). Probabilistic mixture modeling revealed that when given 50ms to shift attention, participants almost always

reported the correct location of the target item, but sometimes made feature errors: specifically, they tended to misreport the color of the item opposite the first-cued item in color space, characteristic of “feature avoidance errors.” These errors indicate that participants were aware they were supposed to report a different item than initially cued, but couldn’t yet access the feature, so strategically guessed a color maximally different from the color still “stuck” in memory. We replicated this pattern in Experiment 2, which was identical to Experiment 1 but with endogenous postcues presented at fixation instead of peripherally, revealing that these errors are not driven by eye movements, but indeed associated with failures to shift covert internal attention. Together, these results suggest that shifting of internal spatial attention occurs more quickly than re-prioritization of features in working memory.

NIH R01-EY025648 (JG)

TALK 6, 12:00 PM, 32.26

VISUAL WORKING MEMORY ACROSS SACCADDES: EVIDENCE FOR INDEPENDENT PRE- AND POST- SACCADIC REPRESENTATIONS

Zexuan Niu¹, Andrew Hollingworth¹; ¹UNIVERSITY OF IOWA

Before the saccade to an object, perceptual information is acquired peripherally and is maintained in visual working memory across the saccade. Post-saccadic information is then encoded, typically from the now-foveated object. How do the pre- and post-saccadic representations interact in object perception? Recent work suggests that the two representations are integrated into a single, composite representation (the Integration Hypothesis), supported by evidence that mean feature report lies at an intermediate value between pre- and post-saccadic values. A plausible alternative, however, is that each representation is maintained independently (the Independence Hypothesis), with participants probabilistically reporting either the pre- or post-saccadic value. In our first experiment, participants made a saccade to a colored disk, whose color changed transsaccadically. They then reported the pre-saccadic color. The integration hypothesis predicts a univariate distribution of color report, centered between the pre- and post-saccadic values. In contrast, the independence hypothesis predicts either 1) a bivariate distribution, with means fixed at the pre- and post-saccadic values or 2) a univariate distribution centered at either the pre- or post-saccadic value. In our data, the independence model generally provided a better fit to the observed distribution of color report than did the integration model, although individual differences emerged with larger color change. In a second experiment, participants indicated whether they perceived one or two colors, and the analyses were restricted to single-color trials to minimize the influence of explicit change detection or strategic inference. Here, the independence model consistently provided a better fit than did the integration model. Typically, participants either

retained both pre- and post-saccadic information (i.e., a bivariate response distribution) or post-saccadic input overwrote pre-saccadic information (i.e., a univariate response distribution centered at the post-saccadic value). Together, the data suggest that pre- and post-saccadic values are not directly integrated across saccades and are instead maintained independently.

TALK 7, 12:15 PM, 32.27

MOTORIC MECHANISMS THAT BIAS GAZE PRODUCE REFLEXIVE SELECTION IN MEMORIZED VISUAL SPACE

Priyanka Gupta¹, Devarajan Sridharan¹; ¹Indian Institute of Science, Bangalore

Although the shared neural mechanisms of eye-movements and covert attention are well-studied, the link between eye-movements and working memory (WM) remains tentative. Whether and how gaze mechanisms affect selection in visual WM is actively researched. Here, we present a cognitive mechanism that biases gaze and produces reflexive (automatic) selection in memorized visual space. We tested the effect of “saccadic tagging” – cueing a location for a subsequent saccade – on spatial selection in WM. In one dual-task, participants (n=24) executed cued saccades to one of four locations (Go), while also maintaining in WM orientations of gratings presented at each of those locations. On 40% of trials, participants were strongly incentivized to plan saccades but refrained from executing them (No-go). Saccadic tagging produced robust spatial selection in WM, quantified by the mean absolute error (MAE) of orientation recall (cued: $29.39^\circ \pm 1.51^\circ$, uncued: $32.49^\circ \pm 1.27^\circ$, $p < 0.001$). We note three remarkable aspects of this selection. First, spatial selection occurred for the tagged location regardless of whether the saccade was executed (Go) or merely planned (No-go) toward it. Second, selection occurred despite each of the four locations being equally task-relevant (equal recall probe validity). Finally, nearly identical selection occurred also in a manual response task designed to avoid systematic spatial correspondences between the cue and button locations. In sum, modality-generic motoric preparation sufficed to produce robust, reflexive selection in memorized visual space. We sought putative physiological underpinnings of this mnemonic selection. Microsaccades – small, fixational eye-movements – were strongly biased toward the tagged location ($p < 0.001$), across both modalities. Moreover, the direction of the first microsaccade post-cue predicted the extent of reflexive selection in WM, irrespective of motor modality. Our results link motoric mechanisms that bias gaze to those that produce selection in visual WM. The findings may be relevant for diagnosing and treating disorders of WM.

Talk Session: Sunday, May 17, 2026,

2:30 – 4:15 pm, Talk Room 1

Development

Moderator: Cameron Ellis, Stanford University

TALK 1, 2:30 PM, 34.11

DEVELOPMENTAL CHANGES IN VISUAL CLUTTER IN INFANTS' AND TODDLERS' VISUAL WORLDS

Minju Kim¹ (minjukim023@gmail.com), Linda B. Smith¹; ¹Indiana University Bloomington

The everyday visual world is messy, noisy, and cluttered. Clutter is well-known to limit information extraction from a scene. Nonetheless, during the first two years of life, infants rapidly acquire visual categories of faces, objects, and scenes. Understanding the properties of clutter in the everyday environment where visual development occurs is essential to a comprehensive understanding of visual learning and category formation. Here, we quantify visual clutter in infants' daily experiences and its temporal structure. We analyzed 500,000 egocentric images from 40 infants across five age groups (1-3, 6-8, 10-12, 15-20, 22-26 months). Recordings captured naturalistic indoor and outdoor contexts, providing a dense sample of the scenes in front of infants' faces. We used four converging measures: Feature Congestion as defined by Rosenholtz, the number of visual segments, semantic category diversity, and foregrounding. Distributions were highly skewed; therefore, we discretized scores into seven ordered categories based on percentiles ranging from Extremely Low to Extremely High. Temporal structure was examined by analyzing bouts of consecutive frames within the same clutter category. We observed evidence of a systematic restructuring of clutter, with convergent patterns across all four measures. For example, the Feature Congestion measure revealed that the youngest infants experienced the highest frequency of sustained, extremely low-clutter periods. In contrast, these long "pockets of simplicity" declined dramatically across development, becoming rare in the oldest age group. Concurrently, high-clutter exposures increased in both frequency and duration. Critically, this developmental trajectory—from sustained periods of uncluttered input to more frequent high-clutter input—was observed across measures, suggesting a systematic reorganization of the visual environment. The nature of the attentional tasks young perceivers need to solve changes with development. We will discuss the roles of infant visual development, early "sticky" attention, and foregrounded entities (faces and objects) as contributors to these developmental patterns.

TALK 2, 2:45 PM, 34.12

DO YOU SEE WHAT I SEE? INVESTIGATING THE DEVELOPMENT OF VISUAL DIET IN CHILDREN AND TEENS USING MOBILE EYE-TRACKING IN NATURAL ENVIRONMENTS

Elizabeth Jiwon Im^{1,2} (ejim@stanford.edu), Kalanit Grill-Spector^{1,2}; ¹Stanford University, ²Stanford Wu Tsai Neuroscience Institute

Face recognition and word reading abilities improve during childhood and adolescence, in correlation with increases in selectivity for faces and words in the ventral temporal cortex. At the same time, selectivity for hands decreases. We hypothesize that these developments may be linked to changes in children's visual diet across development. However, to date there has been no investigation of children's and teenagers' naturalistic visual diet, and how it develops remains unknown. For the first time, we investigate children's (4-6yrs) and teenagers' (14-17yrs) visual diet in natural environments (home: n=10; classroom: n=6) using mobile eye-tracking glasses (FOV=103°×77°). We introduce a novel validation procedure for natural viewing, estimating a measurement accuracy of $1.01^\circ \pm 1.69^\circ$ in children (n=10) and $1.10^\circ \pm 1.52^\circ$ in adults (n=10). We then leverage computer vision models to analyze egocentric scenes and gaze behavior. Our data show that the prevalence of faces and hands is similar between children and teenagers. However, fixation behaviors are different. While both age groups foveate more on faces than hands, there is a significantly higher proportion of faces in the central 5° of teenagers' visual diet than that of children's (14.9%±7 in teenagers vs 6.1%±2.3 in children, p=0.032, permutation test). Moreover, faces more often appear in the upper visual field for children (median y-position: $19.4^\circ \pm 5.2^\circ$), while there is no vertical bias for faces in teenagers. In contrast, hands trend towards being in the lower visual field for teenagers (median y-position: $-9.27^\circ \pm 10.89^\circ$), while there is no vertical bias for hands in children. Together, our findings suggest that, despite shared environments, children and teenagers fixate and interact with the world differently, with visual field biases continuing to develop from early childhood into adolescence.

This work was supported with funds from NIH R01 EY 023915 awarded to KGS

TALK 3, 3:00 PM, 34.13

COMPROMISED GESTALT GROUPING IN LATE-SIGHTED CHILDREN

Lukas Vogelsang¹, Marin Vogelsang¹, Priti Gupta², Pragya Shah², Jeonghwan Cheon³, Purva Sethi², Stutee Narang², Suma Ganesh⁴, Pawan Sinha¹; ¹Massachusetts Institute of Technology, ²Project Prakash, Dr. Shroff's Charity Eye Hospital, ³Independent Researcher, ⁴Pediatric Ophthalmology, Dr Shroff's Charity Eye Hospital

Holistic perceptual organization and Gestalt-like grouping are critical for making sense of the complex visual environment. Past research has investigated the developmental timelines for acquiring these foundational capabilities, highlighting their importance in normal visual maturation. Here, we examined the consequences of curtailing early visual experience on perceptual grouping skills later in life. Specifically, we studied a unique population of visual observers – children from rural India who were born blind but gained sight through surgical intervention late in childhood. We adapted a paradigm previously used with normally-sighted infants. Participants viewed carefully designed square arrays of elements arranged in a uniform grid, where grouping was induced through similarity in luminance or figural form along alternating rows or columns, and reported whether they perceived vertical or horizontal organization. Longitudinal assessments of 16 patients from pre-operative to many months post-operative status revealed persistently poor performance. Despite notable gains in other visual abilities, participants exhibited no improvement over time, remaining substantially below matched controls. This deficit was consistent across all grouping cues. Complementary computational simulations suggest a candidate account for these findings: the deficits may arise from commencing visual experience with higher-fidelity input than is available to neonates. Deep networks trained on developmentally inspired degraded-to-undegraded input trajectories formed representations that more strongly encoded holistic organization in Gestalt stimuli. In contrast, networks trained without this developmental progression, akin to the experience of the late-sighted, failed to develop such organization. Together, these results provide evidence that early-onset blindness yields lasting impairments in Gestalt-like grouping, and suggest that not only the timing but also the degraded quality of early visual experience may play an important role in the emergence of holistic organization. Besides helping formulate clinical prognoses for late sight treatments, these findings have implications for understanding typical development and the design of developmentally-aligned computational training procedures.

NIH R01EY020517

TALK 4, 3:15 PM, 34.14

EARLY DEPRIVATION SELECTIVELY IMPAIRS CARDINAL ORIENTATION SENSITIVITY AND REDUCES THE OBLIQUE EFFECT

Md. Salman Sarkar¹ (optomsarkar@gmail.com), Bat-Sheva Hadad¹, Subiksha Saravanan², Shashikant Shetty², Amit Yashar¹; ¹University of Haifa, Israel, ²Aravind Eye Hospital, India

Background: Visual perception varies as a function of environmental regularities, demonstrating greater sensitivity to prevalent stimuli. In orientation processing, the visual system exhibits a well-known asymmetry in orientation sensitivity known as the oblique effect—

greater perceptual sensitivity to cardinal (horizontal and vertical) orientations than to oblique ones. However, whether and how the visual system learns basic environmental regularities during early development—and how monocular or complete deprivation affects this process—remains unknown. Objective: We address this question by examining orientation sensitivity in individuals with a history of early visual deprivation due to dense cataract, either in both eyes (bilateral) or one eye (unilateral), and either from birth (congenital cataract) or acquired later in development (developmental cataract). Method: Participants aged 6–34 years (congenital-bilateral: $n = 43$; congenital-unilateral: $n = 15$; developmental-unilateral: $n = 46$; typically sighted: $n = 61$) completed a two-interval forced-choice (2IFC) orientation discrimination task across cardinal (0° and 90°) and oblique (45°) orientations. Thresholds were measured using a staircase procedure. Visual acuity and age were controlled for in the analysis. Results: Participants with early visual deprivation, either in both eyes or in one eye, showed a reduced standardized oblique effect, driven by poorer discrimination of cardinal orientations. The results remained consistent even after controlling for participants' age and visual acuity as covariates. Finally, participants' age showed a positive correlation specifically with cardinal orientation sensitivity in the typically sighted group. Conclusions: The results reveal that the visual system learns basic environmental regularities during development. Visual deprivation, in one or both eyes, is detrimental to the learning of these regularities. These findings suggest that, from birth and throughout development, the visual system adjusts its sensitivity based on the statistics of the environment.

Israel Science Foundation

TALK 5, 3:30 PM, 34.15

THE ROLE OF VISION IN INFANTS' MULTISENSORY LOCALIZATION AND BODY REMAPPING

Monica Gori^{1,2}, Helene Vitale², Andrew Bremner³, Caudio Campus²; ¹IHMC, USA and IIT, Italy, ²IIT, Italy, ³University of Birmingham, UK

Background. The mechanisms underlying the development of multisensory integration in sensory deprivation remain unclear. In recent years, our work has highlighted the importance of sensory interactions in scaffolding multisensory development. We have shown that vision plays a critical role in shaping both auditory and tactile integration. In the absence of vision, auditory spatial representations and body representations are altered, making audio-tactile integration more challenging for blind individuals compared to sighted peers. Methods. We examined behavioral responses and event-related potentials (ERPs) in 12 sighted (S) and 12 severely visually impaired (SVI) infants. Participants were presented with unisensory (auditory, tactile) and multisensory (congruent, incongruent) stimuli delivered to their hands in both uncrossed and crossed positions. Neural activity was recorded using a high-density 128-channel EGI system. Results and Discussion. Together, the unisensory results suggest that SVI

infants remain anchored to body-centered coordinates in early development, whereas sighted infants shift toward an external reference frame, revealing a two-step tactile processing mechanism. Indeed, SVI infants showed heightened contralateral somatosensory responses in an early time window, which correlated exclusively with their behavioral reactions and showed no evidence of tactile remapping. In contrast, sighted infants exhibited a later ipsilateral somatosensory response aligned with their behavioral performance. This highlights vision's specific contribution to developing somatosensory remapping from early life. Regarding multisensory processing, congruent stimuli elicited activation of integrative centro-parietal (CP) and fronto-central (FC) regions within 180–220 ms in both groups, indicating preserved multisensory integration. However, in an earlier window (105–120 ms), sighted infants showed greater temporo-parietal activation (associated with auditory processing), whereas SVI infants showed stronger centro-parietal activation (associated with somatosensory processing). Consistent with our previous behavioral work (Gori et al., *Curr.Biol.* 2021), these results support the hypothesis that visual deprivation shapes early multisensory spatial representations.

Acknowledgement: This work is funded by EU H2020, ERC StG MySpace (No. 948349).

TALK 6, 3:45 PM, 34.16

ECCENTRICITY BIAS IN THE INFANT HOMOLOG OF ADULT SCENE-SELECTIVE CORTEX

Sarah S. Tung¹ (sstung@stanford.edu), Emily M. Chen¹, Jason D. Yeatman², Cameron T. Ellis¹; ¹Department of Psychology, Stanford University, Stanford, CA, USA., ²Graduate School of Education, Stanford University, Stanford, CA, USA.

A central question in vision science is how human visual cortex becomes organized into distinct regions that selectively respond to content such as faces, scenes, and words. One hypothesis proposes that selectivity is scaffolded from 'protomaps'—biases for features, like viewing location and spatial frequency, that covary with stimulus category. By hypothesis, protomaps are present at birth and shaped by experience, though their initial feature biases may persist. Indeed, in adults, category-selective regions are biased towards parts of the visual field where those categories typically appear. For example, scene-selective regions are more sensitive to peripherally presented stimuli than foveal stimuli. To test the protomap hypothesis, we conducted fMRI with awake behaving infants (age: 2-6 months, N=7) and collected independent measurements of category selectivity and eccentricity biases. Infants viewed blocks of scenes and words presented either foveally (6°) or peripherally (40°) in the visual field. We found evidence of eccentricity organization in early visual cortex in both infants and a validation sample of adults (N=7). In scene-selective visual cortex (anatomically defined in the collateral sulcus), we find scene-selective responses in adults but not in infants. Yet, consistent

with the protomap hypothesis, infants (and adults) did show a significant eccentricity bias, responding more strongly to peripheral than foveal stimuli. Even voxels with the strongest evidence of scene-selectivity showed a significant eccentricity bias, suggesting that category selectivity and eccentricity are linked in infancy. Altogether, these results are consistent with the hypothesis that eccentricity serves as a scaffold for category-selectivity to emerge.

TALK 7, 4:00 PM, 34.17

A MULTIVARIATE APPROACH TO UNDERSTANDING THE EMERGENCE OF THE VISUAL WORD FORM AREA

Kelly J. Hiersche¹ (hiersche.1@buckeyemail.osu.edu), Lauren A. Rydel¹, Zeynep M. Saygin¹; ¹The Ohio State University

The visual word form area (VWFA) is unique among ventral visual areas in that it shows protracted development, emerging only after a child can read. Prior work demonstrated connectivity to high-level language regions informs the eventual location of the VWFA, but the functional profile of the proto-VWFA is still unclear. Univariate word selectivity may emerge from initial multivariate representations that differentiate between visual words and other visual categories (e.g., faces, objects) or between linguistic vs. non-linguistic auditory stimuli. We scanned 111 children ages 3-12 years (24 longitudinal, including prereaders who transitioned into readers) on two runs of a high-level visual fMRI task and a subset on two runs of an auditory language task. We used functional regions of interest, multivariate representation analyses, and representational similarity analyses to understand the functional profile of emerging VWFA. We find that while pre-readers lack univariate word responses, they do show multivariate representations for written words. Representational similarity analyses showed that activation to words is dissimilar from other stimulus categories in the pre-VWFA and that this dissimilarity is almost identical to that of the VWFA in readers. The proto-VWFA in prereaders did not show multivariate (or univariate) representations for auditory language, much like results in readers. Finally, our longitudinal data supports these results, showing the proto-VWFA holds a multivariate representation for words, but not for auditory language, before the univariate VWFA emerges. Overall, our results suggest that multivariate representations may scaffold univariate selectivity of the VWFA as a child learns to read; interestingly, these do not seem to emerge from auditory representations to language, suggesting that any responses of this region to auditory language in older children or adults are likely a result of experience and challenge the theory that the proto-VWFA is part of the distributed language network.

Alfred P. Sloan FG-2018-10994, OSU College of Arts and Sciences, Women in Philanthropy Award, Chronic Brain Injury Pilot Award, R01 HD110401, (all to ZMS), NSF-GRFP DGE-1343012 (to K.J.H.)

**TALK SESSION: SUNDAY, MAY 17, 2026,
2:30 – 4:15 PM, TALK ROOM 2**

**Face and Body Perception: Emotion,
identity**

Moderator: Nancy Kanwisher, MIT

TALK 1, 2:30 PM, 34.21

**MACAQUES IN ACTION: REALISTIC INDIVIDUALIZED 3D
MONKEY AVATARS WITH NATURALISTIC MOTION**

Lucas Martini¹ (lucas.martini@uni-tuebingen.de), Alexander Lappe¹, Anna Bognár², Rufin Vogels², Martin A. Giese¹; ¹University of Tuebingen, ²KU Leuven

Realistic and well-controlled synthetic stimuli have been crucial for studying the mechanisms of face perception in monkeys. To investigate the neurocomputational processes underlying body perception, an equivalent level of stimulus control is desirable, especially for dynamic bodies. However, this goal is challenged by the complexity of recording realistic movements without markers and modeling dynamic 3D shape deformations of bodies in these animals. **METHODS:** We recorded eight rhesus macaques using 16 high-resolution color cameras to reconstruct their 3D appearance via surface-based avatar models. These models include internal skeletal structures that drive surface deformations, allowing us to align synthetic, computer-generated macaques with real video recordings using keypoint predictions and temporal masks in an automated fashion. Using labeled data, we further generated subject-specific avatars that more accurately capture each individual's appearance and movement dynamics. **RESULTS:** Our dataset contains over 750 macaque actions and interactions, each annotated with ethogram-based action labels. The alignment of our surface models shows qualitative and quantitative improvements over established methods, based on different quality measures of 3D fidelity. This reconstruction pipeline is also designed to make large-scale 3D surface modeling tractable. To assess whether the reconstructions capture critical information about body shape, we evaluated action recognition performance across a wide range of visual feature encodings, with and without our 3D pose descriptions. Across our benchmark, adding 3D pose features substantially improves action classification in mean average precision, including action categories ranging from locomotion to social interaction. **CONCLUSION:** We present a large-scale 3D motion-capture dataset of eight rhesus macaques and corresponding subject-specific synthetic avatar models. These avatars can be posed arbitrarily, include color appearance, and pave the way for fully controlled, experimental studies of body perception in nonhuman primates.

European Research Council (2019-SyG-RELEVANCE-856495). International Max Planck Research School for Intelligent Systems (IMPRS-IS).

TALK 2, 2:45 PM, 34.22

**FAMILIARITY AND TRAIT EMPATHY MODULATE
PRIMARY SOMATOSENSORY RESPONSES TO
VICARIOUS THREAT**

Naama Zur¹ (nrz9@georgetown.edu), Minwoo Lee¹, Maria Czarnicka¹, Jingrun Lin², James Coan², Casey Kenyon Brown¹; ¹Georgetown University, ²University of Virginia

Vicarious threat refers to experiencing another's threat as if it were one's own. Simulation theories propose that somatotopically matched activation in primary somatosensory cortex (SI BA3b) plays a crucial role in representing another's bodily threat experience. However, whether SI BA3b also contributes to perceivers' own affective responses to vicarious threat remains unclear. Addressing this gap, we used functional magnetic resonance imaging (fMRI) to test whether SI BA3b shows increased activation during vicarious threat and whether this response is associated with factors known to amplify vicarious emotional experience: self-reported affective trait empathy and familiarity with the target. We hypothesized that SI BA3b would show greater activation for threat than for safety, track affective ratings, be stronger for familiar partners than for strangers, and scale with trait affective empathy. Ninety-seven participants (ages 23–26) underwent fMRI while holding hands with either a familiar partner or a stranger in separate runs. In each run, safety or threat cues indicated whether the other individual would receive an electric shock. Contrast estimates for [threat > safe] in SI BA3b were extracted (small-volume-corrected $p < .05$) and analyzed in relation to trait empathy and self-reported affective ratings. Contrary to predictions, SI BA3b showed decreased activation during threat trials compared with safety trials, regardless of target familiarity. Intriguingly, during the partner condition, but not the stranger condition, this deactivation was less pronounced among individuals with higher trait affective empathy. Affective ratings did not correlate with SI BA3b activity. Together, these findings suggest that SI activation may not directly encode perceivers' affective responses to vicarious threat. Instead, SI activation appears to be modulated by familiarity and emotional trait empathy, indicating a more nuanced role for SI in vicarious emotional experiences of threat.

TALK 3, 3:00 PM, 34.23

**THE STS RESPONDS PREFERENTIALLY TO
COMMUNICATIVE SIGNALS**

Emalie McMahon¹, David Rahabi¹, Nancy Kanwisher¹; ¹Massachusetts Institute of Technology

Our capacity to perceive and interpret communicative signals—whether directed toward us or to another person—is essential for navigating the social world. A recent large-scale naturalistic study found that regions in the superior temporal sulcus may underlie this ability. Regions that had been previously shown to respond to faces (fSTS) and third-person social interactions (SI-STS) both responded more to videos depicting two people communicating with each other than interacting physically or not interacting at all. However, the communicative and noncommunicative videos differed in many ways, leaving open alternative accounts. We, therefore, conducted a controlled fMRI study in which participants were scanned with fMRI while they viewed videos in which the same people in the same setting interacted communicatively or noncommunicatively, or performed independent noncommunicative actions. We independently localized the face-selective (fSTS) and social interaction selective (SI-STS) regions in each participant individually. Both regions showed significantly stronger responses to communicative interactions than to either noncommunicative interactions or independent actions. Is this response specific to viewing third-party communication? Or does it generalize to viewing one person communicating directly to the viewer? We answer this question using close-up videos of faces communicating with another person off-screen (third-person), communicating directly with the viewer (first-person), or performing a non-communicative action like eating or brushing teeth. Both the fSTS and SI-STS responded just as strongly when viewing one face communicating, whether with an unseen partner or directly with the viewer. Together, these findings demonstrate that these STS regions are tuned to detect both third-party and first-person communicative signals.

This work was funded by the MIT Simons Center for the Social Brain

TALK 4, 3:15 PM, 34.24

THE PERIPHERY IS SUFFICIENT FOR NATURAL, DYNAMIC EMOTION PROCESSING

Jefferson Ortega¹ (jefferson_ortega@berkeley.edu), David Whitney¹; ¹University of California, Berkeley

Traditional models of emotion perception posit that high-resolution foveal vision is more informative than peripheral vision for understanding others' emotions. However, these findings rely largely on static, decontextualized images of facial expressions. Here, we revisited the role of the periphery in affective processing, using dynamic, context-rich videos along with a gaze-contingent paradigm. Using the Inferential Emotion Tracking (IET) paradigm (Chen & Whitney, PNAS, 2019), participants (N = 76) continuously rated the valence and arousal of target characters in 11 video clips under two gaze-contingent viewing conditions: Fovea-Only (8° central window visible; periphery masked) or Periphery-Only (8° central scotoma; periphery visible). Performance was compared to an independent Control group that did not have any of their visual field masked. We

found that participants in the Fovea-Only condition exhibited significantly lower emotion tracking accuracy in the IET task and had lower inter-subject agreement in their emotional inferences compared to controls. Conversely, IET task accuracy and between-subject agreement were not significantly different between the Periphery-Only and the full-field Control group. These results upend the conventional understanding of retinotopic contributions to affective processing, demonstrating that peripheral vision is sufficient, and perhaps indispensable, for inferring others' emotions in dynamic, naturalistic environments.

Supported in part by the National Institute of Health (grant no. R01CA236793) to D.W and (grant no. F99NS141343) to J.O.

TALK 5, 3:30 PM, 34.25

DYNAMIC VISUAL RECONSTRUCTION REVEALS DIMENSIONAL REPRESENTATIONS OF FACIAL EXPRESSION PERCEPTION

Yong Zhong Liang¹ (yongzhong.liang@mail.utoronto.ca), Tyler Roberts¹, Gerald Cupchik¹, Jonathan Cant¹, Adrian Nestor¹; ¹University of Toronto

Everyday interactions rely on the ability to parse a rich spectrum of dynamic facial expressions, ranging from canonical emotions to conversational signals, such as disbelief or “yeah-right” smiles. Yet, the representational structure of this broader set, and its temporal unfolding, remain incompletely understood. Here, we combine EEG decoding, behavioral similarity judgments, and frame-based visual reconstruction to characterize how diverse dynamic expressions are represented. Healthy adults (n = 14) viewed 24 one-second videos (14 emotional, 10 conversational expressions) that evolved from neutral to an apex configuration, and separately judged pairwise similarity among the same stimuli. Multivariate EEG patterns supported reliable decoding of all expressions, and representational similarity analysis revealed a low-dimensional neural space that closely tracks a behavior-based space organized primarily along affective valence and arousal. Time-resolved decoding indicated that expression-specific information emerges shortly after stimulus onset and peaks well before an expression configuration reaches its apex, suggesting that neural processing capitalizes on early, diagnostically informative cues. Sliding-window EEG reconstruction further recovered dynamic visual representations that preserve fine-grained distinctions between closely related expressions (e.g., happy satiated vs. schadenfreude), with reconstruction accuracy mirroring the anticipatory decoding profile. Behavioral data yielded convergent static reconstructions with a comparable dimensional organization. Together, these results move beyond proof-of-concept reconstruction to reveal how dynamic emotional and conversational expressions jointly populate a shared, anticipatory representational space.

TALK 6, 3:45 PM, 34.26

INTACT MOTION PROCESSING BUT ALTERED SOCIAL INTERACTION SELECTIVITY IN AUTISM

Hannah Small¹, Haemy Lee Masson², Ericka Wodka^{3,4}, Stewart Mostofsky^{3,4}, Leyla Isik¹; ¹Johns Hopkins University, ²Durham University, ³Kennedy Krieger Institute, ⁴Johns Hopkins School of Medicine

Autism is characterized by altered social interaction and communication, yet reliable neural correlates of these differences remain elusive. Prior work has also highlighted differences in visual perception, including weaker motion processing and biological motion perception, in autism. In the neurotypical population, a region in the right posterior superior temporal sulcus (pSTS) responds selectively to observed social interactions in different dynamic visual displays. Because of its integrative role in social processing, altered selectivity in this region may link visual processing differences to downstream social communication behaviors in autism. For the first time, we collect fMRI responses from matched autistic (n=20, 13 F) and neurotypical (n=34, 17 F) participants while they view silent videos of point-light figures involved in simple social interactions or independent actions. We find significantly reduced selectivity for social motion in right pSTS in autistic participants at the group and individual level. This reduced social visual processing is specific, as we find no differences in general motion processing in the motion-selective middle temporal area (MT) between groups. Ongoing work will characterize fMRI responses in these regions to a naturalistic, audiovisual movie in the both groups of participants. Together these results suggest social interaction perception in the pSTS may be specifically altered in autism, and highlight the dissociation between social and general motion processing in the human brain.

TALK 7, 4:00 PM, 34.27- REMOTE PRESENTATION

FEATURE-SPECIFIC EFFECTS OF OCCLUSION AND FEEDBACK IN FACE RECOGNITION

Maryam Karimi¹ (imariyakarimi@gmail.com), Jalaedin Noroozi^{1,2*}, Mohammad-Reza A. Dehaqani^{1,3*}; ¹Institute for Research in Fundamental Sciences, Tehran, Iran, ²Institute for Convergence Science and Technology (ICST), Sharif University of Technology, Tehran, Iran, ³School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran

Early face processing in inferior temporal cortex is driven by specific diagnostic features, particularly the eye region. However, when such features are missing due to occlusion, this local feature-based representation becomes insufficient. Recent evidence indicates that under these conditions, feedback from prefrontal regions supports the recovery of face representations by modulating activity in inferior temporal cortex. It remains unclear how feedback depends on missing

facial features, contributes to their recovery, and whether its role changes with the degree of occlusion. To examine, we employed a two-alternative-forced-choice (2-AFC) paradigm combined with backward masking to examine how occlusion and stimulus-onset-asynchrony (SOA) influence feedback processing in face recognition. On each trial, participants briefly viewed a face for 35 ms with parametrically occluded features, followed by a phase-scrambled mask at SOAs ranging from 35 to 250 ms. Sixty-six participants took part in one session comprising six blocks, alternating masked and unmasked blocks. Each session included 11 FaceGen-generated identities, each presented at six occlusion levels. Our study demonstrates a double dissociation between feature type and the effects of feedback across different levels of occlusion. Recognition of the eyes was most sensitive to disruption, whereas recognition of the mouth and nose was less affected. Both occlusion and mask presence reduced recognition accuracy, with the strongest impairments observed under high-coverage conditions. In addition, greater disruption of feedback processing at short SOAs, indicating that at these early time points, feedback signals have not yet fully propagated through the cortical hierarchy. Whereas at longer SOAs, feedback information has already reached allowing incomplete information to be integrated and recognition performance to return toward normal levels. We suggest that occluding the eyes leads to the largest decrease in subjects' performance. This indicates that feedback processing likely contains more information about the eyes than other features. Furthermore, higher-order regions (the prefrontal cortex), may preferentially encode features that are more salient.

TALK SESSION: SUNDAY, MAY 17, 2026,

5:15 – 7:15 PM, TALK ROOM 1

Perceptual Training, Learning and Plasticity

Moderator: Aaron Seitz, Northeastern University

TALK 1, 5:15 PM, 35.11

INCREASE IN NEURAL CONDUCTION SPEED FOLLOWING TREATMENT FOR EXTENDED VISUAL DEPRIVATION

Pawan Sinha¹, Chetan Ralekar², Charles Shvartsman¹, Suma Ganesh³; ¹Massachusetts Institute of Technology, ²Indian Institute of Technology, Roorkee, India, ³Dr. Shroff's Charity Eye Hospital, New Delhi, India

Understanding whether the human visual system retains the capacity for myelin-related plasticity beyond early developmental critical periods is essential for both basic neuroscience and clinical practice. Here, we report longitudinal evidence from a cohort of children who

experienced profound congenital visual deprivation until late childhood and then received sight-restoring cataract surgery. Using flash visual evoked potentials (fVEPs), we examined how the speed of neural transmission in early visual pathways changes over the year following treatment. Across six participants (mean age 12 years), we measured the latency of the P1 component - a well-established marker of conduction speed in the retino-cortical pathway - at five time points from three days to one year post-surgery. Immediately after treatment, P1 latencies were significantly delayed relative to neurotypical adults, consistent with slowed transmission in long-deprived pathways. However, we observed systematic and statistically significant latency reductions over time, converging to typical adult values by one year. A mixed-effects model revealed a strong effect of log-transformed time on latency ($p < 0.001$). The rate of convergence was markedly slower, by nearly an order of magnitude, than that documented in infancy, but the eventual approach to normal ranges demonstrates substantial residual plasticity. Comparisons to published neonatal data suggest that some myelination may have occurred prior to treatment due to residual light stimulation through cataracts. Nevertheless, the pronounced post-surgical latency decreases likely reflect activity-dependent increases in myelination within early visual white-matter tracts. This interpretation is supported by our prior structural imaging work showing myelin-associated changes in similar late-sighted populations. These findings challenge long-standing assumptions about the rigidity of myelination timelines, revealing meaningful plasticity in the visual system well past early childhood. They further point to the possibility that experience-dependent myelin remodeling may serve as a modulator of visual learning across the lifespan.

This work was supported by grant R01EY020517 from NEI(NIH)

TALK 2, 5:30 PM, 35.12

SHORT-TERM MONOCULAR DEPRIVATION FACILITATES BEHAVIORAL TRANSFER OF PERCEPTUAL LEARNING TO ADJACENT RETINAL LOCATIONS

Zahide Pamir^{1,2} (zahide.pamir@bilkent.edu.tr), Tutku Karahan^{1,2}, Berk Yüce^{1,2}; ¹Bilkent University, Ankara, Türkiye, ²Aysel Sabuncu Brain Research Center, Ankara, Türkiye

Topographic reorganization refers to an adaptive neural change driven by sensory deprivation, in which the deprived neurons begin responding to stimuli from neighboring locations. Although this phenomenon has been extensively studied at the neural level, its functional consequences remain unclear. The present study combines behavioral and fMRI measurements to investigate whether perceptual learning that occurs in a later deprived cortical region can enhance performance and induce corresponding neural changes at a nearby, untrained retinal location following short-term, reversible monocular deprivation. Fourteen healthy participants were trained for 15–20 days on a 2-IFC orientation discrimination task at the retinal location

corresponding to the blindspot in the behavioral study. Thresholds were measured at the trained location and at an adjacent near location before and after training, under both undeprived and deprived viewing conditions. After confirming significant perceptual improvement at the trained blind spot location ($p < .001$), 2 (Time: pre, post) $\times$ 2 (ROI: deprived-near, undeprived-near) repeated-measures ANOVA revealed a significant Time $\times$ ROI interaction ($p = .015$), demonstrating that learning transferred more strongly to the near location in the deprived condition than in the undeprived condition. Moreover, eight participants completed fMRI sessions before and after training while performing the orientation-discrimination task. Our fMRI analyses focused on the blind-spot ROI during the presentation of near stimuli in the deprived condition. However, contrary to previous reports, we were unable to replicate findings indicating topographic reorganization in the cortical region corresponding to the blindspot during short-term deprivation, nor did we observe neural correlates of the behavioral enhancement associated with the transfer of perceptual learning under short-term deprivation. Although this study provides the first evidence suggesting a functional benefit of topographic reorganization by showing that short-term monocular deprivation may facilitate the transfer of learned abilities to adjacent, untrained locations, the underlying neural mechanisms remain unclear.

This project is funded by the Scientific and Technological Research Council of Turkey, TUBITAK through 2232-B Program (Project No: 121C139).

TALK 3, 5:45 PM, 35.13

LOW-DIMENSIONAL AND OPTIMISED REPRESENTATIONS OF HIGH-LEVEL INFORMATION IN THE EXPERT BRAIN

Andrea Ivan Costantino^{1,2}, Artem Platonov¹, Felipe Fontana Vieira^{1,4}, Emily Van Hove^{1,2}, Merim Bilalić³, Hans Op de Beeck^{1,2}; ¹KU Leuven, ²Leuven Brain Institute, ³Northumbria University, ⁴Ghent University

Chess expertise requires rapid extraction of relational structure from complex visual arrays. Previous research shows that experts organize chessboards into familiar relational configurations (chunks), exhibit distinctive eye movement patterns, and show activations in distinct brain areas, yet the underlying neural mechanisms remain unclear. We move from classic univariate localization approaches to representational geometry, asking what information experts encode, how it is organized, and where it resides. Using multivariate fMRI with 20 chess experts and 20 novices evaluating chess positions, we show that visual features are processed similarly by both groups, but experts compress this information and shift representational load to nonvisual, domain general control networks. Representational similarity analysis reveals comparable geometry for visual properties across early, intermediate, and posterior visual cortex in both groups. However, manifold dimensionality analyses show that experts systematically compress this information, providing more structured and efficient

neural codes to downstream, higher level regions. Critically, only experts implement a neural geometry that aligns with high level relational properties such as checkmate status and strategic motifs, and these compressed manifolds paradoxically support richer decoding of task relevant features. Meta analytic network correlations reveal that novices rely on visual and semantic networks, whereas experts shift representational load to domain general control systems associated with working memory and navigation, implementing low dimensional, domain tuned, structured representations. Preliminary analyses of deep neural networks trained to play chess show similar geometrical properties. Both trained and untrained networks encode visual properties, but only trained networks encode relational features. However, in these networks relational coding emerges through dimensionality expansion rather than compression, suggesting that compression might reflect biological constraints. These results ground classic cognitive theories in measurable geometry and reveal how visual compression and network reallocation enable efficient extraction of relational structure from complex visual scenes.

TALK 4, 6:00 PM, 35.14

RECURRENT INHIBITION PRESERVES EXISTING REPRESENTATIONS DURING NEW VISUAL LEARNING—EVIDENCE FROM PSYCHOPHYSICS, RECURRENT CIRCUIT MODELS, AND DEEP NEURAL NETWORKS

Yu-Ang Cheng¹, Yuka Sasaki¹, Thomas Serre¹, Takeo Watanabe¹; ¹Brown University

While balancing the acquisition of new information with the stability of existing representations is a fundamental challenge in biological and artificial learning systems, the underlying mechanisms remain elusive. This plasticity-stability dilemma concerns how training on a visual feature affects untrained features, yet little work has examined whether untrained features beyond the vicinity of the trained feature are altered or remain unchanged. Here, using psychophysics, recurrent circuit models, and deep neural network models (DNNs), we investigated how training influences untrained features in visual perceptual learning (VPL). First, we measured how untrained features changed after training. After 3 days of training on an orientation detection task, participants were tested with 18 orientations ranging from -90° to +90° relative to the trained orientation. Performance enhancement was confined to $\pm 30^\circ$ around the trained orientation, whereas orientations farther away showed neither enhancement nor deterioration. To explain why distant orientations remained stable, we modeled the result using a recurrent circuit of early visual areas and performed ablation analyses to identify which plasticity mechanisms were critical. Plasticity in excitatory or readout components produced nonspecific performance increases or decreases, inconsistent with behavior. In contrast, inhibitory-component plasticity generated a normalization-like computation that mitigated these unwanted changes and matched the psychophysical pattern. Next, to test whether this mechanism

generalizes to modern deep learning systems, we introduced a trainable LayerNorm (TLN) module into DNNs. Compared with full fine-tuning, which produced interference with untrained features, TLN-only training preserved untrained orientations and reproduced the psychophysical results. Finally, we examined whether new learning degraded visual function. Linear-probe evaluation on ImageNet revealed substantial accuracy loss after full fine-tuning, whereas TLN-only training maintained performance close to the original model. These converging findings suggest that normalization-like processes mediated by recurrent inhibition serve as a key mechanism for protecting existing representations from interference during new learning.

NIH R01EY019466, R01EY027841 and NSF-BSF BCS2241417

TALK 5, 6:15 PM, 35.15

NEUROCHEMICAL AND BOLD SIGNAL CHANGE IN HMT+ AND ATTENTIONAL NETWORKS ASSOCIATED WITH VISION RESTORATION AFTER OCCIPITAL STROKE

Hanna Willis^{1,2}, Lucy Starling², Isobel Rout², Brendan Sargent², Aaron Kay², Rebecca Millington-Truby², I. Betina Ip², Matthew Cavanaugh³, Sara Ajina⁴, Krystel Huxlin³, Marco Tamietto^{5,6}, Holly Bridge²; ¹École normale supérieure, CNRS, ²University of Oxford, ³University of Rochester, ⁴UCL, ⁵Tilburg University, ⁶University of Torino

BACKGROUND: Damage to early visual cortex after occipital stroke causes visual field loss. Training within the blind field can enhance and restore perception. We used multimodal MRI before and after training to investigate whether changes in the function of visual motion area hMT+ and higher-level attentional networks supports observed improvements in vision and awareness. **METHODS:** Eighteen participants (6 female; aged 24–74 years; >6 months post-stroke) completed at least six months of at-home rehabilitation using a two-alternative-forced-choice motion discrimination and integration task. All participants visited the research centre for two visits (pre-training and post-training). Twelve participants also returned after a further three months without additional training (follow-up). At each research visit, participants completed behavioural tasks and an MRI. To assess transfer of learning, they completed a Gabor detection task in the trained locations (50% and 100% contrast; $\sigma=0.8^\circ$; spatial frequency= $1\text{c}/^\circ$). GABA and glutamate concentrations were measured in ipsilesional hMT+ and a control voxel in sensorimotor cortex. fMRI assessed signal change in hMT+ and the rest of the brain during passive viewing of high-contrast Gabor stimuli in the blind field. **RESULTS:** Participants showed significant improvement in motion discrimination that generalised to Gabor detection ($t(16)=-2.19$; $p=0.044$) and awareness ($t(16)=-2.66$; $p=0.017$). These changes in detection (paired t-test: $t(10)=-1.13$; $p=0.28$) and awareness (paired t-test: $t(10)=-0.76$; $p=0.46$) persisted at a three-month follow up visit. Reduced GABA ($r=-0.60$; $p=0.02$) and glutamate ($r=-0.58$; $p=0.02$) in

ipsilesional hMT+ were linked to improved detection. Increased activity in hMT+ and dorsolateral prefrontal cortex correlated with improved detection, while awareness changes were linked to signal changes in contralesional prefrontal cortex (area 46) and inferior parietal lobule. CONCLUSIONS: Extensive training improves visual detection and awareness for both trained and untrained tasks. Gains are supported by selective changes in brain regions involved in motion perception and attention, suggesting that a broad network supports recovery.

H.E.W was supported by a DPhil scholarship from the Medical Research Council United Kingdom (MR/N013468/ 1), a Waverley Scholarship from The Queen's College, Oxford, UK, and a European Research Council (ERC) Horizon 2020 Research and Innovation Programme Grant No 948366-HOPLA.

TALK 6, 6:30 PM, 35.16

TRIAL-BY-TRIAL EVOLUTION OF VISUOCORTICAL TUNING AND FUNCTIONAL CONNECTIVITY DURING AVERSIVE CONDITIONING: A SIMULTANEOUS EEG-FMRI STUDY

Laura Ahumada¹ (laura.ahumadaher@ufl.edu), Andrew Farkas¹, Faith E. Gilbert¹, Hanna M. Engle¹, Judith Escobar¹, Jourdan J. Pouliot¹, Katherine McCain¹, Arash Mirifar¹, Mingzhou Ding¹, Andreas Keil¹; ¹University of Florida

Responses in visuocortical neurons reflect learning and experience about the visual world. The time course and large-scale mechanisms underlying learning in the human visual cortex, however, remain elusive due to a range of methodological limitations. In the present study, simultaneous EEG and fMRI signals were recorded to address some of these limitations. A sample of 25 healthy observers participated in an aversive conditioning paradigm, where one Gabor patch (conditioned stimulus, CS+) was paired with an aversive electrical pulse (unconditioned stimulus). Three additional Gabor patches with different orientations served as generalization stimuli (GS1, GS2, and GS3) and were never paired with the electric pulse. The patches flickered at 15 Hz to generate steady-state visual potentials (ssVEPs). The experimental session consisted of three phases: habituation, acquisition, and extinction. Tuning in the visual area was measured by modeling, at the single-trial level, the ssVEP amplitude across the four Gabor patches, using a Different-of-Gaussian function (Ricker model). Results showed that visuocortical tuning resembled a generalization trend from the CS+ to the most similar patches in the initial part of acquisition, while from middle acquisition through extinction, the tuning narrowed, representing sharpened tuning of the ssVEP amplitude relative to the CS+. This electrophysiological sharpening predicted the increment in BOLD activity in early visual areas (V1/V2). On the other hand, inter-site phase locking of the ssVEP was strong between occipital and parietal sensors, and between occipital and frontal sensors, particularly at the

beginning of acquisition. EEG-BOLD integration showed this early strong connectivity between V1/V2 and areas like the superior frontal cortex, orbitofrontal cortex, and the ventral visual pathway. Overall, these findings are consistent with the notion that changes in tuning to a conditioned visual stimulus are supported by extra-striate inputs, while the late narrowing of the tuning bandwidth in V1/V2 involves local suppressive activity.

TALK 7, 6:45 PM, 35.17

THE COMPLEX ROLE OF MIND WANDERING IN IMPLICIT STATISTICAL LEARNING

Teodóra Vékony¹ (teodora.vekony@pdi.atlanticomedia.es), Bianka Brezóczi^{2,3,4}, Bence C. Farkas^{5,6,7}, Gábor Csifcsák⁸, Péter Simor^{9,10}, Dezső Németh^{1,11,12}; ¹Gran Canaria Cognitive Research Center, Atlántico Medio University, ²Doctoral School of Psychology, Eötvös Loránd University, ³Institute of Psychology, Eötvös Loránd University, ⁴Institute of Cognitive Neuroscience and Psychology, HUN-REN Research Centre for Natural Sciences, ⁵UVSQ, Inserm, Centre de Recherche en Épidémiologie et Santé des Populations, Université Paris-Saclay, ⁶Pôle Pilotage activités et Projets, Direction Générale Adjointe Enfance Famille Santé (DGAFFS), ⁷LNC2, Département d'études Cognitives, École Normale Supérieure, INSERM, PSL Research University, ⁸Department of Psychology, UiT The Arctic University of Norway, ⁹Institute of Behavioral Sciences, Semmelweis University, ¹⁰IMéRA Institute for Advanced Studies of Aix-Marseille University, ¹¹Centre de Recherche en Neurosciences de Lyon CRNL U1028 UMR5292, INSERM, CRNS, Université Claude Bernard Lyon 1, ¹²BML-NAP Research Group, Institute of Psychology, Eötvös Loránd University & Institute of Cognitive Neuroscience and Psychology, HUN-REN Research Centre for Natural Sciences

Mind wandering (MW) is often regarded as a state of mind associated with impaired cognitive performance, yet recent studies suggest it plays a surprising, adaptive role in statistical learning (the implicit acquisition of environmental regularities). Our prior work established that periods of MW are associated with enhanced statistical learning outcomes and identified a plausible neural mechanism, demonstrating that both efficient statistical learning and increased MW are linked to local slow waves in the waking brain. Building upon this foundation, our latest study directly addresses the complex interplay between MW, inhibitory control, and statistical learning. Healthy young participants completed an online experiment that combined simultaneous probabilistic learning and response inhibition (a go/no-go task). Throughout the task, participants provided continuous reports on their attentional state, indicating whether their focus was on-task or MW. We report a clear functional dissociation: increased MW simultaneously impaired inhibitory control (evidenced by reduced response suppression) but remarkably facilitated statistical learning (improved pattern detection in the visuomotor sequence). Crucially, we show that inhibitory control performance significantly modulates the relationship between MW and statistical learning efficiency. These

findings provide a compelling framework linking off-task thought to the dynamics of predictive cognition and waking neural dynamics, thereby highlighting MW's essential and multifaceted role in optimizing cognitive resource allocation.

This research was funded by the Spanish Ministry of Science, Innovation and Universities (MCIU) – State Research Agency (AEI), grant number PID2024-160183NA-I00.

TALK 8, 7:00 PM, 35.18

GAZE-CONTINGENT SCOTOMA AWARENESS TRAINING IN PATIENTS WITH CENTRAL VISION LOSS

Marcello Maniglia¹ (mmaglia@rit.edu), Jason Vice², Elliot Maxwell², Kristina Visscher², Aaron Seitz³; ¹Rochester Institute of Technology, ²University of Alabama at Birmingham, ³Northeastern University

Macular degeneration (MD) is the leading cause of central vision loss in the Western world. With the aging population worldwide, its prevalence is expected to increase, taking a toll on individuals' quality of life, their caregivers' burden, and the healthcare system. Patients often develop spontaneous compensatory strategies, most commonly adopting a preferred retinal locus (PRL), a peripheral region used as a new fixation point and as a new oculomotor reference. However, PRL development is often inefficient, limited by patients' poor awareness of their scotoma's size, shape and location, as well as our lack of mechanistic understanding of PRL development. Studies using simulated scotomas, a framework for lab-based investigations of central vision loss, show that scotoma visibility and contour characteristics can accelerate oculomotor adaptation, suggesting that awareness of visual field loss may promote PRL development. To test the translational potential of this hypothesis, we designed a visual training paradigm to enhance patients' awareness of their blind region using eye-tracker-guided gaze-contingent displays. In Experiment 1, a group of patients with MD trained on a visual discrimination task while a gaze-contingent display drew the perimeter of their individual scotoma on a computer screen in real time. In Experiment 2, a new cohort of patients with MD trained on a modified version incorporating a trained retinal locus (TRL) component, intended to guide PRL development to a specific peripheral location chosen based on their individual structural and functional fundus image exams. A control group of patients was trained on the same visual task without the gaze-contingent component. Results reveal substantial heterogeneity in response to these interventions. We present this variability in terms of changes in oculomotor behavior and visual performance across training conditions and discuss its implications for laboratory-inspired rehabilitation protocols

NEI R21EY033623-01

TALK SESSION: SUNDAY, MAY 17, 2026, 5:15 – 7:15 PM, TALK ROOM 2

Action

Moderator: Cristina de la Malla, Universitat de Barcelona

TALK 1, 5:15 PM, 35.21

CONTACT-POINT SELECTION IN GRASPING UNDER DEGRADED VISION

Carla Lanca^{1,2}, Zoltan Derzsı^{1,2,3}, Robert Volcic^{1,2,3}; ¹Division of science, New York University Abu Dhabi, ²Center for Brain and Health, New York University Abu Dhabi, ³Center for Artificial Intelligence and Robotics, New York University Abu Dhabi

Successful grasping relies not only on motor execution but also on the selection of effective grasp contact positions on the object surface from available visual information. When vision is degraded, the visuomotor system must compensate to maintain successful object interaction. Under full vision, we hypothesized that participants would engage in refined planning, selecting object regions and approach trajectories that maximize stability and overall grasp quality. To test this, we examined how visuomotor choices change when vision is reduced using controlled blur. Participants (n=23) performed reach-to-grasp movements toward 3D-printed objects presented at different orientations and two levels of contrast while hand kinematics were recorded. Under full-vision conditions, behavior showed clear signatures of visually guided selection of grasp contact positions. Movements reflected active consideration of optimal approach paths and contact regions. Grip shaping unfolded later in the reach, consistent with the use of detailed visual cues to refine grasp contact positions up to the moment of grasp. Importantly, grasping angles sometimes deviated intentionally from each participant's natural grasp axis (NGA), reflecting visually driven choices to optimize grasp contact quality. When vision was degraded, behavior shifted toward a more conservative strategy. Movement initiation slowed and maximum grip aperture occurred earlier, indicating reduced reliance on fine-grained visual monitoring. Grasps were less clustered around the center of mass and grasping angles aligned more closely with each participant's NGA, suggesting increased reliance on habitual grasp configurations. Together, these patterns suggest that when vision is available, the sensorimotor system prioritizes grasp quality by refining the selection of grasp contact positions. Under degraded vision, the system simplifies movements, trades spatial precision for robustness, and reverts to natural grasp strategies to maintain successful object interaction. The findings highlight the central role of visual information in shaping the selection of grasp contact positions and in structuring grasping behavior.

We acknowledge the support of the NYUAD Center for Artificial Intelligence and Robotics and the NYUAD Center for Brain and Health, funded by Tamkeen under the NYUAD Research Institute Awards CG010 and CG012.

TALK 2, 5:30 PM, 35.22

DISSOCIATING THE EFFECTS OF HAPTIC OBJECT SIZE AND ORIENTATION ON GRASP APERTURE AND ORIENTATION

Robert Whitwell^{1,2}, Angus Lau¹; ¹Department of Physiology & Pharmacology, Western University, London ON Canada, ²Department of Psychology, Western University, London ON Canada

When we reach for an object, visual input about the object's size and orientation informs the hand's in-flight grasp aperture and orientation, respectively. The visuomotor relationships between these object properties and their corresponding grasp features rely on distinct but overlapping areas of intraparietal cortex. Nevertheless, grasps do not rely solely on a visual analysis of the target. For example, a persistent absence of tangible, real object induces variable and exaggerated scaling of grasp aperture to object size in normally-sighted participants, and in the case of 'DF', who has visual form agnosia, an absence of grip scaling altogether. Is object size and orientation coded independently in the haptic domain like they are in the visual domain? We used a dual adaptation paradigm to address this question. Normally-sighted participants (N=96) reached for visual virtual targets whose real (haptic) sizes and/or orientations were congruent or incongruent with their visual counterparts. Adaptation to incongruent haptic object orientation and size was assessed by measuring aftereffects on grasp orientation and aperture, respectively. Using the logic of dual adaptation, when compensatory updates require changes to the same underlying circuit, interference ensures one or both aftereffects are null. By this same token, independent circuitry affords both compensatory updates to occur and, accordingly, aftereffects for both manipulations. Consistent with independence, haptic changes in target size and orientation induced aftereffects on grasp aperture and orientation, respectively. Furthermore, aftereffects on grasp orientation and aperture under dual adaptation conditions were not significantly different from the effects observed when each haptic target feature was manipulated in isolation. Moreover, inconsistent with a simpler visuomotor memory based explanation, aftereffects on grasp aperture and grasp orientation extended to non-trained visual target sizes and orientations. Overall, these findings indicate that haptic object size and orientation contribute independently to updating motor plans for grasp orientation and aperture.

This work was supported by an NSERC Discovery Grant to RLW (RGPIN-2022-05050)

TALK 3, 5:45 PM, 35.23

VISUAL REPRESENTATIONS REFLECT THE POTENTIAL FOR ACTION

Michael L. Paavola¹, Kade T. Tanke¹, J. Toby Mordkoff¹, Cathleen M. Moore¹; ¹University of Iowa

The ultimate function of vision is the guidance of action. Using the Simon Effect as a tool, we tested the hypothesis that spatial representations include information about whether and how objects can be acted on in addition to their locations in space. The Simon Effect is the finding that responses made with effectors that are on the same side of space as the stimuli on which the responses are based tend to be faster and more accurate than responses made with opposite-side effectors. Critically, Simon Effects occur despite location being irrelevant to the task. The task may be, for example, make a right-hand response to orange stimuli and a left-hand response to blue stimuli; location is irrelevant. The Simon Effect, therefore, provides a measure of spatial information that is incorporated into visual representations independent of instruction. The question addressed here was what is the nature of that spatial information? Using virtual reality to present stimuli at variable distances from the observer, we showed that the Simon Effect was substantially reduced when stimuli appeared at unreachable distances. In another study, again using virtual reality, we presented colored balls that started far away from the observer on the left or right side. The balls moved toward the observer in either a straight trajectory toward the hand on the same side or in a crossed trajectory toward the hand on the opposite side. The Simon Effect for cross-trajectory balls was almost completely reversed relative to same-trajectory balls. Thus, the effective spatial information was not where the balls were in space, which was the same across trajectory conditions, but where they would be if acted on. These studies demonstrate that spatial representations of objects in three-dimensional dynamic environments reflect the potential for acting on the objects, not simply where they are in space.

TALK 4, 6:00 PM, 35.24

PANTOMIMED ACTIONS RECRUIT INTUITIVE KNOWLEDGE ABOUT VISUOMOTOR FEEDBACK

Sholei Croom¹ (scroom1@jhu.edu), Chaz Firestone¹; ¹Johns Hopkins University

Visually guided behavior arises from a complex synergy between perceptual and motor systems; when someone grabs a cup to take a drink, for example, the mechanics of their reach are updated online in response to evolving perceptual input. To what extent are ordinary observers aware of this aspect of others' goal directed actions? Here, we explored these questions through "pantomimed actions", in which people perform actions with imaginary objects. We created a stimulus set of videos where agents performed both genuine object-directed actions (e.g., stepping over a box), and pantomimes of those actions

(e.g., stepping over an imagined box). We asked both (a) whether naive observers who watch these videos can distinguish real actions from pantomimed actions, and also (b) which kinds of information underwrite this performance. In Experiment 1, subjects watched raw video of real and pantomimed actions side-by-side, with a black 'censor bar' covering the real (or imagined) object's location. Under these conditions, subjects were able to distinguish the two action types at rates above chance; for example, they could tell whether someone was interacting with a real (vs merely imagined) box, and also whether someone was shuffling between two real (vs merely imagined) poles. Moreover, subject text responses reflected rich inferences about which features of the movement should be diagnostic. However, in Experiment 2, subjects viewed the same actions but with body movements instead depicted by simple 'pose skeletons' (dots connected by lines on a black background) generated from the original videos. Under these conditions, observer performance dropped to chance, despite the kinematic information being preserved across experiments. Together, these results suggest that ordinary people can relate differences in action kinematics to differences in sensory conditions, but that this capacity must be grounded in contextual information about the actor's relation to their environment.

TALK 5, 6:15 PM, 35.25

DISSOCIABLE SYSTEMS FOR VISUALLY GUIDED NAVIGATION AND REACHING IN HUMAN PARIETAL CORTEX

Hee Kyung Yoon¹, Yaelan Jung¹, Daniel Dilks¹; ¹Emory University

The parietal cortex is widely thought to support visually guided actions, but whether it contains distinct regions specialized for different action classes – such as navigation, reaching, and grasping – remains unknown. Prior work implicates the superior parietal lobule (SPL) in navigation and the superior parietal occipital cortex (SPOC) in reaching, yet whether these regions are truly dissociable is unclear because each has never been tested for the other function. Here, we addressed this question using fMRI in human adults. Participants viewed four types of stimuli: Dynamic Scenes (video clips of first-person motion through scenes), Static Scenes (static images taken from these same movies, rearranged such that first-person motion could not be inferred), Contextual Reaching (video clips of first-person reaching motion on a scene background), and Isolated Reaching (video clips of the same actions on a black background). A clear double dissociation emerged: SPL responded significantly more to Dynamic than Static Scenes – consistent with its role in visually-guided navigation – and, critically, more to Dynamic Scenes than either reaching condition. By contrast, SPOC responded significantly more to both reaching conditions than to either scene condition. Resting-state functional connectivity further supported this double dissociation: SPL showed stronger connectivity with “leg-motor cortex” than with “arm-motor cortex”, whereas SPOC showed the opposite pattern.

Together, these findings reveal two distinct parietal systems: SPL for visually-guided navigation and SPOC for visually-guided reaching – clarifying how the parietal cortex organizes visually guided action.

R01 EY29724

TALK 6, 6:30 PM, 35.26

ENVIRONMENT AND TASK STRUCTURE NATURAL GAZE BEHAVIOR: SPATIAL AND SPATIO-TEMPORAL PATTERNS IN STAIR VS. HILL CLIMBING

Mahdis Dadfar¹ (dadfarm@mail.uc.edu), Kathryn Bonnen², Trenton D Wirth¹; ¹Department of Psychology, University of Cincinnati, ²Department of Optometry, Indiana University

During locomotion over complex terrain, surface geometry constrains gaze behavior: walkers adapt their strategies to locomotor demands, with gaze tuned to foothold availability (Matthis, Yates & Hayhoe, CB 2018; Wirth & Matthis, VSS 2022). But how does a highly regular, constructed surface like a staircase, compared to a natural slope, constrain the spatio-temporal structure of gaze during walking? This study (N=5) examined gaze strategies as participants ascended and descended 30m of stairs (18°) and an adjacent sloped hill (22°). A mobile eye tracker and IMU (Pupil Neon, Berlin) recorded eye elevation and head pitch during each walk; their sum defined gaze angle. Because stairs impose greater foot-placement constraints, we predicted more structured gaze patterns on the stair than on the hill. Mean gaze angle showed strong effects of terrain and direction, with steeper, more downward-angled gaze on stairs and during descent, and a significant interaction (all $p < .01$). We analyzed spatio-temporal gaze behavior using recurrence quantification analysis (RQA) of z-scored gaze angle, assessing recurrence rate (RR; $\epsilon = 0.07$, $L_{min} = 5$) in a mixed-effects model. RR was higher on the staircase than the hill $F(1,192) = 60.64$, $p < .001$, with more revisitation during descent, with a larger descent-ascent difference on stairs (all $p < .001$). This finding supports our hypothesis by demonstrating greater spatial regularity on stairs. Next, we locked RR at 0.05, allowing ϵ to vary. We found only walking direction to be significant; ascending walks exhibited greater determinism, entropy, and laminarity (all $p < .001$), indicating more regular and complex gaze sequences during ascent, independent of terrain. Additionally, mean gaze autocorrelation (lag 0.1-5s) displayed only a terrain $\times$ direction interaction ($p \approx .015$), suggesting context-specific temporal persistence. Taken together, terrain predictability and locomotor demands jointly shape gaze dynamics, but in different ways. Stair walking produced more spatial structure (revisitation) and was direction-sensitive, while ascent elicited greater spatio-temporal structure within those recurrences.

University of Cincinnati Office of Research Collaborative Pilots Grant 1022124

TALK 7, 6:45 PM, 35.27

RETINAL CURL AS A DIRECT CONTROL VARIABLE FOR LOCOMOTION: EVIDENCE FROM GAZE-CONTINGENT BIASES

Joan Lopez-Moliner¹ (j.lopezmoliner@ub.edu), Kontessa Ioanna Zorpala¹; ¹Universitat de Barcelona

Traditional models of visually guided locomotion assume that the brain must recover the Focus of Expansion (FOE) by filtering out rotational flow (curl) caused by eye movements. We propose an alternative control procedure where the visual system utilizes retinal curl directly to steer locomotion, rendering the explicit recovery of heading direction unnecessary. We exposed participants ($n=10$) to visual stimuli simulating walking along straight and curved paths (1 m/s), presented on a large screen (2.03 m x 1.16 m; PROPixx at 120 Hz) displaying a natural-like textured floor generated with simplex noise. Participants maintained fixation on a ground target with different eccentricities to their path, a natural behavior that introduces sustained retinal curl. Participants continuously reported their perceived direction of motion via a rotational encoder. To isolate the role of rotational flow, we employed a real-time manipulation: in critical trials, we cancelled the curl component centered on the fovea while preserving translational flow, effectively removing the rotational signature of gaze stabilization. Under natural flow conditions, participants exhibited systematic heading biases opposite to the direction of gaze (e.g., fixating left induced a rightward bias). Crucially, these biases vanished in the 'cancelled curl' condition, identifying retinal curl as the specific driver of the perceptual shift. We successfully modeled these trajectories using a simple feedback controller that steers locomotion proportionally to the instantaneous mean retinal curl, without any explicit reconstruction of heading direction. The model fits demonstrate that a system reacting solely to retinal curl—treating it as an error signal rather than noise—automatically produces the observed behaviors. These findings challenge the assumption that the visual system must decompose flow into rotation and translation. Instead, by exploiting the geometry of gaze stabilization, the brain may simplify navigation into a low-dimensional control task where recovering the true heading vector is computationally superfluous.

JLM was supported by Grant PID2023-150081NB-I00 funded by MICIU/AEI/10.13039/501100011. KIZ was supported by fellowship PREP2023-001890 from the Ministry of Science and innovation

TALK 8, 7:00 PM, 35.28

VISUAL MODEL OF COLLISION AVOIDANCE GENERALIZES TO MUTUAL AVOIDANCE

Kyra Veprek¹, William Warren¹; ¹Brown University

Walking pedestrians rely on visual information to avoid collisions with moving obstacles. A visual model based on the obstacle's (i) change in bearing angle and (ii) optical expansion rate closely reproduced avoidance behavior with single and multiple moving obstacles, and crossing flows of pedestrians (Bai, 2022; Veprek & Warren, VSS 2024, 2025). However, we developed the model for participants avoiding non-responsive obstacles that moved on linear trajectories. Here, we investigate whether the model generalizes to mutual collision avoidance between two responding pedestrians. We recorded pairs of participants who simultaneously walked across a room and avoided colliding with each other. Participants walked between virtual targets in a mixed-reality environment (Meta Quest 3), where the physical room and the other participant were visible. Dyads started each trial at different locations on a circle ($r = 2.75\text{m}$) and we manipulated their angle of crossing (60, 90, 120, 150, 180 degrees) and walking speed ('fast' and 'normal') to vary the potential collisions. We tested the visual model against the mutual avoidance data under three model-fitting conditions: (1) fixed parameters previously fit to single-avoidance data (Bai, 2022), (2) refitting the optical-expansion threshold for initiating avoidance, and (3) refitting all parameters to the mutual-avoidance data. Position error, defined as the mean Euclidean distance between predicted and observed trajectories (0.30m with fixed parameters), was only marginally reduced ($\sim 0.01\text{ m}$) by refitting. This result indicates that the single-avoidance model also explains mutual avoidance behavior. Consistent with model predictions, the difference in Time-to-Collision-Point, defined as the projected time to the intersection of the two paths, was a strong predictor of which participant passed in front of the other (logistic regression, accuracy = 92.96%). These findings show that the same visual control principles governing single-obstacle avoidance extend to mutual pedestrian interactions.

This research was supported by NIH R01EY029745.

TALK SESSION: MONDAY, MAY 18, 2026, 8:15 – 9:45 AM, TALK ROOM 1

Attention 3

Moderator: Barry Giesbrecht, University of California, Santa Barbara

TALK 1, 8:15 AM, 41.11

DEEP MIND-WANDERING AMPLIFIES ATTENTIONAL CAPTURE

Shivang Shelat^{1,2}, Daniel D. Thayer¹, Jonathan W. Schooler¹, Barry Giesbrecht^{1,2}; ¹University of California, Santa Barbara, ²Institute for Collaborative Biotechnologies

Failures of human attention take many forms. One example is when our attention drifts away from a task toward irrelevant thoughts: mind-

wandering. Another example is when our attention is pulled toward a salient visual distractor at the expense of a target: distractor capture. In this experiment, we test how mind-wandering and spatial attention are linked by interleaving thought probes into the additional singleton paradigm. Participants ($n = 68$) indicated a bar's orientation in a target shape among nontarget shapes (e.g., the diamond among circles), and one nontarget was a color singleton distractor on half of the trials. Periodic probes asked participants to report the depth of their mind-wandering. The distractor did not affect accuracy but slowed response times (RTs) to the target ($b = 22.77$, $SE = 2.78$, $p < .0001$), while deeper mind-wandering reduced accuracy ($b = -0.019$, $SE = 0.0066$, $p = .0033$) but did not affect RTs. Critically, on correct trials, there was a significant interaction between distractor presence and mind-wandering on RTs ($b = 0.87$, $SE = 0.40$, $p = .030$). The RT distractor cost was larger when participants reported deep mind-wandering before the trial onset. On the other hand, when participants were not mind-wandering (i.e., when they were strongly focused), the distractor cost shrunk. This suggests that the classic behavioral cost of distractor capture in the additional singleton paradigm is modulated by mind-wandering.

This work is supported by a National Science Foundation Graduate Research Fellowship awarded to the first author under grant number 2139319.

TALK 2, 8:30 AM, 41.12

MODULATIONS OF PERCEPTUAL PRIORITY OF FOVEAL AND EXTRAFOVEAL CONTENT DURING FIXATION

Sanjana Kapisthalam¹, Martina Poletti¹; ¹University of Rochester

Naturally, perception unfolds in a continuous fixation–saccade cycle. Saccades reposition the fovea several times per second, and fixations (typically 200–300 ms) constitute a stable and relatively long time window in which visual information is acquired. Yet, fixation is often treated as a temporal unit. Here we examine how visual discrimination changes over the course of fixation after saccade landing. Observers made a saccade to the center of the display and reported the direction of a brief (50 ms) orientation change of a continuously visible Gabor (2° diameter, either at 2 or 8 cpd). The orientation change could occur any time up to 450 ms after saccade landing. Gabors were either presented in isolation at a given eccentricity—either foveally (1°) or parafoveally (8°)—or were presented simultaneously at both eccentricities. In the latter case, only one, randomly picked, Gabor changed orientation. Stimulus contrast was adjusted to achieve threshold performance for each eccentricity and spatial frequency. When stimuli were presented in isolation, orientation-discrimination performance steadily declined over the 450 ms following saccade landing. This decline was comparable for foveal and extrafoveal stimuli. In contrast, when stimuli were presented simultaneously at both eccentricities, discrimination declined for foveal stimuli, but increased for parafoveal stimuli over the course of fixation. This pattern

showed an early foveal advantage followed by a later parafoveal gain and occurred only for high-spatial-frequency stimuli (8 cpd). For low spatial frequency stimuli (2 cpd), performance remained stable for both foveal and parafoveal stimuli throughout fixation. These results show that even during brief fixations, visual discrimination is not constant but changes substantially depending on stimulus eccentricity and spatial frequency. Immediately after saccade landing, the visual system prioritizes high-resolution foveal content, but over the course of fixation this prioritization shifts toward extrafoveal stimuli, even when no further saccade is planned.

Meta Reality Labs, NIHEY001319 to the center for visual science at University of Rochester

TALK 3, 8:45 AM, 41.13

TEMPORAL DYNAMICS OF REWARD PROCESSING IN VISUAL DECISION-MAKING

Christopher J.H. Pirrung¹ (chris.pirrung@nih.gov), Dalia Abdo Kahin¹, Chris Baker¹; ¹Laboratory of Brain and Cognition, National Institutes of Health

In visual decision making, visual information and reward processing interact to shape learning with visual and frontal cortex implicated in each, respectively. The Reward Positivity (RewP) is a sensitive M/EEG marker of reward, typically measured at frontocentral sensors/electrodes. The RewP is likely generated by a distributed cortical network whose individual nodes contribute heterogeneous information. Given the broad set of contributors to this signal, using a univariate approach such as ERP analysis may not capture the full scope of this signal. Multivariate pattern analysis (MVPA) has shown increased sensitivity to neural signals and allowed for successful classification of perceptual stimuli based on evoked neural activity. Here we tested whether MVPA of MEG data can be used to temporally decode rewarding and non-rewarding feedback in individual participants. Participants completed a visual discrimination task followed by a visual cue (blue or yellow screen) that indicated monetary reward. Feedback-color mapping changed between runs. MVPA was first used to classify color at each timepoint, irrespective of feedback. This sanity check model accurately predicted color from 70–350ms. Feedback was accurately predicted from 200–700ms in a model including both blue- and yellow-reward trials. A cross-decoding model, in which the training set included rewards of one color and the testing set included rewards of the other color, made systematic errors at early timepoints (70–250ms) in line with color processing, but successfully predicted feedback at later timepoints (350–700ms). This suggests that we are successfully decoding the response to the cognitive content of the feedback, not just the perceptual attributes. Interestingly, preliminary source estimation shows the ability to decode feedback not only in expected frontal areas, but also early visual areas. This method could allow for differential decoding of perceptual and

cognitive attributes of reward, enabling us to explore how perceptual elements of reward guide decision-making.

ZIA MH002893

TALK 4, 9:00 AM, 41.14

TWO PATHS TO GUIDE ATTENTION: DISSOCIATING STATISTICAL AND REWARD LEARNING

Carola Dolci¹ (carola.dolci@ugent.be), Tom Verguts¹, Elisa Santandrea², Nico Boehler¹; ¹Department of Experimental psychology, Ghent University, ²Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona

The human ability to anticipate and efficiently respond to future events relies on learning regularities that shape attentional priorities. Two major experience-dependent phenomena, statistical learning (SL) and reward-based learning (RL), are known to bias attention toward frequently occurring or highly valued stimuli. Yet, it remains unclear whether these processes operate through shared or distinct neural pathways. To address this question, we conducted a two-session EEG study in which participants performed a visual search task designed to independently manipulate SL and RL. In the SL session, targets appeared more frequently in a specific spatial location, whereas in the RL session, correct responses at certain locations yielded higher rewards. To probe latent attentional biases established through learning, each search display was preceded by a task-irrelevant “ping,” allowing us to measure early reactivation of spatial priorities. Behavioral results demonstrated robust facilitation in both learning contexts: participants responded faster and more accurately when targets appeared in high-frequency (SL) or high-reward (RL) locations. Multivariate EEG decoding revealed a clear dissociation between learning types. SL induced spatially specific neural modulation during the ping display, whereas RL did not. Rather, ERP analyses revealed that RL modulated a later lateralized component related to the discrimination of the target (the SPCN), reflecting greater cognitive effort and discrimination demands at low- (vs. high-) reward locations. Hierarchical drift diffusion modeling supported these distinctions by revealing complementary effects on decision dynamics and discriminability across the two learning types. Together, these findings show that SL and RL, despite producing similar behavioral enhancements, rely on (partially) dissociable neural mechanisms. SL shapes anticipatory visual processing even before target onset, reflecting the encoding of learned spatial regularities. In contrast, RL primarily impacts later value-driven selection and decision-related processes during the search itself.

TALK 5, 9:15 AM, 41.15

UNIQUE CONTRIBUTIONS OF SELECTIVE AND SUSTAINED ATTENTION TO NEURAL IMAGE REPRESENTATIONS

Anna Coriveau^{1,2}, Dongfang Tian¹, Matthieu Chidharom^{1,2}, Edward K. Vogel^{1,2,3}, Monica D. Rosenberg^{1,2,3}; ¹Department of Psychology, University of Chicago, ²Institute for Mind and Biology, University of Chicago, ³Neuroscience Institute, University of Chicago

Attention prioritizes relevant information from the environment but is not stable from moment-to-moment. While relevant information receives an overall processing benefit (selective attention), sustained attentional state fluctuates on top of attentional selection leading to trial-to-trial variability in processing. To what extent do these separable components of attention affect the fidelity of visual information represented in neural EEG signals? Participants performed four 15-minute blocks (600 trials/block) of a continuous performance task in which stimuli were images overlaid on oriented Gabor patches. In two blocks, images were task-relevant. On each trial, participants responded with a button press when an image was from a frequent category (food and vehicles) and withheld a response when it was from the infrequent-category (animals, 10%). During the other two blocks, participants responded based on Gabor orientation. To quantify differences in sustained attentional state, trials were split based on whether they occurred during in-the-zone (low response variance) and out-of-the-zone (high response variance) states within both image-relevant and image-irrelevant blocks. Participants made more commission errors when out-of-the-zone ($p=0.013$). We compared EEG logistic regression classification (food vs. vehicles) for relevant, irrelevant, in-the-zone, and out-of-the-zone trials. Classification performance was higher for relevant vs. irrelevant images, with evidence of a difference lasting from 250-600ms after trial onset. However, sustained attention influenced classification differently for image-relevant and image-irrelevant blocks, suggesting an interaction between selective and sustained attention. During image-relevant blocks, classifier performance was higher for in-the-zone vs. out-of-the-zone trials (300-450ms after onset). However, when images were irrelevant, in-the-zone classification under-performed out-of-the-zone classification in the initial period following stimulus onset but this pattern reversed later in the trial. Results reveal unique representational time courses for different selective and sustained attentional states providing further support for a distinction between attentional subcomponents.

Office of Naval Research Multidisciplinary University Research Initiatives (MURI) N00014-23-1-2768 to E.K.V. and M.D.R.

TALK 6, 9:30 AM, 41.16

THE “ATTENTIONAL BLINK” IS NOT ALWAYS ATTENTIONAL: DIFFERENT MECHANISMS UNDERLY TASK-SWITCHING AND SINGLE-TASK DESIGNS

A well-documented limitation of visual processing is the transient deficit in responding to a second visual target (T2) when presented within the specific temporal interval of 100-400 ms after a first (T1) in Rapid Serial Visual Presentation (RSVP). The mechanisms of this “attentional blink” (AB; Raymond et al., 1992) have been investigated using various target-defining and response features, sometimes manipulating target-defining features within a single trial (e.g., identify the only white letter, then indicate whether a black X appeared) and sometimes keeping the target-defining feature constant (e.g., look for two target letters among numbers). We provide evidence that although the transient T2 deficit in the former “task-switching” paradigm may indeed be due to “attentional” limitations, deficits in the latter “single-task” paradigm may result from neural habituation of the target-defining feature (Rusconi & Huber, 2018). Specifically, observers were presented with two target words defined by their semantic category (e.g., colors) among nontargets (e.g., clothes) and a lag-dependent T2 deficit was observed even when participants were instructed to report only the second instance of a target. When T1 was defined by a different category from T2 (e.g., numbers for T1 and colors for T2), ignoring T1 did not produce a lag-dependent T2 deficit; therefore, different mechanism may underly performance deficits between these two tasks. A follow-up experiment was conducted in which participants categorized only the final word presented, which appeared at various intervals after another word of the same category. We observed lag-dependent changes in RT, demonstrating that encoding T1 to memory is not required for eliciting the transient T2 deficit for single-task designs. Finally, we demonstrate a strong reduction in an ERP component associated with late perceptual processing (the N170) during the AB in a single-task design. Together, these results suggest that different mechanisms underly single-task and task-switching AB paradigms.

TALK SESSION: MONDAY, MAY 18, 2026, 8:15 – 9:45 AM, TALK ROOM 2

3D Shape and Space Perception

Moderator: Michele Rucci, University of Rochester

TALK 1, 8:15 AM, 41.21

INTEROCULAR SUPPRESSION WEAKENS BINOCULAR DISPARITY INTEGRATION IN THE DORSAL VISUAL STREAM

Rong Jiang¹, Ming Meng²; ¹School of Psychology, South China Normal University, Guangzhou, China, ²Department of Neurobiology, The University of Alabama at Birmingham, Birmingham, USA

Binocular vision includes two complementary processes: integrating binocular disparity signals and suppressing interocular conflicts. Behavioral studies indicate that these processes influence each other, but the underlying neural mechanisms remain unclear. Here, we used fMRI and multivariate decoding to investigate how interocular suppression affects disparity representations in the human dorsal visual stream. Participants performed a central fixation task in the scanner. We manipulated the central visual field (eccentricity 0.4°–2°) to induce binocular fusion or rivalry, and measured neural responses to near vs. far disparities in correlated or anticorrelated random-dot stereograms in the peripheral region (eccentricity 2.5°–3.5°), yielding four conditions: correlated fusion (CF), correlated rivalry (CR), anticorrelated fusion (AF), and anticorrelated rivalry (AR). Within-condition decoding revealed robust near–far discrimination for CF across V1–V3, V3A, and IPS. Rivalry (CR) significantly reduced decoding accuracy in V3, V3A, and IPS, while decoding for anticorrelated conditions (AF, AR) dropped to chance. Cross-condition decoding showed that near–far decoding accuracy between CF and CR was significantly above chance level, suggesting that rivalry weakens, but does not remap, disparity representations. In V3A, cross-decoding accuracy between CF and AF was significantly below chance, indicating an inverted disparity code. This inverted disparity code was also disrupted by the interocular suppression process, as cross-decoding accuracy between CF and AR returned to chance level. Informational connectivity analyses revealed strong coordination among ROIs under CF. This coordination was significantly reduced within the V3A-centered network during CR, while connectivity within early visual areas (V1–V3) remained unchanged. Together, these results not only characterize how interocular suppression weakens disparity-selective signals and their large-scale coordination in the dorsal visual stream of typical human observers, but also provide insights that may guide future research on the neural mechanisms underlying stereoscopic deficits in binocular disorders.

TALK 2, 8:30 AM, 41.22

THE ROLE OF VISUOMOTOR CONTINGENCY IN STEREOSCOPIC DEPTH PERCEPTION

Jie Z. Wang¹ (jwang255@ur.rochester.edu), Y. Howard Li¹, Jonathan D. Victor², Michele Rucci¹; ¹University of Rochester, ²Weill Cornell Medical College

Eye movements occur continually during visual fixation and profoundly influence visual sensitivity. Yet their functions in 3-D visual processing have been little explored. Recent work has shown that the incessant inter-saccadic motion of the two eyes introduces temporal modulations in disparity signals that the visual system exploits to establish 3-D representations. Here we show that, in addition to the presence of disparity modulations, an important component of this process is the alignment between incoming disparity changes and motor expectations from eye movements. In a forced-choice procedure, observers reported whether the upper or lower half of a vertically

slanted planar surface appeared closer. Stimuli were random-dot stereograms viewed through a stereoscope. Eye movements were continually measured with a digital DPI eye-tracker, and a custom apparatus for gaze-contingent display updated stimuli in real-time to allow for disruption of the natural visuomotor contingency. Trials alternated among three conditions. In the “Stabilized” condition, disparity modulations were eliminated by moving the stimuli with the eyes. In the “Reconstructed” condition, stimuli moved to counteract ongoing eye movements and recreate the retinal motion generated by previously recorded eye traces, yielding disparity modulations largely uncorrelated with current eye movements. In the “Flipped-vergence” condition, the sign of vergence-induced disparity was inverted and version-induced disparity was eliminated. Performance was greatly impaired under retinal stabilization, replicating previous findings. Discrimination improved in the Reconstructed condition, demonstrating that passive exposure to disparity modulations is beneficial, but does not suffice to support normal sensitivity. Strikingly, performance in the Flipped-vergence condition was even more impaired than under stabilization. These results indicate that stereopsis relies not only dynamic disparity signals, but also on their congruence with ongoing oculomotor signals. These findings point to a critical role of visuomotor contingency in depth perception.

Supported by NIH EY18363 and P30 EY001319

TALK 3, 8:45 AM, 41.23

BENCHMARKING HUMAN AND DNN BIASES IN MONOCULAR DEPTH ESTIMATION

Yuki Kubota¹, Taiki Fukui¹; ¹Communication Science Laboratories, NTT, Inc.

Human depth perception from a single image is systematically biased, yet the characteristics of these distortions remain insufficiently understood. Meanwhile, modern monocular depth estimation (MDE) models achieve high physical accuracy, raising a central question: to what extent do such models reproduce—or diverge from—human perceptual biases? To address this, we constructed two human-annotated depth datasets using established benchmarks: NYU (indoor scenes) and KITTI (outdoor scenes). These datasets enabled direct comparisons between human observers and 69 deep neural networks (DNNs), spanning diverse architectures, training strategies, datasets, and output formats. Human data were obtained by asking participants to report the absolute distances to four simultaneously marked target points in each photograph. Model accuracy was quantified using scale-invariant RMSE. Human–model similarity was defined as a partial correlation between model and human error patterns, obtained by repeatedly correlating fixed model errors with split-half averages of observer errors while controlling for ground-truth depth. We further applied an affine decomposition that isolates per-image affine distortions (scale, shift, horizontal shear, and vertical shear) from

residual error. Across both datasets, humans showed robust and systematic deviations from physical ground truth, as indicated by high split-half human–human partial correlations of error patterns (NYU: 0.808; KITTI: 0.671). Examining the relationship between accuracy and human similarity revealed a clear pattern: similarity increased with accuracy up to approximately human-level performance, but declined for models surpassing that range—indicating a distinct accuracy–similarity trade-off. Notably, this trade-off was substantially more pronounced in the KITTI dataset. Overall, our findings demonstrate that human-like behavior in MDE does not emerge simply by improving metric accuracy. Instead, the divergence suggests that the strategies used by state-of-the-art DNNs, which potentially rely on dataset-specific cues to maximize precision, fundamentally differ from the generalized perceptual heuristics employed by human observers.

TALK 4, 9:00 AM, 41.24

FINAL VISION POINTS YOU TO THE RIGHT (OR WRONG) PLACE: A TWO-MODE FRAMEWORK FOR DEPTH-GUIDED ACTION

Carlo Campagnoli¹, Fulvio Domini²; ¹University of Leeds, ²Brown University

We present a re-analysis of a previous reaching study showing that the classic distinction between planning and online control does not reflect a shift from internal representation to sensory feedback, but rather a context-dependent selection between two environmentally constrained control modalities. Even when behaviour appears to express an internal estimate of target depth, the motor outcome remains determined by the constraints available at the moment the movement must be completed, not by the accuracy or persistence of any internal representation. And, critically, when such a representation is forced to operate, it is systematically wrong. In a first experiment, visual feedback was removed shortly before movement completion. Under these conditions (where no late correction was possible) endpoints reflected only the quality of the sensory information available earlier in the reach. Binocular viewing produced near-veridical performance, whereas monocular viewing generated a consistent overshoot. This pattern does not reveal a privileged “internal model” guiding action. Instead, it reveals a regime in which, absent usable feedback, the system is constrained to act on an impoverished and biased mapping of depth. The second experiment preserved visual feedback until the end and manipulated whether the final segment provided binocular or monocular information. Here, the behaviour previously attributed to an internal depth estimate vanished entirely. Instead, endpoints were dictated solely by the final visual configuration. Switching from monocular to binocular vision produced accurate reaches regardless of earlier viewing, whereas switching from binocular to monocular vision produced overshooting despite prolonged access to reliable disparity. In other words, the system did not rely on a stored internal depth estimate: it replaced it whenever

environmental information permitted. Together, these findings challenge interpretations based on stable internal representations guiding action. Reaching behaviour appears to reflect contextual dominance of environmental constraints, not the fidelity of an internal depth estimate (accurate or otherwise).

TALK 5, 9:15 AM, 41.25

HUMAN-LEVEL 3D SHAPE PERCEPTION EMERGES FROM MULTI-VIEW LEARNING

Tyler bonnen¹ (tyler.ray.bonnen@gmail.com), Jitendra Malik¹, Angjoo Kanazawa¹; ¹UC Berkeley

Humans can infer the three-dimensional structure of objects. Modeling this ability has been a longstanding goal for the science and engineering of visual intelligence, yet decades of computational methods have fallen short of human performance. Here we evaluate a novel class of neural networks that, for the first time, match human accuracy in 3D perception experiments. These models are trained with multi-view image sequences and corresponding self-motion cues—visual-spatial information analogous to human sensory inputs in natural environments. To evaluate this novel modeling approach we leverage an existing 3D perception benchmark (MOCHI), which reveals a considerable gap between humans and standard computer vision models using a concurrent visual discrimination ('oddy') task. We determine the performance of these multi-view models by developing a zero-shot evaluation approach, then compare to human (n=350) responses to the same images. These multi-view models match human-level 3D shape inferences, even without task-specific training or fine-tuning. Remarkably, independent model readouts predict human error patterns and reaction times, revealing an emergent correspondence between model dynamics and human perceptual processing. Our work introduces a novel modeling framework to formalize and evaluate theories of human visual perception, demonstrating that human-level 3D abilities emerge in neural networks trained with naturalistic visual-spatial data.

This work is supported by the National Institute of Neurological Disorders and Stroke of the National Institutes of Health (Award Number F99NS125816)

TALK 6, 9:30 AM, 41.26 - REMOTE PRESENTATION

MRD: Using physically based differentiable rendering to probe vision models for 3D scene understanding

Benjamin Beilharz¹, Thomas S.A. Wallis²; ¹Technical University of Darmstadt, ²Center for Mind, Brain and Behavior (CMBB), Universities of Marburg, Giessen, and Darmstadt

While deep learning methods have achieved impressive success in many vision benchmarks, it remains difficult to understand and explain

the representations and decisions of these models. Though vision models are typically trained on 2D inputs, they are often assumed to develop an implicit representation of the underlying 3D scene (for example, showing tolerance to partial occlusion, or the ability to reason about relative depth). Here, we introduce MRD (metamers rendered differentially), an approach that uses physically based differentiable rendering to probe vision models' implicit understanding of generative 3D scene properties, by finding 3D scene parameters that are physically different but produce the same model activation (i.e. are model metamers). Unlike previous pixel-based methods for evaluating model representations, these reconstruction results are always grounded in physical scene descriptions. This means we can, for example, probe a model's sensitivity to object shape while holding material and lighting constant. We assess multiple models in their ability to recover the scene parameters of geometry (shape) and bidirectional reflectance distribution function (material). The results show high similarity in model activation between target and optimized scenes, with varying visual results. Qualitatively, these reconstructions can make clear the physical scene attributes that models are sensitive or invariant to. MRD holds promise for advancing our understanding of both computer and human vision, enabling us to efficiently answer the question of how physical scene parameters cause changes in model responses.

Funded by the European Union (ERC, SEGMENT, 101086774). Views and opinions expressed are however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

TALK SESSION: MONDAY, MAY 18, 2026, 10:45 AM – 12:15 PM, TALK ROOM 1

Visual Search: Attention, mechanisms, models

Moderator: Jeremy Wolfe, Brigham and Womens Hospital / Harvard Med

TALK 1, 10:45 AM, 42.11

NEURAL REPRESENTATIONS OF SPATIAL PRIORITY DURING VISUAL SEARCH IN FEATURE-SELECTIVE AND PARIETAL CORTEX

Daniel D Thayer¹, Thomas C Sprague¹; ¹University of California, Santa Barbara

Attention is directed to locations that are goal-relevant and image-salient (but irrelevant). To identify important locations, priority map theory posits that feature-selective retinotopic maps (e.g., color or

motion) index salient (Thayer & Sprague, 2023) and relevant (Thayer & Sprague, 2025) locations based on their preferred feature dimension, which is subsequently integrated into a feature-agnostic priority map, where the most important location guides attention. Even though feature dimension maps are crucial for directing attention, it is unclear how neural responses within these maps reflect the priority signals associated with relevant and salient stimuli during visual search. Here, we used a covert search task to evaluate how concurrently present relevant and salient items are represented in neural feature dimension maps. On each trial, participants were cued to search for a target defined by a specific color or motion direction in a subsequent search array containing 8 colorful moving dot stimuli. All array items had homogenous features, except for the target, which differed solely on the cued feature dimension, and the occasional salient distractor, which was either a unique color or motion direction. We used an encoding-based multiple regression analysis with activation patterns in feature-selective retinotopic regions (motion [TO1/TO2] and color [hV4/VO1/VO2] maps) to assess the response to individual search array items. Targets and distractors were represented in neural feature dimension maps and representations were stronger when defined by the preferred feature dimension of each region. Furthermore, the time course of feature-selective responses was consistent with behavioral response times. However, target responses in parietal cortices (IPS0/IPS1) correlated most strongly with behavior, suggesting that IPS0/IPS1 may be an integrated priority map and ultimately determine what is important in the visual field. These results indicate that neural feature dimension maps are crucial for computing attentional priority and that activation profiles in parietal regions guide behavior.

Funding: Alfred P Sloan Research Fellowship, National Eye Institute R01-EY035300

TALK 2, 11:00 AM, 42.12

DISTINGUISHING SERIAL AND PARALLEL SEARCH USING NEURAL SIGNATURES

Güven Kandemir^{1,2}, Docky Duncan^{1,2}, Dirk van Moorselaar³, Jan Theeuwes^{1,2,4}, ¹Vrije Universiteit Amsterdam, ²Institute Brain and Behavior Amsterdam (iBBA), ³Universiteit Utrecht, ⁴William James Center for Research, ISPA-Instituto Universitario

How do neural responses differ when one searches for a red rose among red poppies or white tulips? When a target is highly salient, as in a red rose among white tulips, search can take place based on the salient feature. By processing all displayed elements in parallel, the target simply pops out, resulting in rapid detection. Conversely, when target-distractor similarity is high, as in a red rose among red poppies, each item must be processed individually or in small groups, slowing detection rate with increasing distractor count. Despite the long-standing distinction between parallel and serial search, neural correlates of these strategies have rarely been contrasted, since

interpretation is typically confounded by differences in visual displays. In this study, we contrasted neural correlates of search strategies for visually identical displays. For this, we biased participants toward parallel or serial search in subsequent blocks by varying the similarity of the target and the distractors, prompting a focus on feature or conjunction. Embedded among inducer trials, test trials were identical in all blocks and could be searched with either strategy. Behavioral analyses from 24 participants confirmed successful induction of distinct search modes. EEG decoding from 17 posterior electrodes revealed that attentional deployment differed across search conditions in test trials, with significant generalization between test and inducer trials using the same strategy. The target location was represented differently across strategies, including differences in scalp topography. Comparing these location representations revealed that during parallel, but not serial, search, observers switched strategy when the target was not detected early. In addition, we found condition-specific differences in the neural representation of the target itself, consistent with more advanced stages of processing. These findings demonstrate that parallel and serial search are strongly influenced by history effects even under identical visual stimulation, resulting in distinct neural dynamics during visual search.

European Research Council (ERC) advanced grant [833029–LEARNATTEND] and Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) SSH Open Competition Behaviour and Education grant [406.21.GO.034]

TALK 3, 11:15 AM, 42.13

WHAT IF YOU NEVER SEARCHED FOR THE SAME TARGET TWICE?

Jeremy Wolfe¹, Cailey Tennyson², ¹Brigham and Womens Hospital / Harvard Med, ²Brigham and Womens Hospital

In standard visual search experiments, observers typically look for the same target over many trials. Using these repeated searches, researchers have studied observers' search templates, intertrial priming, distractor suppression, etc. These are important research topics. However, in the real world, many searches are one-time events. After all, we do not search for the VSS registration desk 100 times in a row. Do the conclusions drawn from blocks of repeated search apply to one-time searches? We had observers search for 6-dimensional conjunction targets. Targets were whole items with an enclosed part. Whole and part each had a color, shape, and orientation. On each trial, targets shared a particular number of features (0-5) with all distractors. 24 observers were tested for 300 trials, 50% target present. Set sizes: 6, 9, & 12. On each trial in the unique target condition, a novel target was shown before the trial, never to be shown again. The comparison condition was a single target search with the same target repeated for all trials, again sharing 0-5 features with distractors on each trial. Observers had unlimited time to encode each unique target. Results: Present RTs and Miss

error rates were similar for unique and repeated targets. Absent RTs and False Alarms were higher in the unique condition, especially when targets and distractors shared 4 or 5 features. Perhaps observers simply guessed that targets must be present when they became very similar to distractors. A second experiment used only distractors that shared five features with each trial's target. RTs were significantly slower in the unique condition, but RT x set size slopes were not significantly steeper. False alarms were higher for unique targets. Conclusion: Observers pay a cost when targets are new on every trial, but the fundamentals of search behavior remain similar to classic repeated target tasks.

NIH-NEI EY017001

TALK 4, 11:30 AM, 42.14

FREQUENCY OVER TIME REVISITED: LEARNING SCHEDULES AND THE FATE OF EARLY BIAS

Thiti Chainiyom¹, Timothy F. Brady², Chaipat Chunharas^{1,3,4}; ¹Cognitive Clinical and Computational Neuroscience Center of Excellence, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand, ²Department of Psychology, University of California San Diego, San Diego, CA, USA, ³Division of Neurology, Department of Medicine, Faculty of Medicine, King Chulalongkorn Memorial Hospital, Thai Red Cross Society, Bangkok, Thailand, ⁴Chula Neuroscience Center, King Chulalongkorn Memorial Hospital, Thai Red Cross Society, Bangkok, Thailand

In a world where patterns keep changing, do things that are common early on gain a search advantage later – e.g., will an ex-partner always be easy to spot in a crowd, even long after you've stopped actively looking for them? Visual statistical learning shows that repeated exposure builds familiarity, yet it remains unclear whether early high-frequency items continue to benefit in search once the pattern shifts. We asked whether an early high-frequency target (A) keeps an advantage over a later high-frequency target (B) when total exposure and low-level features are matched. Twenty-four participants completed two visual search tasks using colored circle/square targets (A and B) with controlled luminance contrast. Each task contained 640 trials: 160 baseline trials and 480 experimental trials, divided into three phases. In Experiment 1, Phase 1 introduced an early frequency bias toward A, Phase 2 raised the frequency of both targets, and Phase 3 tested later performance. In Experiment 2, as in previous work, A was frequent, and B was rare in Phase 1; their roles were swapped in Phase 2, and both targets appeared on 50% of Phase 3 trials. We analyzed phase-wise d' and learning slopes. In Experiment 1, early A–B d' differences in Phase 1 showed only a weak, non-significant link to Phase 2 ($r = .32$, $p = .12$), but reliably predicted A–B differences in Phase 3 ($r = .59$, $p = .003$), suggesting that early frequency can leave a familiarity bias that benefits performance later. In Experiment 2, learning slopes were predicted by whether targets were currently frequent or rare; these effects were similar for both targets, and early

A–B d' did not reliably carry over once the frequency swapped. Together, the results suggest that temporal structure and learning schedule can shape how early frequency biases survive or are overwritten over time.

TALK 5, 11:45 AM, 42.15

SEMANTIC GUIDANCE DURING VISUAL SEARCH: THEMATIC FACILITATION VS. TAXONOMIC COMPETITION
Tim Mousseau¹, Joseph Nah², Sarah Shomstein¹, Joy J. Geng^{2,3}; ¹Department of Psychological and Brain Sciences, The George Washington University, ²Center for Mind and Brain, University of California Davis, ³Department of Psychology, University of California Davis

Objects are related to one another at different levels of semantic meaning. Taxonomic relationships arise from shared features and categorical membership, whereas thematic relationships stem from real-world co-occurrence. While semantic relatedness guides visual attention overall, it remains unclear how they differ. In three experiments ($N_s = 331$; 337; 20), participants searched for a cued target in an array of four objects. In the two semantic conditions, the target was paired with a taxonomically or thematically related distractor along with two unrelated distractor objects. In the neutral condition, all four objects were unrelated. In all experiments, participants identified the target significantly faster in the thematic (E1-3: $M = 1206$; 1203; 1120) than in the taxonomic ($M = 1329$; 1329; 1178) condition (all $p < .003$), and both semantic conditions were faster than neutral ($M = 1381$; 1404; 1232; all $p < .003$). Experiment 3 used eye-tracking to test when and how thematic and taxonomic relationships influence attention. Time-course analysis was conducted on fixation proportions toward targets, distractors, and the unrelated objects across 2 seconds. In the first 500 ms, participants fixated taxonomic distractors ($M = 0.118$) 3.5 times more often than unrelated distractors ($M = 0.033$), suggesting early attentional capture ($p < .001$). In contrast, fixations toward thematic distractors ($M = 0.055$) were comparable to unrelated objects ($M = 0.042$; $p = .069$), showing that thematically related distractors were rarely fixated, despite facilitating target-search. Critically, participants were 2 times more likely to fixate on the taxonomic than thematic distractors ($p < .001$). In summary, thematic distractors facilitate attention toward targets without being looked at directly, suggesting that naturally co-occurring objects form perceptual groups that can be detected without overt visual attention. Taxonomic distractors, however, competed with the target by spreading attention between objects, perhaps due to shared low-level features that could not be rapidly discriminated.

TALK 6, 12:00 PM, 42.16

UNLIKE COLOR, SHAPE IS NOT AN INTUITIVELY UNDERSTOOD DIMENSION FOR ATTENTIONAL GUIDANCE

Haiye Liu¹ (haiye2@illinois.edu), Jun-Ming Yu¹, Alejandro Lleras¹, Simona Buetti¹; ¹University of Illinois Urbana-Champaign

Visual search becomes harder as the target becomes more perceptually similar to surrounding distractors. Previous work has quantified this relationship using logarithmic search slopes, which increase as target–distractor similarity increases. However, it remains unclear to what extent subjective similarity judgments reflect the objective search difficulty measured in search tasks. Here, search slopes were obtained for a wide range of color pairs (hue differences of $\pm 15^\circ$ to $\pm 70^\circ$) and 10 distinct shape pairs (including house-down, triangles, diamonds, circles, squares, and stars). A new group of participants rated the perceptual similarity of these same color and shape pairs on a 0–100 scale, enabling a direct comparison between subjective similarity structure and previously measured search efficiency. Color similarity showed a strong and highly predictive relationship with search performance ($R^2=.87$), indicating subjective judgments of color similarity are well calibrated to the effectiveness of color signals in guiding search. In contrast, shape similarity showed only a weak, nonsignificant trend ($R^2=.31$) suggesting that subjective shape similarity judgments do not track well search difficulty. Importantly, comparable search slopes were associated with very different levels of subjective similarity for color versus shape. For example, a distractor color at a relatively close distance to the target yielded a slope of 104 ms/log-unit and was rated as subjectively quite similar (62/100) to the target color. Yet a five-point star distractor with a nearly identical slope was rated very subjectively dissimilar to the target (17/100). These findings demonstrate that color provides a more robust and perceptually aligned signal for guiding visual search than shape, revealing an important asymmetry in how feature dimensions contribute to attentional guidance. Subjective similarity, it turns out, is not always an accurate window into the efficiency of perceptual processes. Observers may sometimes have wrong intuitions about what feature dimension should be more helpful during search.

TALK SESSION: MONDAY, MAY 18, 2026, 10:45 AM – 12:15 PM, TALK ROOM 2

Visual Memory: Encoding, retrieval, capacity

Moderator: Ming Meng, UAB

TALK 1, 10:45 AM, 42.21

REVISITING ICONIC MEMORY WITH A NEW EXPERIMENTAL PARADIGM

Gal Vishne^{1,*}, Miranda Ye^{2,*}, Nicholas B. Turk-Browne³, Michael N. Shadlen^{1,#}, John Morrison^{1,2,#}; ¹Columbia University, ²Barnard College, Columbia University, ³Yale University

We experience the visual world as rich and detailed, yet working-memory studies suggest we retain only a few items at a time. Sperling's seminal 1960s experiments provided the first evidence for a higher capacity short-term buffer, later dubbed "iconic memory." In those studies, participants were flashed a 4x3 letter array and were cued after its disappearance to report only one row, which substantially improved performance relative to a full-report condition. Despite the theoretical importance of this topic, research on iconic memory has slowed, partly due to limitations of the original paradigm: reporting multiple items leads to output interference; identifying letters amplifies processing interference; some letters are easier to identify than others; and there is no way to measure the precision of memories. Here, we introduce a new paradigm designed to overcome these limitations. On each trial, participants were flashed (50 ms) nine colored dots evenly spaced on a circle (4° radius). For uniform perceptual spacing, colors were sampled from an equiluminant circle in perceptual CIELAB space (center $a=20$, $b=38$, radius=60). After a variable delay (0–1000 ms), the fixation changed color to match one of the dots, and participants made an eye-movement to the location of that color. This design offers several advantages, among them: one response per trial avoids output interference; as low-level continuous features, colors create less processing interference and support estimates of memory precision. We found that across trials overall performance declined with array-to-probe delay. Modeling error patterns as a mixture of uniform and von-Mises distributions (following Zhang & Luck, 2008) revealed two distinct mechanisms behind this decay: increase in guessing rate and a reduction in memory precision. These findings demonstrate the prevalence of very short-term, high-capacity visual memory across tasks and provide a new platform for a revival of research on iconic memory.

TALK 2, 11:00 AM, 42.22

SEMANTIC UNDERSTANDING, NOT JUST COHERENT OBJECT STRUCTURE, STRENGTHENS VISUAL WORKING MEMORY

Yong Hoon Chung¹ (yong.hoon.chung.gr@dartmouth.edu), Sam Jung¹, Timothy Brady², Viola Störmer¹; ¹Dartmouth College, ²University of California San Diego

What constrains visual working memory capacity? Recent work shows that memory for real-world objects is markedly better than for abstract stimuli (e.g., colored squares or scrambled objects; Chung et al., 2024). This raises an important, yet unanswered, question: is this advantage driven by perceptual differences between these stimuli –

real-world objects have cohesive form structure and are perceptually more distinct than abstract stimuli – or does it instead reflect the incorporation of semantic knowledge into visual working memory? Here, we test this by comparing memory performance and neural maintenance activity for real objects with counterfeit objects generated by generative adversarial networks (Cooper et al., 2023). Importantly, we confirmed with norming studies that counterfeit objects preserved the object form and visual similarity structure of their real counterparts, only varying in their recognizability. If the real-object advantage is purely perceptual, then matching objects on form and similarity should preserve memory benefits. In contrast, if semantic knowledge contributes to storage, then disrupting recognizability while preserving visual structure should reduce memory performance and reduce neural maintenance signatures. Strikingly, enhanced memory performance emerged only for real objects, with counterfeit objects resulting in the same performance level as scrambled control stimuli (Exp. 1, $N = 50$). Human electrophysiological results further revealed increased neural engagement for real-world objects, expressed in greater CDA amplitudes and more stable patterns over time across encoding and maintenance periods (Exp. 2, $N = 24$). Finally, correlation analyses (Exp. 3, $N = 300$) showed that subjective familiarity of each object predicted memory for real objects, whereas low-level features, specifically colourfulness, predicted memory for counterfeit objects, suggesting distinct mechanisms underlie memory performance across stimulus types. Overall, these data show that visual working memory – long viewed as a primarily sensory system – is strongly shaped by higher-level semantic structure for complex and naturalistic stimuli.

TALK 3, 11:15 AM, 42.23

VISUAL PATTERN COMPLETION OF OCCLUDED OBJECTS PROMOTES ROBUST VISUAL LEARNING

Rosemary Cowell¹, Jennifer Gove¹, David Huber¹; ¹University of Colorado Boulder

Recognizing an occluded object is effortless: visual cortex fills in the missing parts to infer a recognizable whole. Across six behavioral experiments and an fMRI study, we tested whether part-to-whole visual pattern completion constitutes not just a retrieval operation, but a learning mechanism that strengthens object representations. The behavioral experiments employed a “peephole” learning paradigm: first, participants studied whole objects, then each object was assigned either to retrieval practice, cued by a circular aperture offering a peephole view, or to restudy. At final test, retrieval-practiced objects were remembered better than restudied objects – a visual testing effect. The task minimized the utility of verbal/semantic information, suggesting that this testing effect occurs without semantic mediation, contrary to several influential theories. Furthermore, when retrieval no longer required pattern completion—because the peephole cue was removed—the testing effect disappeared, suggesting that pattern completion was responsible for learning. This provides the first

evidence for a non-semantically mediated testing effect for visual stimuli, driven by occlusion-induced pattern completion. To test directly whether pattern completion in visual cortex underpins this novel testing effect, we took the task into the fMRI scanner. We recorded brain activation during peephole-cued recall. We also ran “template scans,” in which participants viewed whole objects. We trained a classifier to map template patterns from visual cortex onto object labels (e.g., lawnmower, clock, car). Next, we took “recall” brain patterns elicited by peephole cues (e.g., of a clock) and asked the classifier to predict the item being recalled. Item prediction was above chance, indicating that visual cortex performs pattern completion during peephole-cued recall. Moreover, classifier accuracy was higher for retrieval-practiced items than restudied or novel items. We conclude that visual pattern completion underpins the observed testing effect and serves a broader function than recognizing occluded objects, providing a potent mechanism for visual learning.

TALK 4, 11:30 AM, 42.24

PUSHING THE ENVELOPE: ACTIVE VIEWING DRIVES BOUNDARY EXTENSION

Hong B Nguyen¹ (nguyh376@newschool.edu), Benjamin van Buren¹; ¹The New School

When we actively move to view a scene, planning those movements requires predicting what we will see next. Sensorimotor predictions lie at the core of the “active learning” advantage, wherein observers better remember information if they can control what they see, and when they see it. Compared to passive viewing, active viewing may increase memory precision because it better aligns incoming visual input with the observer’s current preparatory and attentional state. Here, we asked whether visual predictions generated during active viewing not only enhance memory for what was seen, but also bias memory toward details which were never visible. In Experiment 1, observers viewed scene images by either using their mouse to actively move a small viewing aperture, or passively watching a replay of another participant’s viewing trajectory. After a 3-s viewing period, they completed a recognition memory test in which they discriminated the final view visible through the aperture from a decoy view which was slightly zoomed-in or zoomed-out. Active viewers showed robust boundary extension, more frequently selecting zoomed-out decoys than zoomed-in decoys; passive viewers showed the opposite pattern, consistent with boundary contraction. If active viewing drives greater boundary extension due to motor predictions, active viewers should specifically misremember having seen information in the direction their viewing trajectory was headed. In Experiment 2, we displaced the decoy view forward or backward relative to the participant’s last heading direction. Active viewers’ memory was biased toward unseen information in front of their viewing trajectory, and this bias was stronger than for passive viewers. Thus active exploration drives memory toward predicted future views—not just any unseen content. Together, these experiments demonstrate that sensorimotor

predictions cause us to actively fill in what lies beyond a visible boundary. We conclude that active viewing constructs richer, yet more inferential, memory representations than passive viewing.

TALK 5, 11:45 AM, 42.25

ROBUSTNESS OF PREDICTION BASED FALSE MEMORY

Mahveen Salman Mubarak¹ (mahveen.mubarak@mail.utoronto.ca),
Saba Halabisaz¹, Olya Bulatova², Caroline Yuan¹, Keisuke Fukuda^{1,2};
¹University of Toronto Mississauga, ²University of Toronto

To make sense of the world in real time, the human brain predicts incoming visual events. These predictions, based on learned environmental regularities, enable efficient integration of visual information despite inevitable neural processing delays. Prior work has shown that events that confirm or violate our predictions are well remembered (Atienza et al., 2011; Bein et al., 2021). However, it remains unclear how predictions that are never confirmed or violated influence visual memory. To test this, we had participants perform a color memory task in which they first saw a gray silhouette of a real-world object and were asked to explicitly predict what color it would become using a continuous color wheel. After the prediction, the silhouette either turned into 1) the same color as the prediction (confirmation), 2) the opposite color (180° offset, large violation) or 3) a similar color (45° offset, small violation) to their prediction thus rebutting their prediction, or 4) remained colorless, thus neither confirming or rebutting their predictions. Critically, participants were instructed to remember the actual color of the object, not their prediction. Subsequent memory recall revealed that making an explicit prediction was sufficient to elicit a false memory of the predicted outcome when their prediction was not confirmed. Furthermore, direct rebuttal of the prediction did not eliminate this prediction-based false memory, even with a large violation. A series of follow-up experiments revealed that the prediction-based false memory cannot be explained by action selection (Experiment 2) or pre-existing schemas (Experiment 3), and it is robust against overnight decay (Experiment 4), unlike the memory of the actual outcome. Together, these findings demonstrate the strong influence of explicit predictions on what we remember, underlining the powerful role of expectations in shaping episodic memory.

This research was supported by the Natural Sciences and Engineering Research Council (RGPIN-2024-05727).

TALK 6, 12:00 PM, 42.26

THE FACTORS THAT DETERMINE TRIAL-BY-TRIAL MEMORY BEHAVIOR ACROSS 6 MILLION TRIALS

Wilma Bainbridge¹ (wilma@uchicago.edu); ¹University of Chicago

It is well known that many factors influence memory performance, including participant factors (e.g., individual differences, memory capacity), stimulus factors (e.g., memorability, distinctiveness), and contextual factors (e.g., local image context, trial order). However, little work has compared how important each factor is in determining memory. To test this question, I analyzed 6,223,953 memory decisions made by 13,946 participants on a continuous recognition memory task with 26,107 object images. I examined how 12 different factors commonly discussed in the memory field (e.g., memorability, image context and distinctiveness, previous trial performance, repetition lag, trial number, etc) are predictive of trialwise memory performance. These factors combined explained over half the variance in memory behavior (61.2%). The participant and the image's memorability were the most impactful determinants of memory, with little variance captured by contextual factors, like the experimental distinctiveness of an image. The impact of memorability remained strong even when modeling memorability as predictions from a neural network (ResMem) independent from the experimental context, suggesting this is truly a stimulus-specific effect, rather than something about the stimulus set. Next, I ran 1000 mini-experiments within the data, testing whether the influence of these factors changes when looking at more homogenous or heterogeneous participants or contexts. More diverse participants resulted in less of an impact of context but an unchanged impact of memorability. More diverse images also did not change the impact of memorability on memory performance. Finally, I tested the factors that contribute to false alarms, and discovered that these are influenced by a wider variety of factors (intrinsic false alarmability, but also contextual and participant factors). In sum, this work demonstrates that surprisingly a lot of memory is predetermined and context-independent. I advocate for more models integrating stimulus, participant, and contextual factors in conjunction.

W.A.B. was supported by the Alfred P. Sloan Foundation and the National Science Foundation under CAREER Grant No. 244171

TALK SESSION: TUESDAY, MAY 19, 2026, 8:15 – 9:45 AM, TALK ROOM 1

Visual Search: Real-world scenes, objects

Moderator: Geoffrey Woodman, Vanderbilt University

TALK 1, 8:15 AM, 51.11

PROCESSING OF SCENE SUMMARY STATISTICS IS RAPID AND AFFECTS FEATURE-BASED ATTENTION

Jessica N. Goetz¹, Mark B. Neider¹; ¹University of Central Florida

Numerous studies have suggested that the visual system relies on the rapid processing of summary scene statistics to dynamically tune selective attention mechanisms during visual search (e.g., Becker et al., 2025; Rosenholtz, 2024). We have demonstrated that primary color features are used to guide attention more than secondary color features when observers search for real-world objects defined by more than one color (Goetz & Neider, 2025a). However, we have not investigated whether these effects translate to scenes. In the current study, 27 participants were shown a 200 ms target preview on a white background followed by a search array presented against a phase-scrambled real-world scene. We used the MATCH toolbox (Goetz & Neider, 2025b) to determine the primary and secondary colors in the scenes and real-world object targets. We selected scenes that were close in color to the target's primary color (primary scene), secondary color (secondary scene), and scenes far from the target's primary color (control scene). The toolbox was used to create color-matching distractors to the target's primary and secondary colors. Additionally, there were distractors that matched the target's shape but differed from the target's color. We predicted dynamic tuning of selective attention based on each scene such that the central guiding feature would differ depending on the specific elements of each scene. First saccade served as a proxy to measure visual selection. The results indicated that in control scenes, both colors were used to guide attention (all $ps < .001$), whereas colors and shape were used to guide attention in secondary scenes (all $ps < .001$). Critically, in primary scenes, the secondary color was used most to guide attention (all $ps \leq .007$). The data suggest that search templates are malleable and sensitive to the rapid processing of scene properties, specifically the color summary statistics in a scene.

TALK 2, 8:30 AM, 51.12

VISUAL SEARCH FOR REAL-WORLD OBJECTS IS SHAPED BY THE VARIABILITY OF OBJECT-CATEGORY TEMPLATES

Susan Ajith¹ (susan.ajith@psychiat.med.uni-giessen.de), Daniel Kaiser^{1,2,3,4}, Lu-Chun Yeh¹; ¹Neural Computation Group, Department of Mathematics and Computer Science, Physics, Geography, Justus Liebig University Giessen, ²Center for Mind, Brain and Behavior (CMBB), Universities of Giessen, Marburg, and Darmstadt, ³Center for Applied Computer Science and Data Science (ZAD), Justus Liebig University Giessen, ⁴Cluster of Excellence "The Adaptive Mind", Universities of Giessen, Marburg, and Darmstadt

Object categories vary widely in their appearance. While some are defined by relatively narrow feature configurations across exemplars (e.g., doors), others yield vast feature variations (e.g., bags). When searching for an object, such differences in diagnostic variability could shape the precision of target templates, in turn affecting search efficiency: Categories with low variability should yield narrow templates and efficient search, whereas highly variable categories

should produce coarser templates and less efficient search. Here, we used human drawings as a window into the variability of target templates, where more diverse drawings within a category indexed greater template variability. In two experiments, we then examined the effects of this variability on visual search performance. In Experiment 1, we selected 200 real-world object categories and quantified variability in their exemplar drawings from the THINGS drawings dataset using a deep neural network (DNN). Participants then performed a cued visual search task in which they indicated whether the target category was present or absent. Critically, search performance correlated with object variability: the greater the variability, the longer it took participants to find the target or report its absence. In Experiment 2, we selected 12 real-world object categories and quantified their variability at the individual level. Participants drew four exemplars per category before performing the cued search task. On the group level, object variability again affected search performance, with less efficient search for objects with variable drawings. By computing similarity between each participant's drawings and individual search targets, we found evidence for prioritised templates based on drawing order: participants' first drawing was the strongest predictor of search performance. Further, individuals' own drawings predicted their search better than other participants' drawings, revealing that individual differences in template variability shape search efficiency. Together, our findings provide robust evidence that inherent variability in object category templates shapes search behavior.

SA is funded by the JLU graduate scholarship. LCY is supported by the MSCA programme (101149060). DK is supported by the DFG (SFB/TRR135,222641018; KA4683/5-1, 518483074, KA4683/6-1, 536053998), "The Adaptive Mind", and an ERC Starting Grant (PEP, ERC-2022-STG 101076057)

TALK 3, 8:45 AM, 51.13

SALIENCE BIAS DURING VISUAL SEARCH IN NATURAL SCENES

Hyunwoo Gu¹, Justin L. Gardner¹; ¹Stanford University

Classical attentional capture studies using simplified search arrays have shown that goal- and salience-driven systems can be placed in conflict, making search less efficient. However, unlike artificial arrays, salience and target-relevant features in natural scenes show complex statistical relationships that vary between concordance and conflict. This raises a key question about which system exerts greater influence on gaze selection, and how their influences evolve under conflict in natural scenes. We analyzed published natural visual search datasets using saliency models and vision-language model template matching, finding that human fixation choices are biased by salience features. Furthermore, search templates directly estimated from gaze patterns recapitulated this bias, showing a tendency to be similar to salience templates derived from free-viewing. Using deep stimulus synthesis,

we further demonstrated that this salience bias can be used to control gaze selection patterns during visual search. Consistent with classical attentional capture, we found that mismatch between salience and target features lowers search performance. However, capture was not static; we observed that salience bias was strongest during initial fixations and attenuated as search progressed. To further characterize these dynamics, we fit an observer model that chose fixations from either salience or target maps at each step. We found that the initial probability for salience was higher but the probability of remaining in the target mode increased, indicating a tendency to switch from salience to target drive as search progressed. Finally, analyses of natural stimuli revealed that salience features alone yield above-chance search performance. This stands in contrast to classical search arrays, which are often designed to make salient distractors task-irrelevant, thereby motivating observers to suppress them. Thus, in natural scenes, salience serves as an informative signal that helps constrain search, justifying the initial bias despite the potential for capture.

TALK 4, 9:00 AM, 51.14

WHERE IS MY CEREAL? IMPACTS OF TARGET SALIENCY AND BACKGROUND COMPLEXITY ON VISUAL SEARCH AT HOME

Qingying Gao¹ (kaiagaogy@gmail.com), Aleks Mihailovic¹, Pradeep Ramulu¹, Yingzi Xiong¹; ¹Johns Hopkins University

Efficient visual search in a scene relies on both bottom-up features such as target saliency and number of distractors, and top-down factors such as prior knowledge of target locations. At familiar places such as homes, do bottom-up features still matter when we look for common objects? This question is especially relevant to individuals with vision impairment, who show slower visual search than those with normal vision, even at familiar home environments. Thirty-two participants with vision impairment performed three everyday tasks at their homes, including paying bill, preparing meal, and taking medicine. Each task requires sequential search of predetermined objects (e.g., bowl, cereal, milk, and spoon for preparing meal), and object-level search time was recorded. Photos of home environments containing each object were captured, and target saliency and background complexity were quantified through two methods: 1) computer vision (CV)-based analysis, with target saliency quantified by root-mean-square contrast, color diversity, and itti's saliency; and background complexity quantified by clutteriness, scene congestion, and object density, and 2) human experts' ratings based on a clinically-developed rubric (1-5 for saliency and clutteriness, respectively). Across participants, the search time ranged from 3.97s for medicine to 25.78s for stamp. For CV-based analysis, target saliency, but not background complexity, showed small but significant associations with search time. Higher contrast ($r = -0.14$), color diversity ($r = -0.12$), and itti's saliency ($r = -0.17$) were associated with shorter search time ($p <$

0.05). For human expert-rated features, however, background complexity, but not task saliency, was significantly associated with search time ($r = 0.29$, $p < 0.001$). Our results provide evidence that scene features affect visual search by individuals with vision impairment, even in familiar home environments. Computer vision models and human experts each have their strengths and limitations in estimating environmental features that best predict visual search time.

TALK 5, 9:15 AM, 51.15

AN INVESTIGATION OF THE EFFECTS OF TARGET LONG-TERM MEMORY STRENGTH ON HYBRID SEARCH

Raeavan L. Hanan¹, Melissa R. Beck¹, David A. Tomshe¹; ¹Louisiana State University

Hybrid search requires observers to conduct a working memory task, locating a target in a visual display, while retrieving multiple potential targets from long-term memory (LTM). Prior work has demonstrated that the availability and precision of LTM representations are shaped by exposure, encoding quality, and learning criteria. Further, the availability and precision of LTM representations can meaningfully influence search performance. The present study examines whether strengthening LTM representations of target items improves hybrid search performance. Participants ($N = 144$) were assigned to either baseline ($n=72$) or high ($n=72$) LTM strength conditions. The high-strength condition included more exposure to the items during familiarization phase and required more detailed representations and a higher accuracy threshold to progress from the familiarization to the search phase. These requirements for familiarization encourage more precise and more readily accessible target templates for the high-strength group relative to the baseline group. The strengthened templates should facilitate retrieval into WM, increasing search accuracy and potentially reducing the time for template-to-object comparison during search. The high-strength group ($M = 0.90$, $SD = 0.04$), demonstrated higher search accuracy than the baseline group ($M = 0.85$, $SD = 0.09$), $t(105) = -4.57$, $p < .001$. For search response time, the high memory strength group ($M = 4661.68$, $SD = 785.63$), when compared to the baseline ($M = 4941.46$, $SD = 945.04$), revealed only a marginal effect, $t(137) = 1.93$, $p = .055$. This pattern of effects indicates that improved LTM strength impacts primarily search accuracy rather than search speed, suggesting that target LTM strength impacts target identification during search, but not search guidance. These results highlight the value of strengthening LTM representations for improving accuracy in applied search tasks where correct identification is critical (i.e., security distinguishing contraband from benign items, pharmacists identifying medications, or criminalists recognizing objects during evidence recovery).

TYPO DETECTION AS A WINDOW INTO INTERACTIONS BETWEEN VISUAL SEARCH AND READABILITY

Emily Heffernan¹ (emily.heffernan@mail.utoronto.ca), Kiran Panicker¹, Anna Kosovicheva¹, Benjamin Wolfe; ¹University of Toronto

When scanning this VSS abstract for (hopefully nonexistent!) spelling errors, the reader must balance speed and accuracy. As a form of visual search, typo detection relies on both reading and attentional guidance, but the extent to which the visual properties of text and visual saliency impact this process has received limited attention. In three experiments, we explored how text appearance and attentional (mis)guidance influence search dynamics. In Experiments 1 and 2, participants (N = 17 each) scanned “pseudoparagraphs” of random words for typos in a single-target visual search paradigm. Words were presented in either hard- or easy-to-read fonts. The “hard” font condition used a high stroke contrast in Experiment 1 and a condensed font in Experiment 2. For both experiments, accuracy remained high in each condition, but participants took longer to respond after fixating the typo in the hard-to-read condition. Drift diffusion modelling isolated this change to slower evidence accumulation for the hard-to-read fonts; nondetection time and decision threshold were unchanged across conditions. These findings demonstrate that manipulations to text appearance can interrupt processing efficiency. In Experiment 3, we studied how attentional guidance impacts search in a typo foraging task with 0–4 typos per pseudoparagraph. Pilot participants (N = 9) completed two conditions: a baseline task and a second with imperfect “spellchecker” annotations. Participants were faster and more accurate with annotations. However, eye tracking data indicated a shift in strategy—with annotations, participants had shorter fixations and fixated fewer words. In other words, annotations generally improved search outcomes but decreased quitting thresholds. This novel paradigm highlights how stimulus properties can influence search, with poor readability hampering evidence accumulation and reliance on automated aids impacting quitting threshold. These results point to shared processes across different forms of visual search but also highlight the importance of readability for text searches.

Acknowledgments: This work was supported by a SSHRC Insight Grant to BW and AK.

TALK SESSION: TUESDAY, MAY 19, 2026,

8:15 – 9:45 AM, TALK ROOM 2

Functional Organization of Visual Pathways 2

Moderator: Niko Kriegeskorte, Columbia University

CORTEX: A UNIFIED FRAMEWORK FOR EVALUATING THE PREDICTIVE AND COGNITIVE ALIGNMENT OF VISION MODELS ENTIRELY ZERO-SHOT

Ruolin Wang^{1,2}, Mayukh Deb^{1,2}, Alex Abate³, Alish Dipani^{1,2}, Sanjana Chillarege^{1,2}, Kruthik Ravikanti^{1,2}, Kushal Dudipala^{1,2}, Yuxuan Li^{1,2}, Haider Al-Tahan^{1,2}, Ranjani Koushik^{1,2}, Herrick Fung^{1,2}, Yung-Ying Chen^{1,2}, Nikolas McNeal^{1,2}, Nancy Kanwisher^{4,5,6}, N. Apurva Ratan Murty^{1,2}; ¹Cognition and Brain Science, School of Psychology, Georgia Institute of Technology, ²Computational Cognition, Georgia Institute of Technology, ³Harvard Medical School, ⁴Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, ⁵McGovern Institute for Brain Research, Massachusetts Institute of Technology, ⁶The Center for Brains, Minds and Machines, Massachusetts Institute of Technology

Our understanding of the visual system has relied on hypothesis-driven studies and tightly controlled stimuli to adjudicate between theoretical predictions. In contrast, modern NeuroAI benchmarks prioritize aggregate metrics, such as how well artificial neural networks (ANNs) predict responses to large, uncontrolled datasets. But prediction alone offers limited insight into whether models capture the experimental phenomena and foundational theories of the field. Here we introduce Cortex, a framework for evaluating models of functional ROIs entirely zero-shot (without model retraining). We leveraged Cortex to evaluate 126 leading ANN-based encoding models of category-selective cortex (FFA, PPA, and EBA) on their ability to predict brain responses and their capacity to recapitulate influential findings in the field. Our assessments spanned 8 fMRI datasets and 20 experimental studies. On prediction tests we found that the choice of fMRI mapping data was critical: models mapped using rich naturalistic datasets (like NSD) generalized better than models mapped using smaller datasets. However, systematic gaps also emerged as these models did not predict responses to synthetic or dynamic video stimuli. We also observed remarkable variations between fROIS: EBA was less predictive and harder to model than the FFA or PPA. We also found that vision–language models (like CLIP-ResNet50) tended to generalize better across datasets, as previously reported. Finally, we evaluated how well models replicated prior cognitive neuroscience studies. Here we found that even the best predictive models frequently failed to replicate specific experimental findings. For example, CLIP-ResNet50 that consistently ranked highly on prediction tests failed to reproduce patterns like retinotopy or any of the EBA-specific experimental results. Collectively, these findings show that prediction is a necessary but incomplete proxy for understanding vision. Cortex helps bridge this gap by providing a rigorous in silico platform for hypothesis-testing, exposing key model failures that can serve as a roadmap for future model development.

This work is supported by the NIH Pathway to Independence Award (R00EY032603), the NSF Nexus computation support (Allocation number: SOC250049), and a startup grant from Georgia Tech.

TALK 2, 8:30 AM, 51.22

A RESOURCE OF NEURAL ENCODING MODELS FOR IN SILICO VISUAL NEUROSCIENCE

Alessandro Gifford¹, Domenic Bersch², Daniel Janini¹, Gemma Roig², Radoslaw Cichy¹; ¹Freie Universität Berlin, ²Goethe Universität Frankfurt

In silico neural responses to visual stimuli generated by encoding models increasingly resemble in vivo responses recorded from real brains, enabling the novel research paradigm of in silico visual neuroscience. The fast and economical generation of in silico neural responses allows researchers to test more scientific hypotheses and to explore across larger solution spaces than possible in vivo. Crucially, novel findings from large-scale in silico experimentation are then validated through targeted small-scale in vivo data collection, thereby optimizing research resources. To empower this emerging research paradigm, we introduce the Brain Encoding Response Generator (BERG; <https://github.com/gifale95/BERG>), a resource consisting of diverse pre-trained encoding models of the brain and a Python package to easily generate in silico neural responses to arbitrary visual stimuli. BERG enables researchers to efficiently address a wide range of research questions through in silico visual neuroscience by providing a growing, well documented library of encoding models trained on different neural recording modalities, species, datasets, subjects, and brain areas. We demonstrate BERG's potential for neuroscientific discovery in four ways. First, BERG's encoding models accurately predict neural responses to visual stimuli. Second, these in silico responses reproduce key signatures of visual processing in the brain, such as retinotopy and categorical selectivity (fMRI), or the different dynamics of object exemplar versus concept categorization (EEG). Third, BERG enables data types that are impossible to collect in vivo, such as high spatio-temporally resolved in silico neural responses derived through linear mapping of EEG onto fMRI. Fourth, in a separate study we used BERG to reveal representational relationships between visual areas, which we successfully validated in vivo (Gifford et al., 2025, NatHumBehav). Together, we envision that BERG will empower in silico visual neuroscience, ultimately accelerating scientific discovery. We warmly welcome models, ideas, and collaboration from the vision science community.

TALK 3, 8:45 AM, 51.23

DIFFUSION-BASED STIMULUS OPTIMIZATION REVEALS FUNCTIONAL ORGANIZATION ACROSS HIGHER VISUAL CORTEX

Margaret M Henderson¹ (mmhender@cmu.edu), Andrew F Luo², Sungjoon Park¹, Michael J Tarr¹, Leila Wehbe¹; ¹Carnegie Mellon University, ²University of Hong Kong

Characterizing the fine-grained functional organization of human higher visual cortex remains a significant challenge. Traditional neuroimaging experiments are limited in the number of stimuli they can sample, which may bias results toward particular stimulus attributes. In prior work we developed a novel data-driven tool, termed "BrainDiVE" (Luo et al. 2023, NeurIPS), which addresses these challenges by synthesizing images optimized to activate specific brain regions. BrainDiVE leverages pretrained image diffusion models guided by gradients from an image-computable fMRI encoding model. Here, we validated BrainDiVE experimentally by generating images that targeted several functional regions of interest (i.e., images predicted to maximally activate those brain areas), and presenting them to new human participants (n=12) in an fMRI study. We found that the synthesized images elicited robust and spatially specific responses in the predicted target regions, yielding significantly higher measures of category selectivity relative to natural images. This validates BrainDiVE's ability to capture neural selectivity in human ventral visual cortex, characterizing tuning properties that generalize across participants. Furthermore, we demonstrated fine-grained experimental control by differentially activating two face-selective regions, the occipital face area (OFA) and fusiform face area (FFA), providing further evidence that these regions encode distinct aspects of faces. We also identify a functional gradient within the occipital place area (OPA) along a posterior-to-anterior axis, suggesting a functional topology based on scene properties such as distance and indoor-outdoor. These findings provide new insights into the representational structure of category-selective regions and establish a novel paradigm for targeted exploration of neural selectivity in human visual cortex. More generally, our approach offers a powerful tool for investigating the functional organization of visual cortex at a fine-grained level, exceeding the capabilities of traditional methods across multiple dimensions.

Support was provided by a grant from Apple Inc. to MJT. This work used Bridges-2 at Pittsburgh Supercomputing Center through allocation SOC220017 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation.

TALK 4, 9:00 AM, 51.24

INTERPRETABLE FMRI-TO-IMAGE DECODING REVEALS HOW BRAIN REPRESENTATIONS GUIDE GENERATIVE VISUAL RECONSTRUCTIONS

Pinyuan Feng¹, Hossein Adeli¹, Wenxuan Guo¹, Fan Cheng¹, Ethan Hwang¹, Nikolaus Kriegeskorte¹; ¹Columbia University

Recent works have demonstrated that complex visual stimuli can be decoded from human brain activity. Current approaches first map fMRI signals into intermediate image or text features to then guide a generative model. While effective, this two-stage strategy introduces an information bottleneck and obscures how specific brain regions contribute to the final reconstruction. Here, we introduce NeuroAdapter, an end-to-end visual decoding framework that conditions a latent diffusion model directly on brain representations. Neural activity is transformed into parcel-wise tokens that modulate the generative process through transformer cross-attention, enabling a more transparent link between cortical signals and image generation. We show that our model achieves high-quality visual reconstructions on Natural Scene Dataset (NSD), NSD-Imagery and Deeprecon. We also created a baseline model using a state-of-the-art brain encoder that retrieved images from a large image dataset (ImageNet). On NSD, our model achieves 11.7% average improvement in low-level image quality metrics and 18.6% average improvement in high-level semantic metrics. Furthermore, the decoded images yield predicted neural responses that closely match the measured fMRI patterns, confirming that the reconstructions preserve the brain-relevant visual features. To understand how brain representations shape the unfolding generative trajectory, we further propose the Image–Brain Bi-directional Interpretability (IBBI) framework. IBBI analyzes cross-attention patterns across diffusion steps to show (i) the contribution of individual cortical parcels during denoising process, and (ii) the spatial influence of category-selective regions (e.g., face, body, scene, word region of interests) on evolving image-level features. Visualization of ROI attention maps reveals a systematic temporal pattern that ROI-level influence is broadly distributed across the image during early denoising stages but progressively converges onto semantically relevant pixel regions at later stages of reconstruction. Together, our work establishes an interpretable approach for end-to-end brain-to-image reconstruction, highlighting the potential of decoding to reveal the brain representation of complex scenes.

Research reported in this publication was supported in part by the National Institute of Neurological Disorders and Stroke of the National Institutes of Health under award numbers 1RF1NS128897 and 4R01NS128897. The content is solely the responsibility of the authors.

TALK 5, 9:15 AM, 51.25

CONTEXT-DEPENDENT DISSOCIATION OF SHARED INPUT AND DIRECTED FLOW IN THE VISUAL CORTEX

Yuxuan Xue^{1,2} (yuxuan.xue@yale.edu), Mitchell Morton³, Anirvan S. Nandy^{4,5,6,7}, Monika P. Jadi^{2,4,7}; ¹Department of Electrical Engineering, Yale University, ²Department of Psychiatry, Yale University, ³Interdepartmental Neuroscience Program, Yale University, ⁴Department of Neuroscience, Yale University, ⁵Department of Psychology, Yale University, ⁶Kavli Institute for Neuroscience, Yale University, ⁷Wu Tsai Institute, Yale University

Cortical population activity reflects a mixture of externally driven inputs and internally generated brain states, yet these sources of co-fluctuation can produce similar patterns of correlated variability. This ambiguity poses a fundamental challenge for interpreting whether population coupling reflects true information flow or shared modulation. To dissociate these mechanisms, we investigated how visual stimulation and wakefulness state differentially shape inter-laminar communication in the macaque primary visual cortex (V1). We recorded laminar spiking activity across input and superficial layers in two macaques during both stimulus-evoked and spontaneous (eyes open vs. closed) conditions, and quantified cross-population interactions using reduced-rank regression (RRR) to identify low-dimensional predictive subspaces linking the two layers. Prediction accuracy, principal angles between subspaces, and temporal delay analyses were used to assess the strength, structure, and directionality of inter-laminar coordination. To mechanistically interpret the empirical signatures, we constructed a two-layer recurrent neural network (RNN) with low-rank feedforward connectivity and manipulated the structure and layer-specificity of external inputs. Visual stimulation markedly modulated the inter-laminar predictive subspace and increased prediction accuracy in a directional manner: input-layer activity predicted superficial responses more strongly than undirected models. Additionally, optimal delays showed that input activity consistently preceded superficial activity. In contrast, eye-closed spontaneous activity increased prediction accuracy with no directional enhancement and no consistent temporal lead-lag structure. Simulations reproduced this dissociation: directionality emerged only when stimulus-driven inputs aligned with feedforward connectivity and when the superficial layer exhibited stronger low-dimensional structure relative to the input layer, whereas shared fluctuations produced undirected co-activation regardless of structure. These findings demonstrate that similar correlation patterns can arise from qualitatively distinct mechanisms. Sensory stimulation drives feedforward communication across V1 layers, whereas spontaneous internal dynamics produce shared modulation without directional flow. Our results establish predictive subspace structure as a principled tool for distinguishing communication from common input across cortical circuits.

This research was supported by NIH R01 EY034605, NIH R00 EY025026, NIH R21 MH126072 and SFARI 875855 to MPJ, NARSAD Young Investigator Grant, Ziegler Foundation Grant, Yale Orthwein Scholar Funds, NIH R01 EY032555, NIH R21 MH126072 and SFARI 875855 to ASN, and by NEI core grant for vision research P30 EY026878 to Yale University.

TALK 6, 9:30 AM, 51.26

META-LEARNING IN-CONTEXT ENABLES TRAINING FREE PREDICTION & DECODING OF VISUAL CORTEX FROM NOVEL SUBJECTS

Andrew Luo¹, Mu Nan¹, Muquan Yu¹, Margaret Henderson², Leila Wehbe², Michael Tarr²; ¹University of Hong Kong, ²Carnegie Mellon University

Developing computational models of human visual cortex that generalize across individuals remains a fundamental challenge in neuroscience, as the substantial variability in functional organization typically necessitates collecting large-scale datasets to train individual models for each subject. To address this challenge, we introduce a unified meta-learning framework that reformulates neural modeling as an in-context inference problem, enabling both the prediction and decoding of visual representations in novel subjects without any gradient-based fine-tuning. First, addressing the forward encoding problem, we present a transformer-based architecture that infers the unique functional tuning of individual voxels by simply conditioning on a small set of image-activation examples. This approach yields high-fidelity predictions of voxelwise responses in higher visual cortex that generalize robustly across diverse scanners, acquisition protocols, and subject populations using only a fraction of the data typically required. We then leverage this learned forward model to solve the inverse problem (decoding visual stimuli from brain activity) via a novel hierarchical in-context learning strategy. In the first stage of the hierarchy, the model estimates per-voxel visual response parameters in-context; in the second stage, we construct a context consisting of these estimated parameters and observed brain responses across multiple voxels to perform a "functional inversion" that directly infers the visual stimulus embedding. We demonstrate that this method achieves state-of-the-art cross-subject image retrieval without requiring anatomical alignment or shared stimuli, effectively bypassing the correspondence problem. Furthermore, attention-based analyses reveal that the model spontaneously learns to localize and differentially weight functionally specialized regions: such as face- and place-selective areas to maximize decoding accuracy, mirroring known cortical organization. Together, these findings establish a scalable, data-efficient foundation model for non-invasive brain activity prediction and decoding, offering a principled computational lens for investigating population-level neural representations without the constraints of subject-specific training.

TALK SESSION: TUESDAY, MAY 19, 2026, 10:45 AM – 12:15 PM, TALK ROOM 1

Visual Memory: Individual differences, imagery

Moderator: Joan Ongchoco, University of British Columbia

TALK 1, 10:45 AM, 52.11

A UNIQUE NEURAL SIGNATURE OF INDIVIDUAL DIFFERENCES IN LONG-TERM MEMORY FROM EEG INTER-ELECTRODE CORRELATION WITH EVENT-RELATED POTENTIALS

Chong Zhao¹, Edward K. Vogel¹, Monica Rosenberg¹; ¹University of Chicago

Classic memory models suggest that the processes supporting visual working memory (VWM) play a central role in determining how effectively information is encoded into long-term memory (LTM). Consistent with this view, VWM and LTM performance are typically correlated, leaving open the question of whether LTM performance relies on neural mechanisms that are distinct from those supporting VWM encoding. To address this, we recorded EEG activity while participants performed recognition memory tasks with large set sizes (32 and 128 items), far exceeding typical VWM capacity. Using interelectrode correlation (IC) analyses, we found that IC patterns robustly predicted individual differences in LTM performance across both set sizes, indicating a stable, domain-general neural signature of LTM formation. Crucially, the predictive power of our IC model persisted even after statistically controlling for VWM capacity and attentional control, demonstrating that the model captures variance specifically related to individual differences in LTM performance rather than general cognitive performance. Temporally, IC-based predictive signals emerged shortly after stimulus onset and remained significant for approximately 500–600 ms. Distinct correlation patterns characterized early versus late encoding windows, suggesting that multiple, dynamically shifting neural processes contribute to individual differences in LTM. Together, these results identify a reliable and temporally dynamic neural signature during LTM encoding that tracks individual differences in LTM performance, independent of VWM and attentional control abilities.

We acknowledge the funding from National Institute of Mental Health (grant ROIMH087214); Office of Naval Research (grant N00014-12-1-0972).

TALK 2, 11:00 AM, 52.12

HIPPOCAMPAL VOLUME AND FUNCTIONAL PATTERNS IN EARLY CHILDHOOD REFLECT THE DEVELOPMENT OF ADULT-LIKE VISUAL MEMORY PATTERNS

Xiaohan (Hannah) Guo¹, Yuan Chang Leong¹, Wilma A. Bainbridge¹; ¹The University of Chicago

The memorability effect refers to the phenomenon that adults reliably remember and forget the same visual stimuli as each other. This adult-like consistency emerges in childhood: 4-year-olds show adult-like memory patterns, whereas 3-year-olds remember the same stimuli within age but differ from adults (Guo & Bainbridge, 2024). This study tests whether neural development—particularly within the

hippocampus—drives the emergence of adult-like visual memory by age 4, given that adults' sensitivity to memorability is decodable in the ventral visual stream and medial temporal lobe, including the hippocampus (Bainbridge & Rissman, 2018). We ask: (1) whether brain activity tracks the emergence of adult-like visual memory in children by age 4, and (2) whether hippocampal volume and functional activity explain the degree to which children show adult-like memory patterns. Using an existing fMRI movie-watching dataset (ages 3–adult; $N = 155$; Richardson et al., 2023), we computed a memorability score for each movie timepoint using a deep neural network (ResMem) trained to predict memorability (Needell & Bainbridge, 2022). A cross-validated model trained on adults' brain data successfully predicted movie-frame memorability scores from children's neural data beginning at age 4, but not at age 3, suggesting that adult-like neural sensitivity to memorability emerges by age 4, replicating behavioral findings. Structural analyses indicated that age and hippocampal volume measures—but not overall brain volume—significantly predicted the adult-likeness of children's visual memory patterns. Functional searchlight analyses further revealed that the inferior temporal gyrus reflects movie-frame memorability from age 3 onward, while the hippocampus shows this sensitivity starting at age 5. Together, these results suggest that the emergence of adult-like visual memory patterns is driven by structural and functional development in the hippocampus and IT cortex around age 4.

TALK 3, 11:15 AM, 52.13

ADAPTIVE DEPLOYMENT OF WORKING MEMORY IN OLDER ADULTHOOD

Nir Shalev¹ (shalenir@gmail.com), Liz Atias¹; ¹University of Haifa, Israel

Working memory (WM) supports the maintenance of goal-relevant information during interaction with the environment. Standard laboratory tests consistently report age-related declines in WM, yet they often fail to capture how WM is used in the real world. In everyday situations, we rarely rely on our maximum WM capacity; instead, we balance between forming internal representations in WM and resampling information from the environment. For example, when following a cake recipe, we can either hold several ingredients in mind while searching the pantry or repeatedly look up each ingredient in the recipe. Understanding this adaptive allocation of WM resources is essential for explaining how memory decline affects people in everyday life and for linking lab-based findings with real-world performance. We tested whether younger (20–30 years) and older (65–75 years) adults differ in their flexible use of WM in an immersive Virtual Reality (VR) object copying task that approximates real-world activity. Participants ($N=100$) encoded the identity and location of items on a simulated model board, searched for targets among distractors, and placed them in matching slots. We manipulated movement demands to examine how reliance on WM changes as information sampling becomes more effortful. Older adults were

generally slower and relied on fewer items in memory, yet greater movement demands prompted both groups to adapt and use more memory. Adaptation patterns diverged: younger adults showed sharper increases in encoding time and memory use as movement demands rose, whereas older adults incurred larger costs in completion time. VR-based performance measures correlated with behaviour outside the task, supporting ecological validity: memory use was associated with cognitive ability, and completion time with functional mobility. These findings demonstrate adaptive WM use across adulthood and show how ecologically valid VR assessments can reveal age-related differences that conventional tests overlook.

Supported by the Israeli Science Foundation Research Grant (1073/24)

TALK 4, 11:30 AM, 52.14

PAIN EFFECTS ON VISUAL WORKING MEMORY ARE MODULATED BY THE INTERACTION BETWEEN INDIVIDUAL DIFFERENCES AND ATTENTIONAL PRIORITY

Phivos Phylactou¹, Angdi Chu², Mathieu Piché³, Siobhan Schabrun², David Seminowicz²; ¹University of Nevada, Reno, ²Western University, ³Université du Québec à Trois-Rivières

The effects of pain on cognition are often mixed, with evidence showing bidirectional pain effects on cognitive performance. This has led the field to propose a two-group categorization: individuals who perform better (A-types) in cognitive tasks under pain and individuals who perform worse (P-types). Alternatively, the bidirectional effects may instead be due to an interaction of pain with attentional priority rather than individual differences. To explore this alternative, we performed four experiments (total $n = 110$) that manipulated attentional priority via retro-cues in an orientation change-detection visual working memory (VWM) task, while participants experienced thermal pain. Experiment 1 ($n = 45$) indicated worse VWM performance when pain was introduced after compared to before attentional prioritization ($BF = 4.22$). However, when accounting for individual differences, evidence showed that pain effects were evident when introduced prior to attentional prioritization for A-types ($BF = 353.93$), but after attentional prioritization for P-types ($BF = 28.82$). These effects were absent during an auditory distractor control condition. To further explore whether the dissociable effects were better explained through attentional mechanisms or individual differences, Experiment 2 ($n = 20$) modeled VWM performance as a function of thermal pain intensity. Model fitting comparisons indicated that accounting for individual differences provides a better fit than attentional mechanisms but also showed better model fitting for P-types versus A-types ($BF = 3.49$). We investigated this further in Experiments 3 ($n = 25$) and 4 ($n = 20$) that showed that P-types are more susceptible to VWM pain effects, both during prioritized (Experiment 3: $BF = 4.76$) and unprioritized (Experiment 4: $BF = 24.81$) attentional states. Contrary to our

predictions, these findings illustrate that the nexus between pain and VWM is better understood through an interaction between individual differences and attentional prioritization, rather than one over the other.

This work was partly supported through the 2023 Postdoctoral Associate Recruitment Award of the Faculty of Health Sciences at Western University, awarded to PP and SMS

TALK 5, 11:45 AM, 52.15

CAN YOU IMAGINE IN COLOUR?: MEMORY CONFUSION AS AN INDEX FOR COLOUR IMAGERY VIVIDNESS

Zenith A. Zyn¹, Joan Danielle K. Ongchoco¹; ¹The University of British Columbia

Our perceptual experiences are full of colour—the red of an apple or the blue of the sky—that involve a dynamic interplay between surface reflectance, illumination, and viewing conditions. But consider the colour of your mental images. Some people report their mental images to be rich in colour, whereas others experience colours as dim, even nearly black-and-white. The present study develops an objective property-specific measure of colour imagery using a source memory confusion paradigm. Observers first completed a baseline imagery calibration in which they adjusted the lightness (i.e., L in HSL colour space) of four canonical colours (pink, cyan, orange, green) to match their mental images. These settings were used for abstract shapes that were presented either fully coloured (in the lightness provided by the observer) or merely outlined. In coloured trials, observers studied the shape for a period of time; in outline trials, they imagined “filling in” the shape with a cued colour for the same duration. At test, all shapes appeared coloured, and observers reported whether each shape was old or new—and, for recognized shapes, whether its colour had been seen or imagined. Finally, all observers completed the Vividness of Visual Imagery Questionnaire (VVIQ). Shape recognition was robust, confirming successful encoding for coloured and outlined shapes. But the critical question was whether observers confused outlined shapes as having been seen in colour, where successful imagery would predict greater source-memory confusion. Individual differences in source-memory confusion were predicted by colour-specific VVIQ items (with greater confusion in observers who responded the maximum rating of 5), but not by overall VVIQ scores and lightness settings. These results expand our notion of “vivid” mental imagery: imagery vividness may be better captured by independent visual features, and successful “vivid” imagery may blur (in memory) the line between what was perceived versus imagined.

TALK 6, 12:00 PM, 52.16

THE ROLE OF V1 FOVEA IN VISUAL MENTAL IMAGERY

Yangjianyi Cao¹ (ycao2@uab.edu), Yanna Mao², Shuai Chang³, Li Zhaoping⁴, Ming Meng¹; ¹University of Alabama at Birmingham, United States, ²South China Normal University, China, ³Zhongshan Ophthalmic Center of Sun Yat-Sen University, China, ⁴University of Tübingen, Max Planck Institute for Biological Cybernetics, Germany

Prior research on visual perception suggests that the foveal region of V1 exhibits strong top-down feedback (Zhaoping, 2023, Trends Cog. Sci.) and contains high-level information about objects located in the peripheral visual field (Williams et al., 2008, Nat. Neurosci.). Although visual mental imagery is the cognitive ability to voluntarily generate internal representations without sensory inputs, it activates remarkably overlapping brain regions with visual perception, particularly in the primary visual cortex (V1). However, it remains unknown whether and how the foveal region is involved in representing imagined targets located in the periphery. Using functional magnetic resonance imaging (fMRI), we measured neural activity in individuals with aphantasia and typical imagers during a voluntary imagery task. Participants were instructed to imagine a green-vertical or red-horizontal Gabor patch in their left or right visual field. We employed univariate analysis and multivariate pattern analysis (MVPA) within retinotopically defined V1 to assess BOLD signal amplitude and decode the combined spatial and feature information of the imagined targets. Results revealed that, unlike typical imagers who showed no significant activation in the foveal region of V1, individuals with aphantasia exhibited significant negative BOLD signal changes compared with the no imagery period. Crucially, this signal reduction did not compromise multivariate pattern decoding in the foveal region, as its accuracy remained significantly above chance level. This suggests that, despite the absence of subjective visual experience, aphantasic individuals may recruit alternative neural strategies to engage the foveal region when attempting to generate the specific internal representation. Collectively, these results advance our understanding of visual imagery generation by identifying the foveal region of V1 as a critical hub for top-down feedback modulation. Furthermore, this finding supports the argument that V1 does not merely passively receive external input but also actively shapes subjective experience, thereby serving as a necessary substrate for visual awareness.

TALK SESSION: TUESDAY, MAY 19, 2026, 10:45 AM – 12:15 PM, TALK ROOM 2

Decision Making

Moderator: Alan Stocker, University of Pennsylvania

TALK 1, 10:45 AM, 52.21

A RE-ANALYSIS OF 55 EXPERIMENTS REVEALS TWO FACTORS THAT DRIVE CONFIDENCE FOR ERROR TRIALS

Herrick Fung¹ (herrickfung@gatech.edu), Xinyue Li¹, Dobromir Rahnev¹; ¹School of Psychology, Georgia Institute of Technology

Confidence on correct trials reliably increases as tasks become easier. On error trials, however, empirical results diverge: some studies report higher confidence for easier stimuli, others report the opposite. This inconsistency suggests the presence of overlooked factors that shape confidence computations. Here, we reviewed and reanalyzed 55 experiments (N = 2,804), yielding 77 independent analyses of how error-trial confidence varies with difficulty. We found 41 decreases, 26 increases, and 10 flat patterns. Two factors fully explained this heterogeneity: (1) type of difficulty manipulation and (2) availability of post-decisional information. First, manipulations that increase stimulus reliability (e.g., using higher contrast or lower noise) tend to produce confidence increases for easier stimuli. In contrast, manipulations that vary the judged stimulus feature (e.g., manipulating tilt in an orientation discrimination task) tend to produce confidence decreases. Further analyses suggest that this dichotomy is mediated by whether the difficulty of the manipulation is perceptually visible: visible manipulations provide implicit cues to trial-by-trial difficulty, which are then used to adjust the final confidence rating. Second, the presence of post-decisional sensory evidence strongly drives confidence downward for easier stimuli on error trials, regardless of the difficulty manipulation. This is likely because post-decision evidence typically contradicts the initial choice on errors; since this conflict is stronger for easier stimuli, confidence decreases more steeply for those stimuli. Together, these two factors account for all 77 observed patterns. Our results reveal that a core signature of metacognition, namely how confidence changes with task difficulty on error trials, is not fixed but is determined by factors that are not accounted for by most models and often ignored in experimental design. Finally, we show that standard computational models can be extended to incorporate both factors and include simulations that reproduce the effects observed across all analyses.

This work was supported by the National Institute of Health (award: R01MH119189).

TALK 2, 11:00 AM, 52.22

SEQUENTIAL EVIDENCE ACCUMULATION ADAPTS TO CHANGING LEVELS OF EVIDENCE RELIABILITY

Mengting Fang¹ (mtfang@sas.upenn.edu), Alan Stocker; ¹University of Pennsylvania

Perceptual decision making often involves integrating evidence over time. Our prior work showed that this process is governed by a resource-rational principle where the brain actively balances task performance with cognitive effort. For a fixed level of evidence reliability, participants maintained a stable trade-off strategy. However, our previous study did not test whether participants adapt their strategy

when evidence reliability changes. Given the brain's ability to actively control the accumulation process, we considered two competing hypotheses: (1) the system increases encoding precision when evidence is reliable and reduces it when evidence is less reliable, or (2) the opposite, the system increases encoding precision to compensate for low-quality evidence and lowers it when reliability is high. To test these hypotheses, we conducted a visual estimation task where participants inferred the angular position of an unknown source from 8 sequentially presented, normally distributed stimulus samples. We manipulated sampling noise across three levels. The testing phase included 9 blocks across 3 visits in a balanced Latin square design, with trial-by-trial feedback and a performance-dependent bonus. Noise levels were not disclosed in each block. Results showed that participants' estimation performance decreased as sampling noise increased. Fits with our normative resource-rational model showed that encoding precision scaled with sampling noise, supporting our first hypothesis that the brain actively adjusts encoding precision to match external sensory reliability. Circular regression analysis further revealed that the variance of normalized temporal weights decreased as noise increased, reflecting shifts in temporal weighting strategies (e.g., from recency to uniform/primacy). Changes in trade-off parameters also showed greater tolerance for effort costs under higher noise, partly offsetting the reduced encoding precision. Together, these findings demonstrate that evidence accumulation is governed by a performance-effort trade-off strategy that adapts sensory encoding precision to the reliability of the evidence.

This work was supported by the NSF CRCNS grant IIS-1912232 to A.A.S

TALK 3, 11:15 AM, 52.23

HOW HISTORY MANIFESTS IN THE MACAQUE CORTICAL HIERARCHY

Jongmin Moon¹, Zoe M. Boundy-Singer², Julie A. Charlton³, Robbe L. T. Goris¹; ¹The University of Texas at Austin, ²Massachusetts Institute of Technology, ³Princeton University

Perceptual judgments of ambiguous stimuli exhibit intricate history-dependent biases. These effects arise from a sophisticated computational process that is both efficient and effective and is hypothesized to involve sensory as well as decision-making circuits. To test this hypothesis, we studied neural population activity in the visual and prefrontal cortices of macaque monkeys while they performed a perceptual decision-making task. The animals judged whether a visual stimulus was oriented clockwise or counterclockwise relative to vertical and communicated their decision with a saccade. As expected, perceptual choices of each monkey exhibited a complex history-dependency consisting of a mixture of stimulus repulsion and choice attraction. How does this behavioral pattern emerge from the underlying neural activity in cortical circuits? To address this question, we analyzed the one-dimensional linear component of neural

population activity in V1 (two monkeys) and FEF (two monkeys) that pertains to the formation of a perceptual decision. Neural activity in both areas was strongly influenced by the current sensory stimulus. In contrast, preceding sensory and choice events had a prominent effect on neural activity in FEF, but only a restricted effect in V1. Analysis of the dynamic evolution of FEF activity further revealed a temporal distinction between stimulus and choice history effects. The preceding choice already influenced FEF activity before the stimulus onset, while the influence of the preceding stimulus only manifested after the stimulus onset. This temporal distinction is consistent with theoretical proposals that attribute choice attraction and stimulus repulsion to two distinct computational components of decision-making: perceptual expectation and sensory adaptation. Together, our findings illuminate the physiological basis of the influence that recent experience exerts on current perceptual impressions. In doing so, our work illustrates how a targeted analysis of neural population activity in different brain regions can reveal the neurobiological basis of fundamental brain computations.

NIH R01 EY032999

TALK 4, 11:30 AM, 52.24

INTROSPECTING VISUAL BIASES

Noa Perlmutter¹, Chaz Firestone¹, Ian Phillips¹; ¹Johns Hopkins University

We are aware of the world and its properties; for example, we can see the number, size, and motion of objects around us. Are we also aware of the internal processes that generate these percepts? Whereas it is controversial whether higher-level cognitive processes are accessible to introspection (e.g., knowing why we make various choices), basic perceptual biases are widely assumed to be closed off from introspection. In contrast to this consensus, here we reveal successful introspection of three classic effects in vision and visual memory: numeric underestimation, size contrast, and representational momentum. Experiment 1 investigated numeric underestimation: Subjects briefly saw an array of 11-100 dots, and estimated their numerosity; on a subset of trials, subjects were then asked whether they thought they had underestimated or overestimated. Experiment 2 investigated the Ebbinghaus illusion: Subjects saw a target circle surrounded by flankers of variable size, and adjusted a second circle to match the target; subjects were then asked whether they thought they had underestimated or overestimated. Experiment 3 investigated representational momentum: Subjects saw a fish glide across the screen and disappear, and estimated its final seen location; subjects were then asked whether they thought their estimate was too far left or right. We replicated all three biases: Subjects underestimated numerosity, were influenced by flankers, and extrapolated motion. Strikingly, however, all three experiments also revealed awareness of the effects themselves. For example, in Experiment 1, subjects correctly answered that they underestimated numerosity. And in

Experiments 2 and 3, they showed trial-by-trial awareness of their biases, as computed by detection-theoretic statistics. These findings go beyond successful metacognition of performance (i.e., knowing whether one is performing well or poorly) to awareness of directional effects themselves, suggesting deeper and subtler access to internal mental processes than traditionally assumed.

TALK 5, 11:45 AM, 52.25

DISENTANGLING PERCEPTUAL AND ANALYTICAL PROCESSES IN SOLVING EUCLIDEAN AND NON-METRIC TRAVELING SALESPERSON PROBLEMS

Arman Baradaran¹, Jacob VanDrunen², Sébastien Hélie³, Zygmunt Pizlo¹; ¹University of California, Irvine, ²Westminster Seminary California, ³Purdue University

Human performance on classic traveling salesperson problems (TSPs) depends primarily on perceptual processes. Participants produce near-optimal tours by relying on hierarchical clustering followed by global-to-local tour refinement, though it is unclear to what extent analytical skills like forward planning are involved. We investigated this question using non-metric TSPs in which half of the cities were assigned one of two colors, and switching colors incurred a cost equal to twice the Euclidean distance between cities. Performance on 30 TSPs with 50 cities each was evaluated through two dependent variables: (a) error with respect to optimal tours and (b) number of color switches. Performance varied significantly more in non-metric TSPs than in Euclidean TSPs, with percent errors ranging from 3% to 32%. Several participants adopted a minimal-switch heuristic, producing tours that emphasized clustering of same-colored cities and making only 2 color switches. Others made deviations from color-based clustering that discounted the switching cost, with one participant even making 22 color switches on average. We then compared participants' performance to two versions of a multiresolution graph pyramid model. The first version, which operated on spatial clustering and used cheapest insertion to consider switching costs, underperformed compared to most humans and overestimated the number of switches. The second version had a front-end that used a 3D Euclidean approximation produced by applying multidimensional scaling (MDS) to the non-metric pairwise distances. This latter version mimicked the behavior of the minimal-switch participants, and after parameter adjustments in the MDS process, it also replicated the performance of more analytically-inclined participants. These findings suggest that combining a multiresolution pyramid representation with 3D MDS can capture human behavior in non-metric TSPs. More broadly, non-metric TSPs may serve as a useful proxy for studying how humans integrate perceptual and analytical processes in complex problem solving domains such as chess.

**HIERARCHICAL REPRESENTATIONS IN VISUAL
STRUCTURE LEARNING INTERACT WITH STIMULUS-
DEPENDENT METACOGNITIVE DYNAMICS**

Rochelle Kaper^{1,2,3}, Megan A.K. Peters^{1,2,3,4,5,6}; ¹Department of Cognitive Sciences, University of California Irvine, ²Center for Theoretical Behavioral Sciences, University of California Irvine, ³Center for the Neurobiology of Learning & Memory, University of California Irvine, ⁴Department of Logic & Philosophy of Science, University of California Irvine, ⁵Department of Experimental Psychology, University College London, ⁶Program in Brain, Mind, & Consciousness, Canadian Institute for Advanced Research

Our daily lives afford the ability to construct internal models of multiple spatiotemporal patterns without feedback, in order to predict incoming information and generalize to novel contexts; this process is called visual structure learning. Along with prediction errors, metacognitive error signals guide learning in the absence of external feedback. It is unknown how metacognitive capacity may change over time throughout learning, and how such dynamics would affect learning across structures and stimuli differing in abstractness under uncertainty. To address these gaps, we developed a novel probabilistic visual structure learning paradigm where we probed the dynamics and representation of learning and metacognition under uncertainty. Participants learned probabilistic co-occurrences of item pairs (AA', BB'...) and a probabilistic sequence of those co-occurrences (AA'→BB'...). After each learning phase where participants viewed probabilistic streams of item pairs, participants completed tests of explicit structure learning (forced-choice accuracy) and metacognitive capacity (accuracy-confidence correspondence). Participants completed 2 sessions in counterbalanced order: one with low-level perceptual fractals (n=56), and the other with familiar line drawings (n=57). Mixed effects linear models revealed significantly higher sequence accuracy compared to co-occurrences for fractals, and vice versa for line drawings. We also observed significant metacognitive capacity increases across learning for fractals but not line drawings, and metacognitive capacity was higher for sequence structure despite greater task accuracy in fractals. We modeled the representation of hierarchical structure information using analytical successor representations, revealing a potential tradeoff between metacognitive capacity, representation, and usage of sequential information in learning the co-occurrences that differ among stimuli. Our results reveal distinct learning rates, metacognitive capacity dynamics, and representations of hierarchical information across structures and stimuli varying in abstractness and familiarity, motivating novel task designs that can assess how different stimuli and structures alter how hierarchical information is represented and varying metacognitive capacity dynamics throughout learning.

Funding Source: Canadian Institute for Advanced Research / Research Corporation for Science Advancement

**TALK SESSION: TUESDAY, MAY 19, 2026,
2:45 – 4:30 PM, TALK ROOM 1**

Theory

Moderator: Michael Bonner, Johns Hopkins University

TALK 1, 2:45 PM, 54.11

**ATTRACTIVE AND REPULSIVE PERCEPTUAL BIASES
NATURALLY EMERGE IN GENERATIVE ADVERSARIAL
INFERENCE**

*Hyun-Jun Jeon¹ (hjjeon@unist.ac.kr), Hansol Choi¹, Oh-Sang Kwon¹;
¹UNIST*

Perceptual estimates of orientation exhibit noise-dependent biases: estimates are attracted toward cardinal orientations under high external noise but repelled away from them under high internal noise. Bayesian observer models with efficient coding constraints explain these effects with explicitly specified prior distributions and likelihoods, yet it remains unclear whether such properties can emerge from learning with natural cost functions. We propose a Generative Adversarial Inference (GAI) model in which estimates for noisy stimuli are constructed through generative reconstruction. GAI is realized through an adversarially trained autoencoding architecture with a discriminator that distinguishes real and generated stimulus-latent (x, z) pairs to enforce distribution-level consistency, while a reconstruction loss is applied simultaneously to guide the autoencoder to learn consistent stimulus-level mappings. The model was trained in a toy visual environment consisting of 32×32 Gabor patches whose orientations followed a bimodal distribution favoring the cardinal axes, reflecting natural orientation statistics. As a result of independent training runs with different random seeds (n = 30), the trained model consistently reproduced cardinal bias properties observed in human perception. Specifically, increasing external noise produced systematic attraction toward the cardinals, and internal noise in latent space induced reliable repulsion away from them. The learned latent representation exhibited efficient-coding signatures: Fisher information peaked at the cardinals and matched the implicit prior inferred from noisy reconstructions. Ablation experiments dissociated the contributions of the two training objectives. Reconstruction-only training produced attraction without consistent repulsion, whereas adversarial-only training yielded unreliable orientation estimates with large, unsystematic biases. Only the joint objective yielded a stable orientation manifold capable of converting internal variability into repulsive bias. These findings suggest that the attractive and repulsive cardinal biases traditionally attributed to efficient coding and Bayesian

inference can emerge naturally within a learned generative system, providing an account of their origin in human vision.

This research was supported by the National Research Foundation of Korea (NRF-2023R1A2C1007917 to O.-S.K.).

TALK 2, 3:00 PM, 54.12

HAND-ENGINEERED IMAGE-COMPUTABLE MODELS CAN STILL OUTPERFORM DNNs IN V1 SIMILARITY

Tanish Mendki¹, Sudhanshu Srivastava², Ansh Soni³; ¹University of California, Santa Barbara, ²University of California, San Diego, ³University of Pennsylvania

Task-based Deep Neural Network models (DNNs) are widely used as models of inferotemporal visual cortex (IT), with early work showing a large jump over previous hand-made models [Yamins and DiCarlo, 2014]. However, recent work has suggested that over time, not only has the performance-alignment relationship plateaued but reversed, with high performing models becoming worse models of IT [Linsley et al., 2023]. Here we attempt to see if this reversal extends to earlier cortical regions. We used a 515-image subset of the Natural Scenes Dataset (NSD) and extracted V1 voxel responses from eight subjects, constructing 515 x N matrices of neural data. For each model, we compared latent representations for the same 515 images across all layers and identified the best-fitting “V1 layer” based on subject-averaged alignment scores. We evaluated a broad set of models, including SOTA IT similarity models [Schrimpf et al., 2020], high-task-performant CNN’s and VIT’s, along with other models directly attempting to model V1 [Dapello et al., 2020]. Along with DNNs we also tested traditional, hand-crafted models such as HMAX [Riesenhuber and Poggio, 1999]. Three complementary similarity measures were used: (1) representational similarity analysis, (2) pairwise matching between model units and voxels, and (3) linear predictivity via ridge regression. Surprisingly, we find that the most modern models are equal or worse models of V1 compared to HMAX even as they are better models of IT. Furthermore, HMAX becomes the best model when utilizing representational similarity scores that care about representational geometry and indistinguishable from the best for strict pairwise matching, being only permutation invariant. These findings highlight that progress in task performance has not translated into better mechanistic models of V1, and that classical image-computable models should not be treated as obsolete benchmarks. Re-evaluating hand-engineered approaches, rather than defaulting to DNNs, might be crucial for improving biologically-grounded alignment.

TALK 3, 3:15 PM, 54.13

EFFICIENT CODING ENABLES DATA-EFFICIENT LEARNING

Ananya Passi¹, Brian S. Robinson², Michael F. Bonner¹; ¹Johns Hopkins University, ²Johns Hopkins University Applied Physics Laboratory

Deep learning is the leading approach for building computational models of visual cortex. However, conventional approaches to deep learning require immense training data and fail to explain how the human visual system learns from very little data. One key factor could be unsupervised learning, which is thought to play an important role in biological vision but is uncommon in leading deep neural networks. Here, we ask whether integrating unsupervised efficient coding principles into neural networks can make them trainable in low-data regimes. Previous work has proposed that efficient coding in higher-level vision is achieved when neurons are tuned to the maximally varying axes of natural stimuli. We created a network that learns such an efficient code in a deep hierarchy, with each layer learning from the activations of its inputs—a process that can be implemented through Hebbian learning without any task whatsoever (not even a self-supervised task). We next examined how this network performs when combined with subsequent supervised learning on image classification using limited training data. We additionally evaluated the ability of this network to predict human fMRI responses and behavioral similarity judgments using the NSD and THINGS datasets, which contain data on thousands of natural images. We found that under data-limited conditions, with just tens to thousands of training images in total, a hybrid procedure that combines unsupervised efficient coding with supervised learning strongly outperforms conventional end-to-end supervised learning. Specifically, this hybrid approach yields superior classification performance and superior alignment with visual cortex representations and behavioral similarity judgments. Remarkably, under the most stringent data limitations, conventional supervised learning fails, while our hybrid approach still exhibits robust performance. Together, these results show that biologically inspired efficient coding enables rapid learning from limited data, and they establish a new fully unsupervised learning model of the visual hierarchy.

TALK 4, 3:30 PM, 54.14

ROBUSTNESS ALIGNS THE LOCAL CODING AXES OF ARTIFICIAL AND BIOLOGICAL VISION

Nikolas McNeal^{1,2} (nikolas@gatech.edu), N. Apurva Ratan Murty^{2,3}; ¹Machine Learning, School of Mathematics, Georgia Tech, ²Center of Excellence in Computational Cognition, Georgia Tech, ³Cognition and Brain Science, School of Psychology, Georgia Tech

The frontier of vision neuroscience is shifting toward closed-loop experiments that rely on computational models to actively synthesize stimuli and manipulate neural states. However, current model evaluation metrics (like prediction accuracy and representational similarity) prioritize global alignment across images and neglect the fine-grained “local-coding” representational structure required for

closed-loop control. Here, we expose this evaluation gap and show that models with comparable global alignment can differ profoundly in their suitability for closed-loop model-guided experiments. To characterize local-coding axes, we trained encoding models (N=21 models) on human fMRI responses (3,514 voxels, from NSD) and used targeted adversarial probes to map the local representational geometry around stimuli. We found that even the state-of-the-art brain models (CLIP ResNet-50, ViTs, etc.) were highly fragile and exhibited idiosyncratic local-coding axes that varied between architectures (subspace similarity=0.05, energy overlap=0.01, ceiling=1). To identify models better aligned with the local-coding axes of the brain, we introduced the Control-Consensus Score (CCS). This score quantified the extent to which local-coding axes generalized across models, identifying coding directions that we hypothesized were more brain-like. We found that CCS varied non-linearly with adversarial sensitivity, with marked improvements emerging only once models exceeded a critical threshold of local robustness (logistic regression $R^2=0.98$). Finally, we validated our approach by generating stimulus manipulations for closed-loop experiments. We found that stimulus perturbations based on models with high CCS produced semantically meaningful and interpretable changes in the image, whereas those with low CCS did not. These high-CCS models also demonstrated superior efficacy in closed-loop tests on both macaque and human fMRI data. Together, these results show that global alignment does not imply local-axis alignment. Our work established local-coding axis alignment as a critical dimension of brain-model evaluation, particularly for next-generation studies that depend on precise and interpretable stimulus control.

This work is supported by the NIH Pathway to Independence Award (R01EY032603), NSF Nexus computation support (Allocation number: SOC250049), and a startup grant from Georgia Tech to NARM.

TALK 5, 3:45 PM, 54.15

TRAVELING WAVES ALONG THE CORTICAL DEPTH ARE MODULATED BY VISUAL STIMULATION AND CORRELATE WITH SYNCHRONOUS SPIKING ACTIVITY

Lihao Yan¹, Mitchell Morton¹, Anirvan S. Nandy¹, Monika P. Jadi¹; ¹Yale University

Neural oscillations, or rhythmic fluctuations of neural activity, have played an important role in neuroscience since their discovery. However, oscillations have primarily been studied by averaging across trials, a process that can obscure important spatiotemporal dynamics present in single trials. Recent studies have revealed that neural oscillations often manifest as traveling waves (TWs) that propagate across the cortex. TWs have been shown to play functional roles in sensory processing, working memory, and motor control. However, previous studies have primarily focused on TWs along the cortical surface, and the existence of this phenomenon along the depth of the

cortex remains understudied. Moreover, the neural bases of TWs remain unclear. In this study, we investigated the nature of TWs along the cortical depth by analyzing translaminar recordings in the visual cortex of awake monkeys. We observed local field potential (LFP) TWs along the cortical depth, with their speed and probability modulated by visual stimulation. Further, we found that LFP TW speed correlated positively with local spiking probability. To investigate the underlying neuronal mechanisms, we analyzed concurrent spiking activity using the multi-unit activity envelope (MUAe) signal. Surprisingly, we did not observe MUAe TWs; rather, we observed robust MUAe synchrony across cortical depth, which co-occurred with LFP TWs. Furthermore, the phase of MUAe during synchrony correlated with LFP TW speed. Finally, we developed a minimal model of LFP TWs and MUAe synchrony as a potential linking mechanism that encapsulates our findings. Our results provide new insights into the spatiotemporal dynamics of neural oscillations along the cortical depth and highlight the importance of developing mechanistic insights into the neuronal basis of LFP TWs.

TALK 6, 4:00 PM, 54.16

EARLY COARSE CODING AND LATER HIGH-DIMENSIONAL REFINEMENT IN IT CORTEX REVEAL DISTINCT CONTRIBUTIONS OF EXCITATORY AND INHIBITORY NEURONS

Sabine Muzellec¹, Sachi Sanghavi², Kohitij Kar¹; ¹York University, Department of Biology and Centre for Vision Research, Centre for Integrative and

Applied Neuroscience, Toronto, Canada, ²Department of Computer Sciences, University of Wisconsin-Madison, USA

How do excitatory and inhibitory neurons contribute to the emergence of object representations in primate inferior temporal (IT) cortex? We recorded large-scale population activity with Utah arrays from three macaques viewing 640 images (80 exemplars × 8 objects) and classified neurons as putative excitatory or inhibitory using spike waveform clustering, matching all analyses for cell count. We first asked when each population begins to carry object information. Inhibitory neurons became reliable significantly earlier than excitatory neurons (~12.3ms lead, $p<0.001$), indicating involvement in the earliest feedforward sweep. We next asked whether these early signals share a common representational format. They do not: inhibitory responses showed higher across-image variance and larger within- and between-category dispersion (all $p<0.05$), and population similarity (CKA) between the two groups dropped from ~0.7 to ~0.4 within the first 100 ms, revealing sharply diverging representational trajectories. Do these differences matter for categorization? Not at the level of overall accuracy: category decoding did not differ across early or mid-latency windows ($p>0.22$), suggesting that both populations can support coarse discrimination. However, the representational differences became crucial when examining which images were easy

or difficult. Excitatory neurons aligned substantially more strongly with monkeys' image-by-image accuracy (70–170ms: $\Delta=7\%$, $p<0.001$). To understand why inhibitory neurons lead in time yet fail to predict behavior, we examined representational geometry. By 100–200ms, inhibitory responses collapsed into a low-dimensional, highly correlated, small-radius manifold, providing coarse category separation but limited image-level structure. In contrast, excitatory neurons, the primary projection population, expanded into higher-dimensional, larger-radius manifolds that preserved rich object-specific structure that could support richer primate behaviors. Consistent with this interpretation, feedforward ANNs (e.g., ResNet-18), which implement high-dimensional feature transformations but lack realistic inhibitory circuitry, predicted excitatory responses significantly better than inhibitory responses ($p<0.001$). Together, these results reveal a coarse-to-fine division of labor in the IT cortex.

KK is supported by the Canada Research Chair Program (CRC-2021-00326), SFARI (967073), Brain-Canada Foundation (2023-0259), the Canada First Research Excellence Funds (VISTA Program), and NSERC (RGPIN-2024-06223). SM is funded by Connected Minds Postdoctoral Fellowship (supported by CFREF).

TALK 7, 4:15 PM, 54.17

PARTITIONING SIGNAL AND NOISE (PSN): A MODALITY-GENERAL DENOISING TECHNIQUE FOR NEURAL RESPONSES

Jacob S. Prince¹, Heiko Schütt², Dora Hermes³, Greta Tuckute¹, Ian Charest⁴, Peter Brotherwood⁴, David G. C. Hildebrand⁵, Michael J. Tan⁶, George A. Alvarez¹, Talia Konkle¹, Kendrick N. Kay⁷; ¹Harvard University, ²Université du Luxembourg, ³Mayo Clinic, ⁴Université de Montréal, ⁵University of Houston, ⁶Carnegie Mellon University, ⁷University of Minnesota

Large-scale neural datasets, such as the Natural Scenes Dataset, BOLD5000, and THINGS, have rapidly inspired a growing ecosystem of open data in visual neuroscience. To maximize the number of included stimuli, such datasets typically contain only a few repeated trials per stimulus, which leads to noisy measurements of neural activity even after trial averaging. This persistent noise challenges our ability to make accurate inferences about neural tuning and representational geometry, and limits the utility of the data for computational modeling. We introduce Partitioning Signal and Noise (PSN), a novel low-rank denoising method that is applicable to any dataset with repeated trials. Standard PCA-based denoising retains dimensions with high total variance, but this approach fails to distinguish signal and noise and is therefore susceptible to signal loss and biased neural activity estimates. In contrast, PSN uses a simple generative model to explicitly estimate the signal and noise distributions that underlie a set of data. This enables improved denoising via low-rank reconstruction that focuses on signal-rich

dimensions. We used ground-truth simulations to characterize the bias-variance properties of PSN. Compared to alternatives such as standard PCA, independent components analysis, and simple trial averaging, PSN achieves better recovery of ground-truth signal, while incurring minimal bias. We then applied PSN to diverse visual datasets taken from human fMRI, intracranial EEG, scalp EEG, macaque electrophysiology, and marmoset calcium imaging. In all cases, we observed substantial improvements in the stability of tuning profiles (e.g., for $n=2,091$ FFA-1 voxels in NSD: median tuning correlation between trial splits was $r=0.367$ for trial averaging, $r=0.584$ for PSN). As a consequence of improved signal fidelity, PSN consistently raised the predictive performance of encoding models. Broadly, by addressing the pervasive challenge of noise in large-scale data, PSN enables improved representational analyses, model evaluations, and theoretical insights from existing and future neural datasets.

This research was supported by NSF CAREER BCS-1942438 (TK), funding from the Kempner Institute (TK), and an NDSEG Research Fellowship (JSP).

TALK SESSION: TUESDAY, MAY 19, 2026, 2:45 – 4:30 PM, TALK ROOM 2

Eye Movements: Mechanisms, naturalistic viewing

Moderator: Stefan Pollmann, Otto-von-Guericke-University

TALK 1, 2:45 PM, 54.21

SACCADE PREPARATION CONCURRENTLY DEGRADES FOVEAL PERFORMANCE AND APPEARANCE

Nina M. Hanning¹ (hanning.nina@gmail.com), Hyun Seo Lee², Marisa Carrasco²; ¹Humboldt-Universität zu Berlin, Department Psychologie, ²New York University, Department of Psychology & Center for Neural Science

When we prepare saccadic eye movements, attention shifts toward the upcoming fixation location, enhancing perceptual performance and appearance. Recent work shows that the saccade target benefit is accompanied by a cost at the current fixation, where overall perceptual performance is reduced shortly before we look away. It is unknown whether the foveal costs of saccade preparation are limited to performance or also extend to visual appearance. Here, we investigated the presaccadic time course of foveal performance and appearance using a two-by-two AFC task. Participants prepared horizontal saccades while viewing random dot kinematograms (RDKs) at the presaccadic center of gaze. During saccade preparation, a motion signal of varying coherence was briefly presented, and participants reported both its perceived direction (performance) and

whether its coherence appeared weaker or stronger than a fixed reference signal presented earlier in the trial (appearance) with a single, combined response. Motion coherence was parametrically varied, and the timing of the motion pulse relative to the saccade cue was sampled across several presaccadic intervals. Psychometric functions were fitted separately for performance and appearance as a function of motion coherence and time. Consistent with previous results, performance at fixation deteriorated during saccade preparation: the coherence required to reach 80% correct accuracy increased progressively following the saccade cue, reflecting a growing performance cost. Appearance showed a closely matched pattern: the point of subjective equality (PSE) shifted systematically to higher coherence values over the course of saccade preparation, indicating a gradual reduction in perceived motion strength at fixation. Together, these findings show that saccade preparation produces parallel costs in both performance and appearance at the current center of gaze. Combined with established presaccadic benefits at the saccade target, the results support a redistribution account in which presaccadic attentional gains at the movement goal are mirrored by concurrent perceptual costs at fixation.

This research was supported by a Marie Skłodowska-Curie individual fellowship by the European Commission (898520) to NMH.

TALK 2, 3:00 PM, 54.22

HUMAN SCENE VIEWING FIXATION PATTERNS EMERGE FROM OPTIMIZING SCENE COMPREHENSION

Miguel P. Eckstein¹ (eckstein@psych.ucsb.edu), Shravan Murlidaran¹; ¹UC Santa Barbara

Introduction: Landmark findings of human eye movements during scene viewing include: first fixation at the center (center bias) and frequent fixations on faces, text, gazed objects, and objects critical to scene understanding. Here, we ask whether such signature human fixation patterns emerge from a strategy to optimize scene comprehension for an organism with foveated vision. We developed a foveated Visual-Language Model (VLM-FOV) whose fixation selection is optimized to comprehend scenes. Methods: Participants (N=20) viewed images of real-world scenes (M =277, UCSB) and were instructed to describe the scenes while eye movements were recorded. VLM-FOV incorporated visual loss as a function of eccentricity from the fixation point, mimicking human foveation, and trained it using reinforcement learning to maximize scene comprehension by optimally exploring through fixation selection with a 24K image training set. We compared the VLM-FOV model to saliency-based model fixations (GBVS, Harel et. al., 2006), and a neural network trained on human eye movement data to predict test image fixations (DeepGaze, Linardos, A. et al., 2021). Results: Scene description accuracy for the UCSB image set was superior for VLM-FOV with RL-optimized eye movements than fixation selection based on GBVS saliency or DeepGaze maps. Critically, the VLM-FOV model

showed emergent fixation patterns resembling humans and DeepGaze: center bias for first fixations and frequent fixations on people, text, and objects critical to scene comprehension. The overall prediction of human fixations in the scene description task (UCSB images) for VLM-FOV (AUC = 0.91) was higher than GBVS (AUC = 0.86 ; $p < 0.001$) and comparable, but somewhat inferior to DeepGaze (AUC = 0.93 ; $p < 0.001$), which is expected because Deep Gaze (unlike VLM-FOV) is trained on human fixations. Conclusions: Our findings suggest that the landmark human fixation selection behaviors during scene viewing reflect an eye movement strategy to optimize scene comprehension.

TALK 3, 3:15 PM, 54.23

NEURAL SIGNATURES OF NONLINEAR INFORMATION INTEGRATION ACROSS FIXATIONS - EVIDENCE FROM RECURRENT NEURAL NETWORK MODELLING OF A LARGE-SCALE MEG DATASET

Carmen Amme¹, Sushrut Thorat¹, Philip Sulewski^{1,2}, Malin Braatz¹, Alexander Kroner¹, Peter König^{1,3}, Tim Kietzmann¹; ¹Institute of Cognitive Science, Osnabrück University, ²Vision and Computational Cognition Group, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany, ³Department of Neurophysiology and Pathophysiology, Center of Experimental Medicine, University Medical Center Hamburg-Eppendorf, 20251 Hamburg, Germany

The human visual system relies on rapid eye movements to sample different regions of the visual field. This fragmented information is likely integrated to facilitate successful interaction with the environment. Yet, the search for neural signatures of information integration across eye movements has so far remained unsuccessful (Xiao et al., 2024). Here, we revisit this question by harnessing the power of a large-scale MEG-eyetracking dataset and advanced deep neural network (DNN) modelling. In our dataset, five participants freely viewed 4,080 natural scenes, yielding data for about 200,000 fixations. Using a linear encoding approach, which maps DNN embeddings of fixation patches onto fixation-locked neural activity, we compared models using only the current fixation's embedding with models incorporating embeddings from the preceding fixation. We tested whether local integration between successive fixations reflects (i) a linear remixing of features of past and present or (ii) a nonlinear integration process, as modelled by the Glimpse Prediction Network (GPN; Thorat et al., 2025), a self-supervised recurrent model trained either to (iia) predict visual embeddings for the subsequent fixation, or (iib) predict semantic large language model (LLM) embeddings of the scene captions. Linear remixing of successive fixation embeddings didn't enhance encoding performance after fixation onset. In contrast, GPNs' nonlinear integration did improve the encoding performance when contextualised with the previous fixation, where peak encoding performance at around 120 ms post fixation onset exhibited a 22.69% improvement (averaged across participants and 5 occipital

magnetometers). This improvement was absent for the GPN variant trained to predict LLM embeddings of the scene captions, indicating that the training objective was critical to reveal integration in the brain. Contrasting prior attempts, our findings provide evidence for neural signatures of cross-fixation integration, suggesting that integration operates locally from fixation to fixation instead of converging towards a holistic scene representation.

This work was supported by the European Research Council's (ERC) Starting grant #101039524 "TIME".

TALK 4, 3:30 PM, 54.24

AN ORBITOFRONTAL-ANTERIOR TEMPORAL NETWORK COMPLEMENTS THE ATTENTION NETWORK FOR VISUOSPATIAL ATTENTION IN NATURALISTIC VIEWING

Stefan Pollmann¹, Nico Marek¹, Radha Meghanathan¹; ¹Otto-von-Guericke-University, Magdeburg, Germany

Human neuroimaging experiments have consistently found dorsal and ventral attention networks subserving spatial shifts of covert attention or eye movements. However, these experiments used abstract, artificial stimuli for better experimental stimulus control and clearly defined instructions about where to attend or what to search for. This clearly is not an adequate model of eye movement control in natural environments full of semantic and social information. Here, we investigated fMRI activation and simultaneous eye movements during movie viewing, using the StudyForrest dataset. We correlated representational dissimilarity matrices of eye fixation patterns and BOLD patterns. Searching for intraindividual changes in activation patterns between movie scenes that mirrored changes in fixation patterns, we were able to identify the attention networks defined by previous experiments with artificial stimuli. In contrast, interindividual differences in fixation patterns, indicative of endogenous decisions about where to look, were mirrored by activation pattern differences in orbitofrontal and anterior temporal cortex. A follow-up analysis showed that this pattern was most pronounced for scenes containing faces. We were further able to replicate this pattern in two more datasets (Sherlock; 500 days of summer), estimating fixation patterns from the BOLD data. Thus, we could precisely replicate the role of the attention networks in gaze shifts in a more natural viewing environment. Crucially, we also showed that previous lab experiments, lacking the complexity of free viewing in a naturalistic environment, overlooked the importance of the orbitofrontal - anterior temporal network for individual differences in active social vision. Our findings show that the orbitofrontal - anterior temporal network needs to be considered in studies of active vision, particularly during looking at faces.

Deutsche Forschungsgemeinschaft, PO 548/18-1

TALK 5, 3:45 PM, 54.25

LOCAL MOTION DRIVES GAZE STABILIZATION IN MICE AND HUMANS

Felix Franke¹, Federica Rosselli; ¹University of Basel, Basel, Switzerland, ²Institute of Molecular and Clinical Ophthalmology Basel (IOB), Basel, Switzerland

Reflexive gaze stabilization, mediated by the optokinetic reflex (OKR), is critical for maintaining visual acuity during self-motion. Traditionally, this behaviour is assumed to require the computation of global image motion (pattern motion) to accurately compensate for the visual environment's trajectory. However, the primary visual inputs to the nucleus of the optic tract (NOT)—direction-selective retinal ganglion cells (DS-RGCs)—encode only local edge motion and suffer from the aperture problem. Here, we combine in vivo electrophysiology in mouse NOT, ex vivo retinal recordings, computational modelling, and cross-species psychophysics in humans and mice to determine whether the gaze stabilization system solves this ambiguity or relies on a simpler heuristic. We show that mouse NOT neurons have horizontally elongated receptive fields that cluster along the visual horizon, acting as linear integrators of local retinal inputs. When challenged with ambiguous motion stimuli, where the global motion direction diverges from the local edge vector, NOT neurons fail to encode the true global direction. Instead, their tuning shifts to align with the dominant local motion component, a result accurately predicted by a simple linear pooling model of DS-RGC inputs. Behaviourally, the mouse OKR mirrors this neural error, tracking the vector average of local edges rather than the object's true trajectory. Remarkably, human OKR exhibits a similar systematic bias, rotating towards the local edge orientation away from the global pattern motion. These findings demonstrate that mammalian gaze stabilization relies on a low-latency, subcortical heuristic (linear pooling of local edges) that is evolutionarily conserved from mice to humans.

This work was financially supported by the Swiss National Science Foundation (SNSF) through multiple grants—Eccellenza grant PCEFP3_187001 (to F.F.); Spark grants CRSK-3_220987 (to F.F.) and CRSK-3_221257 (to F.B.R.); the Projects in Life Sciences Grant 310030_220209 (to F.F.).

TALK 6, 4:00 PM, 54.26

V4 NEURONAL ACTIVITY IN RESPONSE TO MICROSACCADES

Shawn M. Willett^{1,2}, J. Patrick Mayo^{1,2}; ¹University of Pittsburgh, ²Center for the Neural Basis of Cognition

Microsaccades are thought to improve the detection of visual stimuli. One-way microsaccades might mediate stimulus visibility by releasing photoreceptors from adaptation by shifting the retinal image.

This idea is supported by the modulation of ongoing activity in many visual areas by microsaccades. Yet the specific modulation seems to depend on the particulars of the study (e.g. stimuli). Indeed, recent work indicates that microsaccade direction is a key parameter in determining the modulation across the visual hierarchy. However, how microsaccade-induced retinal changes affect neuronal activity remains unclear. In this study we quantify the modulation of V4 activity in response to microsaccades while monkeys attended to visual targets. We recorded eye position while monkeys fixated during an orientation change detection task. Monkeys maintained fixation until the orientation of one of two simultaneously presented Gabor stimuli changed. Gabor stimuli were odd-symmetric, full-contrast and counterphased at 10 Hz and presented in the lower left and right visual fields (4.4°-27° eccentricity). Eye position was measured using a scleral search coil sampled at 200 Hz. We detected microsaccades during the fixation epoch (500-5000 ms) using a 6°/s velocity threshold. We recorded the activity of over 3500 V4 neurons from two bilaterally implanted Utah arrays across 54 sessions. We found that microsaccades exerted significant and complex modulations of ongoing V4 activity. On average, roughly 80 milliseconds after microsaccade onset there was a significant suppression of V4 activity. This suppression lasted for approximately 100 milliseconds. Following this suppression was a large rebound enhancement in firing rates from 180-280 milliseconds after microsaccade onset. Importantly, we do not find an effect of microsaccade direction on the average V4 response. This work shows that intermediate visual areas like V4 are still affected by precise movements of the retina and may contribute to the improved visibility of stimuli after microsaccade.

TALK 7, 4:15 PM, 54.27

A UNIFIED CONTINUOUS PSYCHOPHYSICS APPROACH FOR CONCURRENT ASSESSMENT OF OCULOMOTOR BEHAVIOR AND PERCEPTUAL THRESHOLDS

Andrea Canessa¹, Federica Ugoni¹, Francesca Peveri¹, Alice Sansalone¹, Agostino Gibaldi², Silvio P. Sabatini¹; ¹University of Genova, ²University of Modena-Reggio

Abstract: Traditional clinical assessment of oculomotor and perceptual functions often rely on batteries of disjointed, repetitive tasks (e.g., fixed-amplitude saccades, simple motion patterns) that lack ecological validity and are prone to cognitive prediction strategies. We present a novel, unified perceptual and computational framework designed to simultaneously quantify saccadic, pursuit, and vergence performance alongside perceptual thresholds for contrast and stereo within a single, continuous tracking task. **Methods:** The paradigm utilizes a Gabor patch stimulus moving within a viewing frustum defined by Hering's binocular coordinates (version, vergence). Unlike standard deterministic trajectories, target motion is driven by a stochastic engine. This "blind" generation concurrently probes physiologically consistent smooth pursuit, fixation stability and reactive saccades. The

stochastic nature of the task suppresses anticipatory motor planning strategies, yielding robust assessment of vergence stability and fusional ranges, smooth pursuit gain and latency, and saccadic gain, latency and main sequence. A multi-threaded adaptive engine (QUEST+) continuously modulates stimulus parameters (contrast, spatial frequency, binocular disparity) based on real-time gaze and vergence error, effectively closing the loop between behavioral performance and task difficulty. **Results:** The framework, tested on healthy participants, is able to provide a reliable and accurate assessment of oculomotor and perceptual performances by analyzing the continuous stream of version and vergence. Furthermore, the adaptive engine allows for the rapid estimation of contrast sensitivity functions and stereoacuity thresholds alongside the task flow. **Conclusion:** This framework aims at a paradigm shift from discrete, static testing to dynamic, continuous assessment. It represents a first step toward the gamification of sequential psychophysical evaluations, offering an alternative to conventional paradigms that are often unsuitable for fragile or less collaborative populations. By doing so, we provide an ecologically valid and adaptable tool that can be readily tailored to individual needs and extended to different neurological and neurodegenerative conditions.

Research reported in this publication was partially supported by the National Eye Institute of the National Institutes of Health under Award Number R01EY032162. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

TALK SESSION: TUESDAY, MAY 19, 2026, 5:15 – 7:15 PM, TALK ROOM 1

Color, Light and Materials

Moderator: Rhea Eskew, Northeastern University

TALK 1, 5:15 PM, 55.11

BLURRED FOVEAL VISION MAKES THE PERIPHERY APPEAR BLURRIER: RAPID RECALIBRATION OF PERIPHERAL APPEARANCE UNDER DEGRADED FOVEAL INPUT

Seonggyu Choe^{1,2,3} (seonggyu.choe000@gmail.com), Alex R. Bowers^{2,3}, MiYoung Kwon¹; ¹Department of Psychology, Northeastern University, Boston, MA, USA, ²Schepens Eye Research Institute of Massachusetts Eye and Ear, Boston, MA, USA, ³Department of Ophthalmology, Harvard Medical School, Boston, MA, USA

The fovea and periphery differ markedly in sampling resolution, resulting in characteristic declines in acuity and contrast sensitivity with increasing eccentricity. Yet perceptually, the visual field appears

seamlessly clear and coherent. This uniformity is thought to arise from calibration processes that align peripheral appearance with high-resolution foveal vision. When foveal input is degraded, as in conditions causing central vision loss, this calibration hierarchy is disrupted. This raises the question of whether peripheral appearance is recalibrated to match a compromised fovea. Here we measured perceived peripheral blur using the point of subjective equality (PSE), with the fovea (normal or degraded) serving as the reference. Normally-sighted observers viewed naturalistic dead-leaves textures with varied peripheral blur and judged whether the periphery appeared blurrier than the fovea. Foveal degradation was simulated by applying a 1.33-cpd low-pass filter to the central 5° region in a gaze-contingent display, yielding blur level comparable to ~20/450 acuity. PSEs were obtained before-and-after 20-minute adaptation during which participants performed demanding recognition tasks under foveal blur. A separate experiment confirmed that perceived clarity of the normal fovea was comparable to 20/20 acuity. We observed pronounced shifts in both the PSE and the psychometric slope as foveal clarity decreased. Under normal-fovea viewing, a peripheral blur level of 5.23 cpd±0.34 (~20/120) was perceived as matching the intact fovea. This six-fold difference in physical resolution indicates that normal foveal-vision strongly anchors perceived peripheral clarity. In contrast, under degraded-fovea, a peripheral blur level of ~1.39 cpd±0.16 (~20/430) was perceived as equal to the foveal blur. PSEs in degraded-fovea condition were unaffected by adaptation, suggesting this calibration occurs immediately. Our findings suggest that the visual system rapidly down-regulates perceived peripheral clarity when foveal input is degraded, thereby maintaining perceptual balance between the fovea and periphery. They also offer insight into perceptual changes associated central vision loss.

TALK 2, 5:30 PM, 55.12

NATURAL RETINAL STATISTICS AND THE DESIGN OF AN OPTIMAL TRICHROMATIC ARRAY

Deepthi Bannai^{1,2}, Alexander Belsten^{1,2}, Bruno Olshausen^{1,2}, Jacob Yates^{1,2}; ¹UC Berkeley, ²Redwood Center for Theoretical Neuroscience

Human color perception arises from the comparison of signals from long-(L), medium-(M), and short-(S) wavelength-sensitive cones, which form a mosaic that tiles the retina and encodes visual input. Interestingly, the human cone mosaic contains few, spatially distributed S-cones (Curcio et al. 1991). Previous work attributed S-cone sparsity to longitudinal chromatic aberration (LCA), but neglected to incorporate the full optical constraints of the eye, utilize a trichromatic array, or learn spatial arrangements (Garrigan et al. 2010). We address these limitations by proposing an efficient-coding model that optimizes a population of linear-nonlinear neurons to maximize information transfer between its inputs and photoreceptor outputs ((Karklin & Simoncelli 2011). Each cone in the mosaic learns a spatial location and spectral type. Model inputs are derived from

hyperspectral natural images (Arad et al. 2022) processed through a full model of the human optics (Cottaris et al. 2019). Power analysis of hyperspectral, optically-processed, and cone-response images demonstrated that optical and spectral filtering alters image statistics in a wavelength-dependent manner. Hyperspectral images had a $1/f^2$ distribution, which only medium wavelengths retained up to a cutoff after optical processing. Due to high LCA, shorter wavelengths had the lowest amount of power, which translated to S-cones having significantly lower power than L- and M-cones. Optimal mosaics learned by the efficient-coding model had few, distributed S-cones, similar to observed human mosaics. Finally, we observed a negative relationship between degree of LCA at shorter wavelengths and proportion of S-cones in learned mosaics. Taken together, this work provides a framework for describing the formation of an optimal trichromatic cone mosaic. By incorporating natural images processed by realistic model of the human, the model predicts optimal mosaics which match those of the human retina. Future work will apply this framework to the optical systems of other species.

National Science Foundation, grant # 2313149

TALK 3, 5:45 PM, 55.13

ANALYSIS OF NATURAL IMAGES REVEALS RESPONSE POLARITY DIFFERENCES IN THE CONES AND IN THE CHROMATIC CHANNELS

Yangyi Shi¹, Rhea T. Eskew, Jr.¹; ¹Northeastern University

Human visual perception exhibits distinct asymmetries in the processing of increments and decrements. Psychophysical studies show that perceptual scales of achromatic and S-cone increments follow a saturating curve, while their decrements both follow a reversed, inverted S-shaped curve; discrimination thresholds on increment pedestals are higher than on decrement pedestals (Shi & Eskew, 2024, 2025, Gabree, Shepard & Eskew, 2018), and field spectral sensitivities are different for S increment and decrement tests (McLellan & Eskew, 2000). The present study investigates whether these psychophysical polarity differences could be related to natural image statistics. Hyperspectral images of natural and urban scenes (Cauwerts & Jost, 2019; Nascimento et al., 2016) were analyzed to characterize the distribution of cone response contrasts. Spectral data were converted to L, M, and S cone excitations, and then filtered with Difference of Gaussians (DoG) using different spatial scales for each type of cone to simulate cone response contrast. Post-receptoral channel responses were then calculated using weighted sums of cone contrasts. The analysis revealed a consistent statistical dominance of decrements over increments across L, M, and S cones, at multiple spatial scales. In post-receptoral channels, the Red-Green channel showed a prevalence of "greenish" over "reddish" responses, while the Yellow-Blue channel showed a prevalence of "yellowish" over "bluish" responses, with few exceptions. Both Luminance (L+M) and

Achromatic (L+M+S) channels exhibited a strong bias toward decremental responses. These results align qualitatively with previous psychophysical asymmetries. The observed decrement dominance may be attributed to the physics of material reflectance: objects typically exhibit low reflectances due to light-absorbing pigments (Richards, 1982). Furthermore, consistent with the efficient coding hypothesis, the high sensitivity observed at low contrasts for both polarities in psychophysics can be explained by the prevalence of low-contrast content in the natural environment.

NSF BCS-2239356

TALK 4, 6:00 PM, 55.14

DIVERSE AND FUNCTIONALLY DISTINCT CONTRAST RESPONSE FUNCTIONS OF MACAQUE V1 NEURONS

Cai-Xia Chen¹, Shu-Chen Guan¹, Nian-Sheng Ju¹, Sheng-Hui Zhang¹, Xin Wang¹, Shi-Ming Tang¹, Cong Yu²; ¹Peking University, ²Zhejiang University

Contrast response functions (CRFs) of V1 neurons are typically described as fast rising at low contrasts and saturating at high contrasts. These CRFs are well captured by the Naka–Rushton model and are effective in explaining low-contrast discrimination, but how well they support high-contrast discrimination remains unclear. Additionally, intermediate- and high-threshold neurons have been reported, yet their prevalence and functional roles are not well established. We used two-photon calcium imaging to measure CRFs of V1 neurons in five fixating macaques. Among ~10,000 identified neurons, 54.2% were low-threshold ($C_{10} < 0.25$) with steep, shallower, or linear slopes, 21.7% were intermediate-threshold ($C_{10} = 0.25-0.60$) with steep or shallower slopes, and 9.7% were high-threshold ($C_{10} > 0.60$) with steep slopes. Because the Naka–Rushton model's C_{50} is less meaningful for intermediate- and high-threshold classes, we adopted a revised version (Heeger, 1992) replacing C_{50} with a threshold term, such that CRF arises when stimulus contrast exceeds threshold. This revised model fits the data well ($R^2 = 0.74$). To characterize functional roles of each CRF class, we trained binary logistic classifiers to discriminate adjacent contrast levels with high accuracy. Furthermore, ablation experiments revealed that removing intermediate- and high-threshold neurons selectively impaired contrast discrimination at their corresponding contrast ranges. However, removing low-threshold neurons impaired discrimination at all contrasts, though less so at intermediate and high contrasts, likely because low-threshold neurons' sheer number and many show shallower or even linear slopes to convey higher contrast information. CRF slope was also critical: removing steep-slope neurons markedly reduced discrimination across the full contrast range, while removing shallower-slope neurons had minimal impact. We conclude that (1) low-threshold neurons may not fully decode intermediate and high contrasts without contributions from higher-threshold neurons; (2) The revised Naka–Rushton model is more interpretable describing CRFs

of all neurons; (3) Neurons with steep CRF slopes play a dominant role in contrast decoding.

This study was supported by a Ministry of Science and Technology, China, STI2030-Major Projects grant (2022ZD0204600).

TALK 5, 6:15 PM, 55.15

THE GEOMETRY OF COLOR SPACE: SUPRATHRESHOLD DIFFERENCES FROM DISCRIMINATION THRESHOLDS

Fangfang Hong¹ (fh862@sas.upenn.edu), Jason Chow², Phillip Guan², David H. Brainard¹, Alex H. Williams³; ¹Department of Psychology, University of Pennsylvania, ²Reality Lab Research, Meta, ³Center of Neural Science, New York University

A foundational question in psychophysics is whether a geometric perceptual metric can predict both discrimination thresholds and suprathreshold difference perception. Investigation of this question is greatly facilitated by comprehensive measurement of both, but this has been intractable because of a large number of psychophysical trials required. Here, we first fully characterized one participant's color discrimination thresholds throughout a 3D color gamut (~30,000 trials) using a 3AFC oddity task and fitting an innovative Wishart Process Psychophysical Model. This model derives color discrimination thresholds in all directions around any reference stimulus, and characterizes the local geometry of color space. We then evaluated how well this local geometry can be used to predict suprathreshold color differences. Suprathreshold differences were measured in the same participant using a 2AFC task with stimuli closely matched to the threshold task (~7,200 trials). On each trial, a reference stimulus was cued, and the participant judged which of two comparisons appeared more different. We interleaved 12 conditions, each defined by a fixed reference (R_{fixed}) and a comparison (C_{fixed}). For each condition, the second comparison (C_{variable}) was adaptively sampled to measure the contour at the same perceptual distance from R_{fixed} as C_{fixed} . The difference between R_{fixed} and C_{fixed} was varied across conditions from ~2 to ~6 just-noticeable differences (jnds). The measured contours were compared to equal-geodesic-distance predictions derived from the threshold data. Specifically, for each R_{fixed} , we traced Riemannian geodesic paths in many directions, identified where the geodesic distance from R_{fixed} matched that of C_{fixed} , and connected those points to form the equal-geodesic-distance prediction. This prediction holds for short distances between R_{fixed} and C_{fixed} ($< \sim 2$ jnds), but breaks down as this distance increases. The deviations appear systematic, with the measurements falling inside the predicted equal-geodesic-distance contours. These measurements provide important constraints for developing global geometric metrics of color comparison.

Meta

TASK-DEPENDENT COMPENSATION IN ANOMALOUS TRICHROMACY

Fatemeh Charkhtab Basim¹, Erin Goddard², Kenneth Knoblauch³, Kimberly Jameson⁴, Michael Webster¹; ¹Department of Psychology, University of Nevada, Reno, NV, US, ²School of Psychology, University of New South Wales, Sydney, Australia, ³Univ Lyon, Université Claude Bernard Lyon 1, Inserm U1208, Stem Cell and Brain Research Institute, Bron, France, ⁴Institute for Mathematical Behavioral Sciences, University of California, Irvine, CA, USA

A number of studies have shown that anomalous trichromats may have stronger color percepts than predicted by their threshold sensitivity losses. However, the form and nature of this compensation remains poorly understood. Prior work has typically examined isolated tasks, leaving open how compensation varies across different levels of perceptual and cognitive processing. In the present study, we used common stimuli and compensation metrics to quantify anomalous observers' performance across a range of different tasks. 17 anomalous trichromats (11 deuteranomalous, 6 protanomalous) and 15 normal trichromats completed a battery assessment for stimuli that varied in nominal LvsM chromatic contrast. These included color classification (achromatic boundary placement), focal-hue selection, reaction-times, contrast adaptation, and magnitude estimation (MLDS). For each, we computed the contrast-scaling factor required to align an observer's responses with the average values for the color normal observers, normalized by scaling relative to the threshold elevation (resulting in a compensation index ranging from 0=no compensation to 1=full compensation). The average index varied strongly with task, Appearance-based measures (e.g. focal hues) showed the strongest effects, along with contrast scaling (MLDS). Alternatively, contrast adaptation exhibited significant but smaller improvements, with weakest compensation found for reaction times. The degree of compensation also showed large observer variability, which was unrelated to the magnitude of the individual's threshold loss. Across tasks, deuteranomalous observers generally showed stronger compensation than protanomalous observers, particularly for naming and reaction times, whereas group differences were weaker for adaptation, hue selection, and MLDS. Overall, the results show that compensation in anomalous trichromacy may involve multiple distinct processes that are differentially engaged by perceptual tasks.

Supported by EY-010834

TALK 7, 6:45 PM, 55.17

COLOR COMMUNICATION ACROSS THE LIFE SPAN

Angela Brown¹, Delwin Lindsey²; ¹Ohio State University College of Optometry, Graduate Program in Vision Science, ²Ohio State

Color communication is more than a naming game where the informant names each color: the color terms must also be understood by the listener. Here, we studied how color naming and color term comprehension (participants' productive and receptive vocabularies) affect communication about color. *****Human participants, 3–72 years old, played a color communication game. As "senders", they named 15 or 30 color samples presented individually. Then, as "receivers", they viewed or heard "message" terms from their own sender data (self-communication) or from another participant (interpersonal communication), in random order. The receiver's task was to select, from the full array of colors, the sample that corresponded to each message. The number of color terms in participants' data sets as senders, their fraction correct performance for self-communication, and their fraction correct for interpersonal communication, all matured continuously until at least age 21, and were highly correlated with each other. *****Simulations correctly predicted that participants' self-communication choice performance should be directly correlated with the number of color terms they had provided as senders (their productive color vocabulary size), both of which improved with age. By contrast, simulations predicted that participants' interpersonal choice performance should be (1) much worse than self-communication, (2) nearly unrelated to their productive vocabulary size and (3) independent of age. Contrary to these predictions, (1) choice performance for interpersonal communication was only 6% below self-communication, (2) it was strongly associated with the receiver's productive vocabulary size ($r^2=0.76$), and (3) it improved throughout the age range ($r^2=0.70$). Further analyses attributed these effects to an age-related increase in the size of participants' receptive color vocabularies, because participants often understood terms they had not themselves produced as senders. *****Therefore, color naming (productive color vocabulary) and color term comprehension (receptive color vocabulary) are distinct abilities that mature independently across the life span.

NIH T35 EY007151, BCS-1152841, Center Of Science and Industry, Columbus, OH

TALK 8, 7:00 PM, 55.18

NEURAL CORRELATES OF CROSS-MODAL EXPECTATIONS IN MATERIAL PERCEPTION

Amna Malik¹, Katja Doerschner¹; ¹Justus Liebig University

Humans routinely integrate visual and auditory information to infer what objects are made of and to estimate their material properties. Although multisensory facilitation is well documented, most neuroimaging studies have investigated responses to simultaneously presented congruent versus incongruent audiovisual stimuli, where bottom-up integration and top-down processes are tightly intertwined

and difficult to disentangle. Consequently, the contribution of top-down cross-modal expectations remains less explored, particularly in dynamic, complex scenes. Here, we used fMRI to investigate how such cross-modal expectations about material properties are represented in the brain. Participants viewed computer-generated animations depicting familiar objects (a wooden chair, a porcelain teacup, and a jelly) being dropped to the ground, providing visual information that elicited expectations about the object's material and its characteristic impact sound. Just before impact, the object was masked, leaving only the auditory information. On each trial, this auditory information either met the expectations elicited by the preceding visual information (congruent) or violated them (incongruent), and participants judged whether the object was hard or soft. Multivariate pattern analyses showed that audiovisual congruency was most robustly represented in the lateral occipital cortex (LOC). Crucially, LOC activity was modulated even though expectations were violated in the auditory domain and the corresponding impact events were never presented visually. These findings indicate that object-selective visual cortex may maintain anticipatory, visually derived representations of how an object is expected to behave across modalities, and that these representations are updated when incoming auditory information violates those expectations. More broadly, they suggest that LOC may support the flexible use of context-dependent predictions to guide behavior in complex and dynamically changing environments.

This work is supported by the Deutsche Forschungsgemeinschaft (German Research Foundation, DFG) project number 222641018 – SFB/TRR 135 and DFG under Germany's Excellence Strategy (EXC 3066/1 "The Adaptive Mind", Project No. 533717223).

TALK SESSION: TUESDAY, MAY 19, 2026, 5:15 – 7:15 PM, TALK ROOM 2

Motion: Neural mechanisms, models and illusions

Moderator: Alex Huk, UCLA

TALK 1, 5:15 PM, 55.21

FIRST-ORDER, SECOND-ORDER AND NONRIGID MOTION TUNING IN MARMOSET MT/MTC/FST

Krischan Koerfer¹, Alexia Nelms², Markus Lappe¹, Jude Mitchell^{2,3};
¹Institute for Psychology and Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Münster, ²Brain and Cognitive Sciences, University of Rochester, ³Center of Visual Science, University of Rochester

Neurons in area MT show strong direction tuning for first-order (luminance-defined) motion. By contrast, reports of second-order motion tuning in MT are mixed, and the cortical basis of nonrigid motion processing remains poorly understood, while behavioural evidence from pursuit and speed perception suggesting partly distinct pathways (Koerfer & Lappe, 2022; Koerfer et al., 2024). In marmoset monkeys motion-selective areas MT, MTC and FST are fully accessible on the cortical surface. We recorded extracellular single-unit activity using linear electrode arrays across these areas while presenting first-order, second-order and nonrigid motion stimuli during free-viewing. Off-line eye position was corrected to establish retinotopy and map receptive fields. Receptive fields were defined based on the response to brief (50 ms) first- and second-order motion probes (16 directions of motion, 10 deg/s), yielding 147 units with significant receptive fields. For tuning measurements, we presented first-order, second-order and nonrigid motion for 150 ms in 16 directions at 10 deg/s. Firing rates were quantified with a direction selectivity index (DSI), and tuning significance was assessed using a bootstrap and shuffle-based procedure. Across the population, 47 units showed significant direction tuning at this fixed speed for first-order motion. Most recorded units were also significantly driven by second-order and nonrigid motion albeit without direction tuning. DSIs for these conditions did not rise reliably above shuffled controls, even in units that were clearly tuned for first-order motion. These results suggest that, under our stimulus conditions, direction-selective tuning in the marmoset MT/MTC/FST complex is dominated by first-order motion, whereas second-order and nonrigid motion primarily modulate firing rate without strong or consistent direction selectivity. This dissociation constrains models of how different motion pathways contribute to the perception of complex and nonrigid motion and suggests that direction-selective tuning for second-order and nonrigid motion may emerge in areas downstream or parallel to MT/MTC/FST.

Funding: R01-EY030998, EU MSCA SE 101068206

TALK 2, 5:30 PM, 55.22

REPRESENTATIONAL GEOMETRY OF OPTICAL FLOW PATTERNS DETERMINES THE RESPONSE PROPERTIES OF MST NEURONS

Nathaniel Powell¹ (npowell@utexas.edu), Mary Hayhoe¹, Gregory DeAngelis², Xuexin Wei¹; ¹University of Texas at Austin, ²University of Rochester

Neural tuning properties are shaped by stimulus statistics. Efficient coding predicts that distributions of neurons and Fisher information should be allocated according to stimulus statistics. While efficient coding can account for encoding of certain stimulus variables, its prediction seems to be inconsistent with encoding of heading direction in macaque MST (i.e., more neurons are selective to lateral headings, yet the sensitivity to them is lower). We developed computational

models of how the visual system hierarchically processes heading direction. A given heading direction causes an optical flow pattern, which is encoded by motion selective neurons in the model MT layer. Model MST neurons then linearly integrate MT responses. We constructed several variations of models, and investigated the predicted neuronal selectivity and representational geometry. The models used optical flow patterns as inputs that were generated either by real human movement or by uniformly sampled heading directions from analytically derived flow patterns, inconsistent with natural statistics. To test the model predictions, we re-analyzed the response of recorded macaque MST neurons from published datasets. When changing input statistics, model MST responses adapt, as reflected in the change in Fisher information and distribution of neural preferences, consistent with efficient coding. However, when heading direction is uniformly distributed, model MST neurons still exhibit large heterogeneity in tuning properties similar to neural data. This observation cannot be explained by efficient coding. Crucially, we found that the tuning heterogeneity is due to the asymmetry of the stimulus manifold defined by the optical flow. Model MST neurons inherit this asymmetry, causing the heterogeneity in their tuning properties. Importantly, macaque MST neurons exhibit similar asymmetrical structures in their representational geometry. We found that the geometry of the input manifold influenced the tuning properties of downstream areas and could account for the puzzling tuning properties observed in MST.

This work is supported by NIH EY05729 (to Hayhoe), NSF 2318065 (to Hayhoe and Wei) and a Sloan Research fellowship (to Wei).

TALK 3, 5:45 PM, 55.23

CONSISTENCY BETWEEN VISUAL MOTION AND LOCOMOTION MODULATES RESPONSES IN PRIMATE AREA MT

Penny-Shuyi Chen^{1,2}, Declan P Rowley^{1,3}, Alexander C Huk^{1,4}; ¹Fuster Laboratory for Cognitive Neuroscience UCLA, ²Neuroscience Interdisciplinary Program UCLA, ³Department of Ophthalmology UCLA, ⁴Department of Psychiatry & Biobehavioral sciences UCLA

Primate motion area MT has been extensively studied, often in animals whose movements are highly constrained, with them passively viewing or performing artificial tasks when viewing synthetic stimuli. It is therefore unknown whether MT exhibits computations that involve interactions between active behaviors and resulting retinal motions. Here, we employ a virtual reality setup for neural recordings in marmosets to test whether responses in MT are affected by whether retinal motions resulted from the animal's own movements. Two head-fixed marmosets ran on a treadmill. They viewed simulated realistic optic flow patterns (dot fields), with a densely-structured flat ground plane and a less-regular upper zone. In the "Closed-Loop" condition, the stimulus was updated in real time to be consistent with the

subject's concurrent treadmill locomotion. In a "Replay" condition, the stimulus from the "Closed-Loop" was replayed, but the visual display was no longer yoked to the treadmill. We conducted multi-area Neuropixels recordings (MT, MST, V1 and V2) and high-precision digital DPI eye-tracking, and also performed extensive mapping of the receptive fields of individual neurons. As expected, MT activity was strongly driven by retinal motion (number of MT units = 3243). But, surprisingly, the population average firing rate was approximately 20% higher during Replay than Closed Loop. This Closed-Loop versus Replay modulation was evident in highly-responsive and well-tuned single neurons, and was not explained by initial analyses of running modulation, saccade modulation, or arousal states. Furthermore, MT showed higher activity and larger response variance during open-loop Replay compared to (visual-motion-matched) Closed-Loop when there was a mismatch between visual motion and ongoing locomotion. V1/V2 responses (n=1587) did not show this effect. These results suggest that MT responses to self-generated motion signals are suppressed, consistent with a gain-based mechanism that differentiates self-generated retinal motions from those derived from moving objects and scene elements.

Fuster endowment, UCLA

TALK 4, 6:00 PM, 55.24

BETTER, BUT NOT SUFFICIENT: COMPARING VIDEO ARTIFICIAL NEURAL NETWORKS, AND MACAQUE IT DYNAMICS

Matteo Dunnhofer^{1,2}, Christian Micheloni², Kohitij Kar¹; ¹York University, ²University of Udine

Feedforward ANNs trained on images remain the dominant models of visual cortex, despite being limited to static inference. In contrast, the primate visual system is a dynamical system operating in a dynamic world. Recent evidence shows that macaque inferior temporal (IT) cortex encodes both object identity and motion velocity. How well do static image-trained artificial neural networks (ANNs) capture IT dynamic computations? Do video-trained ANNs improve upon these static models? We recorded activity from 131 IT sites (two monkeys) while they passively viewed 920 short (300ms) videos of moving objects. We tested 12 feedforward, 2 recurrent, and 13 video-trained ANNs. Feedforward models were unfolded in time by extracting frame-by-frame features to approximate instantaneous, history-independent computations. Video ANNs outperformed other models in decoding motion-direction (400 clips; video ANNs accuracy=0.69, percent correct; feedforward ANNs accuracy=0.58; Δ accuracy=0.109, $p<0.05$) and motion-speed (320 clips; video ANNs accuracy=0.68; feedforward ANNs accuracy=0.63; Δ accuracy=0.049, $p<0.05$) on naturalistic videos, paralleling IT (0.58;0.57) where late responses carried stronger motion information. In a stress test, we generated 100 appearance-free videos (AFV), where appearance was replaced by

random pixels preserving motion trajectories. IT decoders trained on appearance-based responses generalized to above-chance motion decoding for AFV (accuracy=0.59), whereas all ANNs collapsed to chance-level. We then tested how well ANN features predict neural responses. Feedforward models explained a significant portion of IT variance during early responses (~90–180ms; mean %EV=62.5%) but declined significantly later (~480–570ms; %EV=23.1%). Video ANNs showed modest improvements in this late window ($\Delta\%EV=6.9\%$; $t(24)=7.55$; $p<0.01$), suggesting that temporal training aids modeling late-phase IT dynamics. Interestingly, recurrent models performed comparably to feedforward networks, indicating that shallow recurrence alone does not bridge the gap. Together, these findings show that temporal training promotes closer alignment between ANN and IT dynamics, yet video models remain bound to appearance-based temporal cues and fail to capture appearance-invariant motion encoding.

KK is supported by the Canada Research Chair Program (CRC-2021-00326), SFARI (967073), Brain-Canada Foundation (2023-0259), and NSERC (RGPIN-2024-06223). MD was funded by the European Union (MSCA Project 101151834 - PRINNEVOT).

TALK 5, 6:15 PM, 55.25

UNCERTAINTY-WEIGHTED INTEGRATION OF MOTION AND POSITION: PSYCHOPHYSICAL AND FMRI EVIDENCE FROM THE DOUBLE-DRIFT ILLUSION

Ke Yin¹ (yink23@mail2.sysu.edu.cn), Shijia Zhang², Jiyan Zou³, Ce Mo¹; ¹Sun Yat-sen University, ²South China Normal University, ³Fudan University

Visual information from multiple physical dimensions (e.g., motion, position) is integrated to form coherent percepts, yet this process is complicated by sensory uncertainty that is ubiquitous in complex visual scenes. Normative Bayesian accounts propose that information with low uncertainty should be assigned greater weights during integration, but direct behavioral and neural evidence for such uncertainty-weight integration in vision remains scarce. Here, we leveraged the double-drift illusion (DDI) to test how the visual system integrates motion and position signals with different levels of uncertainty. In this illusion, the perceived change of external stimulus position is biased in the direction of the internal stimulus motion. The magnitude of the DDI thus reflects the relative weighting of internal stimulus motion in the integration process. To independently manipulate the uncertainty of motion and position, we designed a novel random dot kinematogram (RDK) stimulus that robustly induced the DDI, where the uncertainty of motion and position was modulated by varying the motion direction variance and the density of the RDK dots, respectively. In a series of psychophysical experiments, we found that increasing motion uncertainty reduced the DDI, whereas increasing position uncertainty enhanced it, indicating that the two signals were weighted according

to their relative uncertainty. Moreover, similar behavioral patterns were found when the RDK stimulus changed only in its spatiotopic position while its retinotopic position remained constant, indicating the same uncertainty-weighted integration process in both the spatiotopic and the retinotopic reference frame. Multivariate fMRI analyses further showed that the neural representation of the DDI in extrastriate visual areas was attenuated by the increasing motion uncertainty. This effect was most pronounced in V2/3 and gradually diminished up the visual hierarchy. Our findings revealed, for the first time, an optimal uncertainty-weighted visual integration mechanism that is implemented in V2/3.

National Natural Science Foundation of China (32471104 to C.M.)

TALK 7, 6:30 PM, 55.26

WHEN FORM MEETS MOTION: RECURRENT CONNECTIONS ENHANCE BIOLOGICAL MOTION PERCEPTION

Shuangpeng Han¹ (shuangpe001@e.ntu.edu.sg), Ziyu Wang¹, Thomas Serre², Mengmi Zhang¹; ¹Nanyang Technological University, ²Brown University

Biological motion perception (BMP) refers to humans' ability to perceive and recognize the actions of living beings purely from their motion patterns, even when those actions are presented only as sparse point-light displays. Neurophysiological studies indicate that BMP in the human brain arises from coordinated processing across the ventral and dorsal visual streams. The seminal two-stream model of Giese & Poggio (2003) formalized this division of labor, proposing parallel form and motion pathways whose dynamic integration supports robust action recognition. However, this model lacks recurrent connections and top-down mechanisms, which are supported by neurophysiological evidence but absent from both classical theories and modern AI systems. Building on this foundation, we introduce the Two-Stream Motion Perceiver (TSMP), which translates the core Giese & Poggio architecture into a modern deep learning framework while incorporating biologically motivated recurrent connections and top-down modulations from memories. TSMP preserves the separation of form and motion pathways, each equipped with memory mechanisms: the ventral pathway encodes and stores body-form structure, while the dorsal pathway learns and memorizes prototypical motion patterns. During inference, TSMP operates in a recurrent manner, wherein each iteration produces top-down signals that progressively select and refine both motion and form representations. We evaluate TSMP, along with humans and a broad set of AI models, using our curated large-scale BMP dataset, which contains 79,744 stimuli spanning 30 BMP conditions grounded in neuroscience. TSMP exhibits notable improvements in out-of-domain generalisation on BMP conditions, surpassing all existing AI baselines by up to 35% in top-1 accuracy and achieving a marked alignment with

human behaviour. Our work offers a biologically grounded pathway toward models that more faithfully capture the computations underlying human biological motion perception.

This research is supported by the National Research Foundation, Singapore under its NRFF award NRF-NRFF15-2023-0001 and Mengmi Zhang's Startup Grant from Nanyang Technological University, Singapore.

TALK 7, 6:45 PM, 55.27

CAUSAL INFERENCE IN VISUAL MOTION AND POSITION PERCEPTION

Jeongbeen Yoo¹ (jeongbinyoo@unist.ac.kr), Oh-sang Kwon¹; ¹Ulsan National Institute of Science and Technology

Most theoretical frameworks of motion perception overlook the fact that motion typically co-occurs with changes in object position. To address this, we previously developed an object-tracking model (Kwon, Tadin, & Knill, 2015). The model accounts for several interactions between motion and position signals, including the Motion-Induced Position Shift (MIPS), where an object's perceived position is biased in the direction of its internal pattern motion. However, internal pattern motion has also been shown to bias boundary position in the opposite direction in near fovea (Zhang, Yeh & De Valois, 1993), a phenomenon that the original object-tracking model cannot explain. Here, we extend the model to improve its generalizability. The key idea is that the model now considers two ecologically valid forward models to interpret the pattern motion inside a boundary. The first case is a moving object with internal pattern motion. In this case, the boundary belongs to the object, and the boundary motion is coupled with the pattern motion (object model). The second case is pattern motion observed through an aperture. In this case, the boundary belongs to the aperture, and the boundary motion and internal pattern motion are independent (aperture model). We assumed that the visual system applies these two forward models to interpret sensory inputs and optimally estimate the state of the world. Results showed that under the object model, boundary estimates shift in the direction of the pattern motion, whereas under the aperture model, they shift in the opposite direction. Furthermore, if we assume that the forward model with higher likelihood dominates perception, the direction of the bias changes systematically with stimulus properties, consistent with empirical data. Overall, the extended model provides a rational account of how visual motion and position signals interact, reconciling previously contradictory findings by incorporating ecologically valid forward models.

This research was supported by the National Research Foundation of Korea (NRF-2023R1A2C1007917 to O.-S.K.).

TALK 8, 7:00 PM, 55.28

VISIBILITY OF GROUND CONTACT POINT CAN ELIMINATE DEPTH MISPERCEPTIONS CAUSED BY INTEROCULAR DELAYS

Anthony LoPrete¹ (aloprete@seas.upenn.edu), Johannes Burge¹; ¹University of Pennsylvania

Abstract: Blur and luminance differences between the eyes introduce millisecond-scale interocular delays that can cause dramatic depth misperceptions for moving objects. Monovision corrections for presbyopia—which intentionally induce interocular blur differences—are worn by millions of people worldwide. The attendant depth misperceptions could therefore have public safety implications. However, despite the extreme consistency of laboratory measurements (and one high-profile airplane crash), reports of motion illusions (and/or accidents) among monovision wearers are rare in daily life. One potential reason is that in real-world viewing, additional cues (e.g. ground contact points) indicate the true instead of incorrect depths. We used a Magic Leap 2 augmented reality headset to measure how viewing context affects delay-induced depth misperceptions. A trial-randomized experiment presented virtual target objects—sometimes rendered with interocular luminance or blur differences—that were depth-consistent with the natural environment. On half the trials, an occluder blocked visibility of where the target contacted the ground. The target was an upright rectangular polygon that moved left or right across the scene (1.7-3.9mph) at distances ranging from 1m to 2m behind the occluder. The moving target, occluder, and ground were textured with 1/f noise. To increase the strength of ground contact cues, the ground was also littered with randomly positioned virtual 'grass blades'. Textures and grass blades were refreshed on each trial to prevent observers from using consistent visual landmarks. Observers reported the perceived target distance by placing a virtual marker on the ground. When the occluder obscured the target contact point with the ground, target distances were substantially misperceived (e.g. >1.0m); errors increased with target speed. When the occluder was absent, the illusion was nearly eliminated. The data help explain the reduced frequency of effects in real-world viewing, and show that, in certain real-world contexts, monovision can still cause pronounced misestimations of depth.

This work was supported by the National Eye Institute and the Office of Behavioral and Social Sciences Research, National Institutes of Health Grant R01-EY028571 to J.B. and an Ashton Fellowship from the University of Pennsylvania to A.L.

Poster Sessions

FRIDAY AFTERNOON POSTERS IN

BANYAN BREEZEWAY

Visual Working Memory: Performance, influences

FRIDAY, MAY 15, 3:45 – 6:00 PM, BANYAN BREEZEWAY

16.301 LIMITED ENCODING WEAKENS CATEGORY BENEFITS, WHILE LONGER ENCODING REVEALS STRONG SEMANTIC EFFECTS

Payachana Victoria Chareunsouk^{1,2,3}, Yong Hoon Chung⁴, Viola Störmer⁴, Chaipat Chunharas^{1,2,3}; ¹Cognitive Clinical and Computational Neuroscience Center of Excellence, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand, ²Division of Neurology, Department of Medicine, Faculty of Medicine, King Chulalongkorn Memorial Hospital, Thai Red Cross Society, Bangkok, Thailand, ³Chula Neuroscience Center, King Chulalongkorn Memorial Hospital, Thai Red Cross Society, Bangkok, Thailand, ⁴Department of Psychological and Brain Sciences, Dartmouth College

When remembering multiple objects, people often group them into chunks to improve memory performance. Evidence shows that people group items by low-level visual features, but real-world objects also share semantic relationships that meaningfully shape working memory. Because forming a semantic gist may require time, we tested whether shortening encoding duration would weaken the semantic benefit or shift reliance toward high-level category labels rather than object-specific details. In an earlier study with a 5000 ms encoding time, arrays containing more items from the same category produced higher accuracy but also more within-category false alarms, indicating reliance on semantic gist. The present study examined whether this benefit persists under brief viewing. Twelve participants completed a delayed-match-to-sample task with five real-world objects under four semantic-composition conditions (five, four, three, or zero items from the same category). Arrays were shown for 500 ms followed by a 2000 ms delay. Semantic structure influenced memory under both encoding durations, but the strength of the effect differed. With long encoding, same-category arrays produced a robust semantic benefit (estimate = 1.075, $t(14.6) = 3.442$, $p = .0037$). Under brief encoding, semantic benefit was smaller and not reliable (estimate = 0.442, $t(14.6) = 1.414$, $p = .178$). Same-category false alarms also increased from 0.0719 at 5000 ms to 0.1387 at 500 ms. Increases in false alarms under brief encoding reflect either greater reliance on category-level similarity

when item details are weak or more general errors when semantic structure is not extracted. Overall, semantic structure mattered at both durations, but the size of the effect varied with encoding time. Shorter exposure reduced semantic benefit and constrained the use of categorical information.

16.302 CONTEXTUAL DISTINCTIVENESS SHAPES THE SELF-PROCESSING ADVANTAGE

yueyao liu¹ (yueyao.liu@zju.edu.cn), pengmin qin², hui chen³; ¹Zhejiang University, ²South China Normal University, ³Zhejiang University

The self-processing advantage (SPA)—prioritized processing of self-relevant stimuli—is widely considered a stable, intrinsic phenomenon. We challenged this view using a novel design that isolates contextual influences on the SPA while holding task structure constant. Across three experiments ($N=320$), participants searched for a target name among distractor nouns and reported its location. While the name's identity had to be attended to define the target, it was incidental to the report. A surprise test probed automatic memory for the name. Crucially, we systematically manipulated the relative distinctiveness of the self-stimulus by varying the categorical composition of the target-name pool across experiments. The results showed that the SPA is context-dependent. In the first series of experiments, a robust memory advantage emerged only when the self-stimulus was the most distinctive item relative to the other potential targets. This benefit was self-specific and did not extend to other highly familiar names (e.g., a friend's name) under comparable distinctiveness conditions. Most critically, when the self-stimulus was embedded in a mixed context that reduced its relative categorical distinctiveness, the SPA was completely eliminated and the processing advantage was instead biased towards contextually novel information. These results provide direct evidence that the SPA is not a fixed attribute of self-representation. Instead, it reflects a flexible, competitive process dynamically shaped by the immediate processing context. We conclude that the self gains cognitive priority only when intrinsic self-relevance aligns with contextual distinctiveness; otherwise, it can be overridden by other salient signals, such as novelty.

16.303 SEMANTIC AND FUNCTIONAL RELATEDNESS IMPROVE SPATIAL PRECISION BUT DRIVE DIFFERENT MEMORY BIASES

Kanokwanwijiit Wongsook¹, Payachana Chareunsouk^{1,2}, Sedthapong Chunamchai^{1,3}, Chaipat Chunharas^{1,3,4}; ¹Cognitive Clinical and Computational Neuroscience Center of Excellence, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand, ²Medical sciences program, Division of Neuroscience, Faculty of Medicine, Chulalongkorn university, Bangkok, Thailand, ³Division of Neurology,

Department of Medicine, Faculty of Medicine, King Chulalongkorn Memorial Hospital, Thai Red Cross Society, Bangkok, Thailand,
⁴Chula Neuroscience Center, King Chulalongkorn Memorial Hospital, Thai Red Cross Society, Bangkok, Thailand

Previous studies suggest that functional relatedness, referring to items' interactions based on their function, and semantic relatedness, referring to shared contextual associations between items, both improve item recall memory, but their impact on spatial precision and spatial biases remains unclear. We hypothesized that relatedness would improve spatial precision through grouping and simultaneously influence spatial biases: items might be stored separately and later pulled together, creating attractive or repulsive bias, or stored as a single unit, reducing bias. Here, we examined the effects and interaction of functional and semantic relatedness on spatial precision and spatial bias. Across 480 trials, real-world related items were presented under four conditions: functionally and semantically related, functionally related only, semantically related only, and unrelated. Participants were asked to drag items to their original locations after a 1400 ms delay. We quantified spatial accuracy using three Euclidean distance-based metrics: precision, absolute bias, and relative bias. Results from fifteen participants showed higher precision for functional than non-functional related items ($F(1,14) = 71.84, p < .001$). Semantic related items were remembered more accurately than non-semantic related items ($F(1,14) = 7.89, p = .013$). Semantic benefit was observed only for functional related items ($F(1,14) = 33.93, p < .001$) and was not observed for non-functional related items. Non-functional related items showed higher attractive bias than functional related items on absolute ($F(1,14) = 61.24, p < .001$) and relative positions ($F(1,14) = 30.13, p < .001$), indicating stronger attractive biases. Overall, functional and semantic relatedness enhanced spatial precision while non-functional related items show stronger attractive bias, suggesting initial separate encoding and later grouping. Together, these results highlight how high-level relationships structure spatial memory.

The authors wish to acknowledge the valuable contributions of the CCCN members for their support, data provision, and insightful discussions.

16.304 CHANGES OF MIND IN VISUAL WORKING MEMORY RETRIEVAL

Adam Triabhall¹, Ivan Tomic², Paul Bays¹; ¹University of Cambridge, ²University of Zagreb

Storage processes in visual working memory (VWM) have been extensively investigated, but retrieval mechanisms are less well understood. Previous work has drawn a parallel between VWM retrieval and perceptual decision making, hypothesizing that both are underpinned by sequential accumulation of evidence towards decision bounds. In perceptual tasks, strong evidence for accumulation-to-bound has come from "changes of mind", dynamic reversals of initial decisions explained by continuing accumulation of evidence after

response initiation. To test for the existence of changes of mind in retrieval from VWM, we conducted four experiments using a modified cued-recall design in both online and laboratory settings. On each trial, participants were shown an array of visual objects and subsequently required to report the location of one target object – identified by a cue matching one or more of its non-spatial features – via a rapid movement to its remembered location within a circular response space. In different experiments we manipulated speed pressure, set size, and cue feature dimension(s). We developed a novel method for detecting and quantifying changes of mind based on movement trajectories within a circular response boundary, revealing their consistent presence at low frequencies in all experiments. Spontaneous response updating occurred predominantly between locations of different objects that had been present in the memory array, consistent with continuing accumulation of evidence regarding cue features after response movement initiation. Their prevalence increased with manipulations predicted to decrease memory evidence strength, including set size. Importantly, changes of direction away from correctly identified targets toward incorrect locations were less frequent than the converse, suggesting an adaptive mechanism that supports accurate retrieval. To our knowledge, these findings provide the first evidence for changes of mind in VWM, identifying accumulation-to-bound as a fundamental component of the retrieval mechanism, and establishing a new source of behavioral evidence for investigating working memory.

BBSRC; Gonville & Caius College

16.305 SLEEP QUALITY AFFECTS THE OFFLINE STORAGE OF VISUAL WORKING MEMORY

Li Gong^{1,2} (gl199409@126.com), Penglan Liu², Dan Nie², Chaoxiong Ye^{1,2}; ¹School of Education, Anyang Normal University, China, ²Department of Psychology, University of Jyväskylä, Finland

Visual working memory (VWM) is a core cognitive function in humans, with memory representations flexibly stored in distinct representational states (online and offline) depending on task demands. Among these, the offline storage state plays a crucial role in stably maintaining memory representations and supporting the online storage state. Previous research has demonstrated that an individual's sleep quality can predict performance in VWM tasks. However, these studies have primarily focused on the online storage state of VWM. It remains unclear whether sleep quality affects the offline storage state, and whether its influence differs between the two storage states. To address these questions, the present study employed a mixed experimental design, combining the Pittsburgh Sleep Quality Index (PSQI) with a sequential presentation retrieval paradigm to examine the effects of sleep quality on different storage states of VWM. The results revealed significant main effects of both sleep quality and memory storage state on VWM accuracy. Moreover, a significant interaction between sleep quality and memory storage state was observed. Simple effect analyses indicated that, in the offline storage state, participants with high sleep quality exhibited significantly higher

accuracy than those with low sleep quality. These findings suggest that sleep quality differentially affects the distinct storage states of VWM. Specifically, when individuals experience poor sleep quality, the offline storage state appears to be more adversely impacted than the online storage state. This study provides direct evidence supporting the hypothesis of a functional dissociation between the online and offline states of VWM, and further elucidates the intrinsic relationship between sleep quality and VWM. The findings underscore the critical role of sleep in cognitive functioning and contribute to a deeper understanding of this domain.

This work was supported by the Research Council of Finland (former Academy of Finland) Academy Research Fellow project (#355369 to Chaoxiong Ye) and Finnish Cultural Foundation (#231373 to Chaoxiong Ye).

16.306 IMPACT OF SELF-RELATED INFORMATION ON WORKING MEMORY PRECISION: EVIDENCE FROM A PERCEPTUAL MATCHING PARADIGM

Rudan Xue¹, Yang Sun¹; ¹Shenyang Normal University, Liaoning, China

Ample evidence indicates a processing priority for self-related information, manifested as advantages in attentional capture and memory enhancement. Working memory serves as a finite cognitive resource system supporting the short-term storage and manipulation of information, and its performance is modulated by cognitive load and temporal factors. However, while it is established that self-related information can influence memory recognition, whether it similarly affects perceptual precision in working memory remains to be further explored. Therefore, this study conducted four experiments to examine how cognitive load (2 vs. 4 items) and retention interval (1000 ms vs. 2000 ms) modulate the self-advantage effect in visual working memory. All experiments used an associative learning task followed by a delayed estimation task. The “Self” condition showed a stable reaction-time advantage across all experiments, while the “Meaningless” condition also yielded rapid responses, likely reflecting lower semantic interference. In terms of memory precision, self-related information significantly outperformed the “Stranger” condition at the 1000 ms interval (Experiments 1–2). Under high load, precision for “Self” exceeded that for “Friend.” Under low load with a 2000 ms interval (Experiment 3), the main effect of social labels persisted, with both “Self” and “Friend” showing higher precision than “Stranger.” However, under high load at 2000 ms (Experiment 4), “Self” and “Meaningless” labels showed higher precision—surpassing both “Friend” and “Stranger”—demonstrating the stability of self-related processing. Overall, self-related information exhibits a robust advantage in working memory, marked by efficient retrieval. In contrast, friend-related benefits are conditional, depending on available cognitive resources and retention intervals, and incur notable retrieval costs. This study shows that social-information processing in working memory is dissociated, resource-dependent, and evolves

over time; critically, retrieval efficiency advantages are more fundamental and enduring than those in storage precision.

Multisensory Processing: Motor

FRIDAY, MAY 15, 3:45 – 6:00 PM, BANYAN BREEZEWAY

16.307 GAZE DYNAMICS REVEAL THE PROTRACTED DEVELOPMENT OF PERCEPTUAL SEGREGATION OF MULTIPLE TALKING FACES TO SOLVE THE MULTISENSORY COCKTAIL PARTY PROBLEM

Katia steinfeld¹ (katia.steinfeld@unil.ch), Micah Murray^{1,2}, David Lewkowicz³; ¹The Lausanne University Hospital and University of Lausanne, ²The Sense Innovation and Research Center, ³Child Study Center, Yale School of Medicine

Often, we are confronted with several people speaking all at once. Successfully identifying a relevant talker requires us to bind and integrate the audible and visible speech streams of that talker (target) and segregate them from competing talkers (distractors). How these multisensory abilities emerge during early development is poorly understood. We used metrics of gaze dynamics derived from information theory to quantify children's and adults' ability to extract the perceptual cues necessary to solve the multisensory cocktail party problem (MCP). While gaze-tracked, children aged 3–7 years ($N=172$) and adults ($N=37$) viewed four talking faces and heard an auditory utterance that was either temporally synchronized with one face or desynchronized from all faces. Participants' task was to identify the “talking” (i.e., audiovisually synchronized) face. Dynamic gaze behavior was quantified using dwell time, stationary gaze entropy and transition entropy. All children, including 3-year-olds, showed greater dwell time on the synchronized target than on distractors ($F(1,172)=18.76$, $p<.001$), indicating an early sensitivity to audiovisual temporal coherence. However, only from 5 years of age did children reliably concentrate gaze on the synchronized target ($F(4,172)=12.11$, $p<.001$). Beginning at 6 years, synchronized cues restricted the exploration to a subset of distractors ($F(4,172)=3.50$, $p=.009$) and produced a more predictable sequence of transitions ($F(1,167)=12.34$, $p<.001$). Adults showed the strongest dwell-time preference for the target ($F(1,37)=1448.30$, $p<.001$) as well as the greatest reductions in stationary ($F(1,37)=434.79$, $p<.001$) and transition entropy ($F(1,35)=167.81$, $p<.001$) when cues were synchronized. Together, these findings reveal a qualitative shift in dynamic gaze control from 5 to 6 years, marked by greater concentration of gaze on the target and more structured transitions among distractors. We interpret it as reflecting (a) improved multisensory integration, increasing the perceived salience of the target, and (b) the emergence of task-dependent weighting of salience cues, supporting efficient segregation of multisensory scenes.

16.308 EQUATING AND COMPARING THE IMPACT OF VISUOMOTOR TRANSFORMATIONS ON BIMANUAL MULTI-STAGE MOVEMENTS USING IMMERSIVE VIRTUAL REALITY

Rea Gill¹, William Harwin¹, Harun Tugal², Peter Scarfe¹; ¹University of Reading, ²UK Atomic Energy Authority (UKAEA), Remote Applications in Challenging Environments (RACE), Culham Campus, Abingdon, UK

Remote operation, or teleoperation, systems are typically used in unsafe or inaccessible environments, such as in nuclear decommissioning. Here, human operators must overcome visuomotor transformations, which result in a mismatch between their movement and the visual feedback they receive. Despite this being a typical feature of teleoperation systems, we do not know which aspects of these transforms (e.g. altered translation, rotation, or scale) are most detrimental to task performance and user experience. Here, in a bimanual virtual reality task, we have examined the impact of translation, rotation, and scale mismatches between a user's movement and the visual feedback they receive. Participants completed a "pick up and place" task where they picked up a cube with their left hand, passed it to their right hand, then placed it in a box. To equate the transforms, we use a two-part study in which we first determine the 'just noticeable difference' (JND) for each transform using a 2AFC paradigm. Next, using the JNDs (1, 2, and 3) to equate each transform in terms of sensitivity, we examined which (if any) is more detrimental to task performance and user experience. Overall, as expected, we find that increasing transformation magnitude systematically decreases task performance. We present a range of well-established task performance metrics (Bradshaw, Elliott, Watt, Hibbard, Davies, and Simpson, 2003) such as task completion time, smoothness of movement, and time in slow phases of the movement, as well as self-reported measures (workload, effort, ownership, agency, presence). Our results highlight the importance of considering human factors when developing teleoperation systems. Notably, we apply rigorous psychophysical methods to a more complex multi-stage movement than is typical in the field, but better represents real-world use cases.

This research was generously supported by the Southeast Network for Social Sciences Doctoral Training Partnership (SeNSS DTP) and EUROfusion.

16.309 VISUO-MOTOR GAIN ADAPTATION FOR VISUAL STABILITY DURING HORIZONTAL HEAD TRANSLATION WITH AND WITHOUT BINOCULAR DISPARITY

Nobuyoshi Takase¹ (takase.nobuyoshi.og@tut.jp), Hiroaki Shigemasa², Michiteru Kitazaki¹; ¹Toyohashi University of Technology, ²Kochi University of Technology

Perceiving a stable visual world despite head movements requires accurate visuo-motor coordination. Kim et al. (Front. VR, 2021) showed that during horizontal linear head oscillations, increasing visuo-motor gain conflict makes the scene look more unstable. Kitazaki (i-

Perception, 2013) demonstrated that visuo-motor adaptation of visual stability emerges after two minutes of yaw-axis rotation and is not strictly specific to retinal location. However, for linear head movements, it remains unclear how prolonged exposure alters the visuo-motor gain that observers perceive as yielding a stationary scene, and how other aspects of self-motion perception and posture respond to gain manipulation (Teng et al., Sci.Rep., 2025). The present study investigated how the visuo-motor gain for stable perception adapts to different gains, with or without binocular disparity. Twenty-seven participants viewed a virtual cloud of random spheres while their head movements were tracked with a 3D-tracker-embedded head-mounted display. The virtual viewpoint was updated according to the measured horizontal head translation with a gain of 0.5, 1.0, or 2.0. Each session consisted of pre-test, adaptation, and post-test trials. In the pre- and post-tests, participants adjusted the visuo-motor gain during left-right head movements so that the visual environment appeared stable. During the 5-min adaptation phase, they made left-right head movements (30 cm, 0.33 Hz) while viewing the spheres at a fixed gain (0.5, 1.0, or 2.0) either with or without binocular disparity. Each participant completed six sessions (3 gains $\times$ 2 disparity conditions) in random order. We found a significant main effect of gain ($F(2, 52)=5.38, p=.008$), but no main effect of disparity and no interaction. The visuo-motor gain that produced a stable scene in the post-test shifted toward the adaptation gain, both with and without binocular disparity. These results suggest that visuo-motor gain adaptation for stable perception occurs for linear head movements and is largely independent of binocular disparity.

Supported by JSPS Kakenhi Grant Numbers JP22KK0158, Murata Science and Education Foundation

16.310 NAVIGATION IN VIRTUAL REALITY: THE ROLE OF LANGUAGE AND LANGUAGE BACKGROUND

Kavindya Dalawella¹ (kd492@rutgers.edu), Jakub Suchojad¹, Jacob Feldman¹, Karin Stromswold¹; ¹Rutgers University

We investigated how people integrate visual, social, and linguistic information in dynamic environments to perform complex tasks. The environment was a virtual reality (VR) train station populated by moving and static VR "people" and objects. The task was to physically navigate to a location based on a spoken announcement. Participants were assigned a destination city at the start of the experiment. In half of the trials (go trials), the announced city matched participants' assigned city and they proceeded to the announced track; in no-go trials, the city didn't match, and participants returned to a "waiting room". We used two linguistic frames to vary how early a trial could be identified as go/no-go: "City First" ("The next train to CityX is now boarding on Track#") or "Track First" ("Now boarding on Track# is the next train to CityX). "City First" trials could be identified as go-trials earlier, while "Track First" trials required participants to remember track numbers and wait. Participants were university students who said English was their best language. Half were native English (NE) speakers with NE-speaking parents and half learned English in early

childhood, but didn't have NE-speaking parents. Position, velocity, head direction, eye gaze and pupillometry data were collected. Results indicate that task, information order (linguistic frame) and language background all affected participants' behavior. Both groups responded faster in no-go tasks (task effect) and completed trials faster in "City First" trials where they didn't have to wait to begin planning (information order). Notably, language background also influenced performance: participants with NE-speaking parents took less time overall and identified trials as go/no-go faster. These findings indicate adults efficiently integrate visual, social and linguistic information to plan and execute a complex navigational task, but early childhood language experience modulates how quickly they use linguistic information.

Supported by the NSF-NRT grant: Socially Cognizant Robotics for a Technology Enhanced Society (SOCRATES), No. 2021628. Additional support provided by NSF BCS Grants 2324598, 2122119, & Rutgers Research Council Grant.

16.311 IS VISUOMOTOR ADAPTATION BASED ON FEATURES OR OBJECTS?

Samuel McDougle¹, Dawei Bai¹, Brian Scholl¹, Zekun Sun; ¹Yale University

Visuomotor adaptation is thought to be driven by sensory prediction errors – the difference between predicted and observed visual feedback given an outgoing motor command. In typical tasks in this domain, visual feedback is given in the form of objects such as a white cursor moving on a black background. As a result, the ensuing adaptation could be driven either by the persisting lower-level visual feature (a white patch moving across a black background), or by the resulting object representation itself. To find out requires a stimulus that at first blush may sound impossible: a persisting object without any persisting features. That is what we explored in the present project, exploiting "featureless objects." In Experiment 1, subjects moved their cursor in a workspace tiled with blue squares of varying brightness. As they moved, each square their hand entered changed from one random shade to another. These brief transients supported the perception of a persistent, trackable object, even though no surface feature was preserved across frames and no static frame allowed for object identification. Feedback was then rotated relative to the angle of subjects' hand movements, and adaptation was measured via implicit aftereffects (i.e., automatic compensatory adjustments to the induced errors). Subjects adapted robustly, and aftereffects were larger under featureless object feedback versus conventional cursor feedback. In Experiment 2, we extended this finding to another form of featureless object feedback: random pixel shuffling within a static background. We again observed adaptation, and observed a canonical cosine-shaped generalization function when aftereffects were measured at nearby reaching directions not experienced during training. In Experiment 3, we discovered that different forms of featureless object feedback could be tied to the same visuomotor memory, observing complete transfer of adaptation between the two

feedback contexts. We conclude that implicit visuomotor adaptation is fundamentally an object-based process.

NIH R01 NS134754 (S.D.M.)

Multisensory Processing: Visual-tactile

FRIDAY, MAY 15, 3:45 – 6:00 PM, BANYAN BREEZEWAY

16.312 PUPILLARY RESPONSE TO SURFACE PROTRUSION INTENSITY ACROSS VISUAL AND TACTILE, VISUO-TACTILE MODALITIES

Sayoon Kim¹, Chai-Youn Kim¹; ¹Korea University

Pupillary response is known to vary parametrically with sensory intensity across modalities, reflecting automatic physiological tracking of stimulus magnitude (Wang & Munoz, 2015). While this relationship is well established in vision and audition (Binda & Murray, 2015; Zekveld et al., 2018), evidence in the tactile domain remains limited and inconsistent (Gusso et al., 2021). The present study examined how pupil dynamics change as a function of protrusion degree from a surface across visual, tactile, and visuo-tactile modalities. Stimuli were constructed as six degrees of protrusion, ranging from a flat surface (S0) to progressively increased intensity (S1–S5), presented as 3D-rendered images in the visual condition and 3D-printed objects in the tactile condition. Ten participants first completed a discrimination task to assess their ability to distinguish protrusion degrees in the visual and tactile modalities. Seven of these participants proceeded to the main experiment, in which they rated the perceived intensity of the stimuli in the visual, tactile, and visuo-tactile conditions while pupil diameter was continuously recorded with an eye tracker (Eyelink 1000, 1000Hz). Baseline-corrected pupil diameter was analyzed in two post-stimulus windows: an early window (500–1000ms) and a late window (1500–2000ms). A positive correlation between protrusion intensity and pupil dilation was significant in both time windows for the visual and visuo-tactile conditions, with the visual condition showing a particularly strong effect in the late window ($r = 0.681$, $p < 0.0001$). In the tactile condition, a positive correlation was found in the early window ($r = 0.415$, $p = 0.0412$) only when the flat surface (S0) was excluded. These findings indicate that pupillometry reflects the systematically varied intensity of the same physical surface property across visual, visuo-tactile, and, to some extent, tactile modalities.

Supported by NRF-2023R1A2C2007289

16.313 VISUAL-TACTILE PERCEPTIONS OF GARMENTS: PSYCHOPHYSICAL INSIGHTS INTO RELATIONSHIPS BETWEEN MATERIAL APPEARANCE AND VISION AND TOUCH

Molly Talbot^{1,2} (mn18met@leeds.ac.uk), Kaida Xiao^{1,2}, Ningtao Mao^{1,2}; ¹University of Leeds, ²Leeds Institute of Textiles and Colour

Fashion retail is an increasingly online market. Online, consumers rely on visual information without touch to infer garment feeling in a process termed visual-tactile perception. Vision acting as a proxy for touch introduces discrepancies between visual predictions and tactile perceptions. The relationship between visual appearance and tactile sensation is complex and not fully understood for garments as complex three-dimensional stimuli. This study aimed to examine whether visual judgements of tactile properties align with actual touch perception by analysing correlations across conditions, validating insights from prior fabric research while expanding understanding into three dimensions, considering both garment substance (surface texture, fibre composition) and structure (construction, mechanics) properties. A psychophysical scaling experiment was conducted with 37 garments representing varied categories and constructions. 20 participants rated each garment on 10 bipolar tactile dimensions (e.g., warm-cool, stiff-flexible, smooth-rough) using a 14-point interval scale under four conditions: touch-only, viewing on-screen, viewing in real life, and combined vision and touch. On-screen stimuli comprised high-resolution images and videos captured using a Sony Alpha series camera and presented on an EIZO display to preserve material appearance features like gloss, texture, and colour. Post-condition interviews provided qualitative insights into associations between appearance cues and tactile judgements. Findings reveal clear distinctions between surface and mechanical properties. Surface cues, including the presence and uniformity of gloss, showed strong associations with smoothness and slipperiness, while surface density evidenced the strongest cross-modal correlations. Mechanical properties were mixed: stretchiness and flexibility were visually predictable, whereas recoverability from creases and deformation showed poor agreement. Across conditions, visual-tactile judgements correlated most strongly. This study reveals vision's strong relationship with touch in perceiving properties of surface and density and their inconsistencies in perceiving structure and mechanics, highlighting both the predictive capabilities of consumers' vision and priors and the limitations of vision as a proxy for touch.

This work is funded by the Leeds Institute of Textiles and Colour (LITAC)

16.314 INTEGRATING HAPTIC CUES DISTINCTIVELY IMPROVED JUDGEMENT OF ANGER AND DISGUST FACIAL EXPRESSIONS

Cameron Mavericks Choo¹ (cameronm001@e.ntu.edu.sg), Hong Xu²; ¹Interdisciplinary Graduate Programme (Neuroscience), Nanyang Technological University, ²School of Social Sciences, Nanyang Technological University

Anger and disgust are often confused in facial-expression recognition tasks due to overlapping visual cues for emotional judgements. Integrating information from different sensory modalities may yield more reliable and accurate affective cues—multisensory integration. In this study, we investigated whether visual-haptic integration could improve the performance to differentiate these emotions. Forty-seven

participants participated in a 2 (emotion: anger, disgust) x 3 (modality: visual, haptic, visual-haptic) within-subjects emotion categorization task incorporating a "none of the above" option to mitigate possible forced-choice artifacts. Accuracy, response latency and confidence level of their judgements were also measured. Stimuli consisted of facial expressions from the NTU Asian Face Database and affective human haptic gestures developed and validated in our lab. Trial-level responses were analyzed to assess perceptual performance across conditions. Poisson regression with model comparisons were used to assess classification patterns. We found that modality played a significant role in response accuracy ($\chi^2(12) = 566.08$, $p < .001$). Post-hoc tests with Bonferroni correction indicated that visual-haptic integration significantly increased correct identification frequencies for both emotions relative to visual trials (p 's $< .001$). Correspondingly, misclassification rates (i.e., wrong choice of emotion) were reduced for anger ($p < .001$), but not for disgust ($p = .098$), suggesting that uncertainties in identifying disgust (i.e., 'none of the above') were mitigated while misclassifications of disgust as anger persisted even with visual-haptic integration. Mixed-effects models further revealed that this multisensory enhancement extended beyond simple error correction: visual-haptic conditions yielded higher accuracy probabilities, shorter response latencies, and higher confidence levels compared to unimodal conditions (p 's $< .001$). These findings demonstrate that haptic cues effectively disambiguate facial expressions via emotion-specific mechanisms, shifting from reducing false perceptions of anger to enhancing signal salience of disgust. Furthermore, it suggests that multisensory integration provides distinct functional benefits across different affective categories in emotion perception and judgement.

Neuroscience, Interdisciplinary Graduate Programme, Nanyang Technological University, Singapore

Eye Movements: Cognition

FRIDAY, MAY 15, 3:45 – 6:00 PM, BANYAN BREEZEWAY

16.315 IS THERE A COMMON ATTENTIONAL MECHANISM BETWEEN READING AND VISUAL SEARCH?

Grace Sinclair¹, Joseph Schmidt¹; ¹University of Central Florida

Navigating via signage or searching for specific words involves aspects of reading and visual search. Both tasks involve peripheral pre-processing that informs upcoming eye movements. However, these tasks require different oculomotor behaviors, search typically exhibits a leftward attentional bias (Durgin et al., 2008), whereas left-to-right readers preferentially spread attention rightward to integrate upcoming textual information (Schotter et al, 2012; Rayner, 1998). We asked if larger perceptual span sizes in reading (the area from which visuospatial information is extracted from the upcoming text outside of

central vision; McConkie & Rayner, 1975) predict increased guidance of spatial attention to target words (as measured by the proportion of trials where the target was fixated first) during a visual search task. We predicted stronger associations at near eccentricities and on the right side of space due to the rightward spread of attention during left-to-right reading. Observers read through sentences with various gaze-contingent moving window sizes while eye movements were tracked to obtain span size. They then searched through arrays containing one target word and five distractor words. We found that larger span sizes result in stronger guidance on the right side of space, and the leftward search guidance bias was only present for individuals with smaller span sizes. This suggests that there is a partially common attentional mechanism, and larger span size provides a compensatory mechanism that results in more similar search guidance on each side of space. Next, we asked if this relationship extended to non-linguistic searches. To assess this, we repeated the study with search arrays containing one real-world target object and five categorically distinct distractors. Early results suggest that the relationship may not extend to non-linguistic information. This demonstrates that readers with larger spans may use linguistic information more effectively, but this advantage does not extend to other types of searches.

16.316 THE EFFECTS OF TEMPORAL LIGHT MODULATION ON READING SPEED AND EYE MOVEMENTS

Tom Foulsham¹ (foulsham@essex.ac.uk), Arnold Wilkins¹; ¹University of Essex

The widespread use of LED lighting, which can flicker at a high rate, has raised questions about the effects of this temporal light modulation (TLM) on visual tasks such as reading. In particular, TLM has been linked to visual discomfort and may disrupt eye movements. This may be because during a saccade the TLM can generate a spatial pattern known as a “phantom array”. This may interact and interfere with the pattern formed by the letter strokes. We measured reading speed and eye movements in two tasks. In the first task, participants read aloud monosyllabic, high-frequency words presented in a random order as a paragraph. In the second task, participants read meaningful paragraphs, silently, before answering a series of comprehension questions. In both tasks the text was presented on paper and illuminated by LEDs with TLM at 120 Hz, 600 Hz or 60 kHz. Illumination was modulated as a square wave, with a duty cycle of 30%, conditions in which the phantom array is maximally visible. Participants showed slower reading speeds with text containing words with high horizontal autocorrelation. There were also small effects of TLM, with slower reading in both tasks at 600 Hz, where the phantom array is easily visible. We discuss further nuanced effects of TLM on fixation duration and position, as well as interactions with individual differences in visual sensitivity. We conclude that striped text and TLM both slow reading, perhaps due to disrupted eye movements.

16.317 THE EFFECTS OF AUDIO PLAYBACK SPEED DURING READING-WHILE-LISTENING ON EYE MOVEMENTS AND COMPREHENSION.

Ryan Ringer¹ (ryan.ringer@ucdenver.edu), Mia Jannika Lim¹, Carly Leonard¹; ¹University of Colorado, Denver

Reading proficiency relies on the intricate interplay between oculomotor behavior and attention to build a representation of the text in working memory. Reading-while-listening (RWL), in which reading is paired with an audio recording of the text, is frequently used both in and out of the classroom to help young readers and second-language learners develop this skill. While some previous research has shown that RWL may aid comprehension, other studies suggest that supplemental auditory input may be interfering. For instance, if the rate of the auditory playback does not match reading speed, the text and recording would then become incongruous and potentially interfere with comprehension. Alternatively, any concurrent auditory processing could interfere with reading. Thus, the conditions in which RWL improves versus impairs reading performance remain an open question. The current study addresses this question by testing whether the audio playback speed of the text moderates reading comprehension and eye movement behavior. Twenty-four fluent English-speakers read eight passages on various topics while their eye movements were recorded. Participants heard audio recordings of passages at either slow (135 wpm), medium (180 wpm), or fast (225 wpm) playback speeds, as well as a silent condition. Passages were followed by multiple-choice questions to assess comprehension. The results showed the presence of auditory playback had several impacts on eye movement behavior. Relative to silent reading, mean fixation duration increased and words per minute decreased for all playback speeds, suggesting RWL reduced processing speeds. Saccade amplitude and word skipping decreased during RWL, suggesting a reduced perceptual span. Lastly, regression behavior increased significantly in the slow condition, suggesting either encoding was impaired or participants backtracked to maintain alignment. Nevertheless, reading comprehension was unaffected by RWL. These results indicate that, while RWL does not necessarily affect reading comprehension, it may interfere with visual processing and eye movement guidance.

This research was sponsored by the Ronald E. McNair Postbaccalaureate Achievement Program, awarded to Mia Jannika Lim.

16.318 THE TRAVELING EYEBALL PROBLEM: SPATIAL OPTIMIZATION IN SCANPATHS

Rithwik J Cherian¹ (rcherian@ur.rochester.edu), Ralf M Haefner¹, Dora Biro¹; ¹University of Rochester

Humans solve the Traveling Salesman Problem (TSP)—a classic optimization problem of finding the shortest path through a set of locations—with remarkable proficiency, producing near-optimal tours with minimal effort and training. This ability is particularly surprising

given the problem's computational complexity, yet the underlying mechanisms remain poorly understood. Theories and models of human performance on TSP emphasize perceptual processes such as boundary detection and hierarchical clustering, but direct evidence linking visual processing to spatial optimization is limited (MacGregor and Chu, 2011). Traditional TSP studies rely on manual reporting via pen-and-paper or mouse clicks, requiring deliberate route construction that engages both perceptual and cognitive processes. We introduce a novel paradigm using eye movements as the reporting modality, allowing us to directly capture the perceptual system's implicit spatial optimization strategies as they unfold in real-time, bypassing explicit planning and working memory demands. We examined scanpath efficiency using a vanishing-dots paradigm where participants explored displays of randomly distributed points (5-25 points) that disappeared after sustained fixation (100ms). This gaze-contingent design provides immediate feedback while reducing visual working memory demands and encouraging complete spatial coverage. Using high-precision eye tracking, we extracted scanpaths and compared them to optimal and heuristic TSP solutions. We found that scanpaths exhibited systematic spatial efficiency. Performance, measured as percent above optimal, ranged from 3-15%, degrading predictably as problem size increased. Participants terminated their routes closer to optimal endpoints than expected by chance, suggesting multi-step saccade planning. Median fixation duration remained approximately constant (230-280ms) within participants across problem sizes, indicating consistent local processing independent of global complexity. These results demonstrate that the visual system produces spatially efficient scanpaths even without explicit goals, establishing eye tracking as a powerful paradigm for studying the perceptual foundations of human spatial optimization abilities with increased ecological validity and temporal resolution compared to traditional manual reporting methods.

16.319 GAZE REVEALS HOW TASK DEMANDS SHAPE LEARNING FROM DEMONSTRATIONS

Inga Ibs¹, Ingolf Tegtmeier¹, Constantin A. Rothkopf¹; ¹Technical University of Darmstadt

In natural, everyday tasks, perception, cognition, and action are tightly intertwined. For example, when learning how to act from demonstrations, people need to select relevant information with their gaze based on their internal beliefs, reason about the demonstrations, and plan their subsequent actions. Here, we investigated how participants select information with their gaze in a controlled experiment that uses a navigation game to manipulate the task difficulty and informativeness of visual demonstrations. Participants (n=20) were shown trajectories of a fictional robot moving through environments with different terrains. They were instructed to infer the robot's movement costs associated with the different terrains and then guide the robot through a new environment. Task difficulty was manipulated by varying the number of terrains, and the informativeness of demonstrations was manipulated by how well an

ideal-observer model could infer movement costs from the provided demonstrations. Analysis of the gaze data reflected a two-phase strategy. Participants first gathered information, and then executed their plan at the end of each trial. Furthermore, attention was directed more toward the most informative parts of the demonstrations, as quantified by the ideal observer model. Gaze allocation was sensitive to the two task manipulations: In complex environments, participants switched more frequently between demonstrations and the environment they had to navigate, with shorter fixation sequences allocated to each surface, whereas in less complex environments, they switched less often with longer sequences. Uninformative demonstrations led to extended switching, reflecting a more thorough search for information. In contrast, informative demonstrations led to slightly longer fixation sequences, particularly in lower-complexity environments. These results indicate that participants adapted their visual attention strategies to support their inference and planning in response to task demands. Finally, using extensions of the ideal-observer model, we show how gaze patterns relate to participants' solutions to the navigation task.

This work was supported by the LOEWE research priority program "WhiteBox" [grant number LOEWE/ 2/13/519/03/06.001(0010)/77] (funded by the Hessian Ministry of Higher Education, Research, Science and the Arts).

16.320 EYE MOVEMENTS TRACK THE EMERGENCE AND CONTENT OF FLICKER-INDUCED VISUAL HALLUCINATIONS

Yueying Dong¹, Ana Chkhaidze¹, Lana Gaspariani¹, Sihan Yang¹, Anastasia Kiyonaga¹; ¹UC San Diego

The seemingly hidden content of the mind is sometimes expressed in observable eye movements. For instance, systematically biased saccade patterns can reveal the visual stimuli that are either imagined or maintained in working memory. However, it is unclear what underlying signal these saccadic signatures reflect, and how much information is contained in them. Here, we use induced visual hallucinations to probe the boundaries of the mental representations that modulate oculomotor activity. Specifically, exposure to bright and periodically flashing light is known to induce hallucinatory visual percepts – often simple geometric motifs and colored shapes. These hallucinations may provide a valuable window into the visual system, as the experience of more or less complex hallucinations is associated with engagement of distinct cortical substrates as well as individual variability in the reported capacity to generate mental images. Here, participants were exposed to a form of stroboscopic stimulation known as Ganzflicker, where a computer monitor alternated at 7.5 Hz between uniform black and red color fields. During the 10-minute exposure period, participants used a joystick to indicate the onset and offset of any hallucinations they experienced, and we tracked their eye movements continuously. After the exposure was complete, participants provided hand-drawn reports of each unique hallucinated percept, along with open-ended narrative recalls and a battery of self-

report measures. We examined both content-general eye movements metrics, like saccade frequency and magnitude, as well as content-specific biases, like the spatial patterning of fixations during hallucinations. We found that saccade frequency gradually ramped up as the onset of a hallucination approached, and also ramped down as hallucination intensity dwindled. Moreover, fixation hotspots during stimulation overlapped with the spatial layout of hallucinated content, as reported in the drawings. Therefore, distinct oculomotor markers track the temporal and spatial properties of experimentally-induced visual hallucinatory percepts.

16.321 THE EFFECTS OF SMART ROAD TECHNOLOGIES ON DRIVERS' STATIONARY GAZE ENTROPY: A DRIVING SIMULATION STUDY UNDER CLEAR AND ADVERSE WEATHER CONDITIONS

Carolina Diaz-Piedra¹, Francesco Angioi^{1,2}, Rocío de Oña², Juan de Oña², David Soler-Ortiz¹, Leandro L. Di Stasi^{1,3}; ¹Mind, Brain, and Behavior Research Center (CIMCYC), University of Granada, Spain, ²TRYSE Research Group, Department of Civil Engineering, University of Granada, Granada, Spain, ³Joint Center University of Granada - Spanish Army Training and Doctrine Command, Spain

Background: Adverse weather is a well-recognized major risk factor for road safety, as it substantially reduces visibility and predictability of road geometry. Conventional road technologies, such as white markings or passive road studs (known as "cat's eyes"), offer limited improvements in visibility. In contrast, smart road technologies (SRT) - such as photoluminescent markings (PRM) and active studs (ARS) - may enhance drivers' situational awareness by increasing luminance contrast and highlighting potential hazardous situations. However, it remains unclear whether these technologies shape drivers' visual exploration strategies and, in turn, affect driving performance. Here, we investigated how different technologies affect stationary gaze entropy during nighttime driving under both clear and adverse weather conditions. Methods: Fifty participants (mean age $\pm$ SD = 36.88 $\pm$ 11.28 years; 23 women) underwent two driving sessions (dry and rainy) each consisting of four driving scenarios (~ 15 min each, one for each technology: white markings, passive road studs, ARS, and PRM). We recorded stationary gaze entropy using the Tobii Glasses 2.0 (50 Hz), driving performance (speed and lateral position), and subjective ratings of complexity with a visual analog scale. Results: Drivers' stationary gaze entropy changed depending on the road technology of the scenario. Such changes were similar in both weather conditions. Post hoc comparisons showed that ARS led to less random (i.e., more stereotyped) visual exploration than classical markings. PRMs did not have such an effect. Finally, while perceived complexity was similar across all scenarios, driving performance improved when driving with PRM. Conclusions: Results show that ARS are valuable visual guidance tools for enhancing driving in poor visibility conditions by increasing drivers' situational awareness. This study supports the introduction of smart technologies as complementary passive

countermeasures to promote better visual strategies, even in adverse weather conditions.

This work is part of the Research Project SARAH (PID2022-141935OB-I00; MICIU/AEI/10.13039/501100011033 & ERDF/UE). FA is supported by a Spanish national pre-doctoral FPI research fellowship (PREP2022-000307; MICIU/AEI/10.13039/501100011033 & ESF Investing in your future).

Eye Movements: Pursuit, vergence

FRIDAY, MAY 15, 3:45 – 6:00 PM, BANYAN BREEZEWAY

16.322 FOLLOWING AN INVISIBLE TARGET: OCULOMOTOR CONTROL OF A BLIND EYE

Stephen Heinen¹ (heinen@ski.org), Devashish Singh¹, Scott Watamaniuk^{1,2}; ¹Smith-Kettlewell Eye Research Institute, ²Wright State University

We have previously shown in normals that an occluded eye moves unpredictably during midline pursuit (Chandna et al., 2021) but moves with the viewing eye during tangent screen pursuit (Heinen et al., 2024). These results suggest that the eyes are controlled independently for vergence, while a unitary signal drives conjugate movements. Here we test the contributions of independent and unitary signals to the pursuit signal by recording from a patient blind in one eye. This unique case allows us to study pure monocular viewing not present when occluding an eye. The patient (57 YO female) had late-onset blindness in one eye due to retinal ischemia and corrected-to-normal acuity in the seeing eye. For midline pursuit, an accommodative stimulus, a central letter "X" (0.34 deg at 67 cm) was surrounded by a peripheral letter array and moved back and forth periodically. For tangent screen pursuit, a dot (0.34 deg at 57 cm) moved left and right across the screen periodically. Eye movements were measured with an EyeLink 1000. We found that the seeing eye followed the target accurately in both paradigms. However, while the blind eye exhibited vergence during midline pursuit, its amplitude and temporal relationship with the seeing eye were highly variable. In contrast, the blind eye moved accurately and was temporally synchronized with the seeing eye during tangent-screen pursuit. The results provide additional evidence that while a single command does not drive the eyes during vergence, a single command does appear to drive them during conjugate eye movements. We believe the blind eye is controlled without vision by relying on an imperfect cortical representation of the target (a "virtual target") derived from input from both eyes. The results have implications for oculomotor neural pathways, and clinical strabismus interventions.

Supported by NIH R01EY034626, Smith-Kettlewell Eye Research Institute and Wright State University

16.323 NEURONAL REPRESENTATIONS OF SACCADDES AND PURSUIT IN THE PRIMATE FRONTAL EYE FIELDS

Kendra K. Noneman^{1,2}, Emily Lopez^{1,2}, Matthew A. Smith¹, J. Patrick Mayo²; ¹Carnegie Mellon University, ²University of Pittsburgh

Natural visual behavior depends on transforming incoming visual information into coordinated eye movement commands. To track unpredictably moving objects, primates alternate between fast, discrete saccades and slower, continuous smooth pursuit—underscoring the tight coupling between these two sensorimotor systems to support seamless gaze control. However, these eye movement sub-types have largely been studied independently and are often thought to be driven by distinct regions of the frontal eye fields (FEF). FEF contains saccade-responsive neurons that respond to visual stimuli or the (saccadic) movement. The distinction between sensory and motor responses for pursuit-responsive neurons is less clear because pursuit depends on the rapid interaction of both to continuously track a moving visual target. In this study, we examined how saccades and pursuit are represented at the level of single neurons and neural populations in the macaque FEF by simultaneously recording activity from neurons with diverse response properties. We classified neurons according to their responsiveness along: 1) a saccade-pursuit continuum, and 2) a more conventional visual-motor (saccadic) continuum during a memory-guided saccade task. Most FEF neurons responded to both saccades and pursuit rather than exclusively to one type of eye movement. At the population level, correlated variability was higher for neuron pairs with similar saccade-pursuit or visual-motor selectivity compared with pairs that were more dissimilar. Our findings show that saccade and pursuit representations overlap more than previously thought, providing new insight into how FEF coordinates flexible visuomotor transformations.

Whitehall Foundation, R01EY035673, P30EY008098, T32EY017271

16.324 MICROFLUCTUATIONS IN OCULAR ACCOMMODATION ACROSS TARGET DISTANCE AND LIGHT SPECTRUM

Nitin Negi¹, Benjamin Chin¹; ¹Rochester Institute of Technology

The human visual system uses a process known as accommodation to achieve clear images on the retina. Accommodation dynamically changes the shape of the crystalline lens, thus adjusting the eye's refractive power. Notably, accommodation is not stable. Even under steady viewing conditions, the visual system exhibits temporal variations in accommodation known as microfluctuations. A longstanding hypothesis in vision science is that low-frequency components of microfluctuations generate odd-error feedback signals that assist with accommodation. We further hypothesize that the visual system may increase its microfluctuations under conditions when accommodation is more difficult, such as when light entering the eye is narrowband. We tested this hypothesis with an existing dataset (Chin et al., 2025) of continuous accommodation measurements in which participants viewed small ($\sim 1^\circ$), foveally presented words for 3

seconds. The light spectra of stimuli were varied by mixing pairs of narrowband display primaries with varying ratios of the red and blue primaries. The green primary was either absent or fixed to half the luminance of the other two primaries. Each subject completed 216 trials corresponding to 12 unique spectra, 3 target distances (1.5, 2.5 and 3.5 D), and 6 trials per combination of spectrum and distance. Accommodative state was quantified as defocus aberration in the eye's wavefront, which was recorded at 30 Hz with a Shack-Hartmann wavefront sensor. We calculated total power in the low frequency components (0.1-0.6 Hz) on each trial via a power spectrum analysis of the time series, then averaged across all trials and participants for each combination of spectrum and accommodative distance. A two-factor repeated measures ANOVA indicated a significant effect of target distance on microfluctuations ($F(2,180)=4.99$, $p<0.01$), but not of color ($F(9,180)=0.29$, $p=0.98$). There was no interaction between the two factors ($F(18,180)=0.2$, $p=1.0$). Future work will explore the potential impact of other factors on microfluctuations.

Startup funds from the Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology

16.325 MONOCULAR DRIFT IN THE COVER TEST IS DRIVEN BY BINOCULAR COMMANDS CONSISTENT WITH HERING'S LAW

Jorge Otero-Millan¹, Michelle Kelly Phan, Reem Almagati; ¹University of California, Berkeley, ²University of California, Berkeley, ³University of California, Berkeley

The cover test is a clinical method used to measure heterophoria. It measures how much the eyes deviate when binocular fusion is interrupted, typically by covering one eye. During the test, the uncovered eye remains aligned with the target while the covered eye drifts towards the patient's phoria. This monocular drift might appear to contradict Hering's law, which states that eye movement commands are inherently binocular. Here, we asked what happens when not only binocular disparity cues are removed, but motion cues are also eliminated by flashing or strobing the target to the viewing eye. Using virtual reality, we simulated a cover test and recorded binocular eye movements in eight subjects under two conditions: a monocular continuous target, and a monocular target flashed for 50 ms once per second. Our data show that, once motion cues are removed, the underlying drift is in fact binocular: both eyes drift in opposite directions, and conjugate saccades return the "uncovered" eye to the target whenever it flashes. Furthermore, the drift velocity of the covered eye was larger in the continuous-target condition than in the flashing condition (2.1 ± 0.4 times), consistent with a conjugate command that uses motion feedback to stabilize the viewing eye and, consequently, is also applied to the covered eye increasing its drift velocity. These results demonstrate that eye movements during the cover test remain fully consistent with Hering's law and highlight the critical role of motion signals in maintaining monocular fixation and counteracting the drifts driven by heterophoria.

16.326 THE CONTRIBUTION OF DISJUNCTIVE SACCADDES TO VERGENCE RESPONSES OF ADULTS AND CHILDREN

Bhagya L Marella¹ (bhagya.opt@gmail.com), Stephanie A Biehn¹, T Rowan Candy¹; ¹Indiana University School of Optometry, Indiana University, IN, USA

Purpose: Humans experience asymmetric vergence demands during natural viewing and must calibrate responses during development. The roles of disjunctive saccades and slower vergence in habitual responses are not well understood in typical and atypical development. This study determined contributions of the two movements in 7-yr-olds and adults for different stimulus ratios. **Methods:** A 16deg radius spatially 1/f dynamic (120z) was presented dichoptically using a PROPixx projector. A 2deg fixation target was presented monocularly to the left eye. Step changes in disparity (1.5s) were introduced asymmetrically to stimulate disjunctive saccades and vergence (combinations of +3 to -3deg symmetric lateral and depth). An EyeLink 1000 (SR Research) made binocular recordings. Vergence and saccade responses were identified using velocity criteria and assigned to pre-saccade, intra-saccade and post-saccade periods. **Results:** For both children (7yrs) and adults, much of the vergence response amplitude was driven by post-saccadic vergence (median (SE): 40% (2.0) & 49% (1.8)), in children and adults, respectively), followed by pre-saccadic vergence (median (SE): 36% (2.1) & 25% (1.6)) and then intra-saccadic vergence. Duration of the disjunctive saccadic component (median: 22ms in both children and adults) was shorter than slower vergence during the pre- and post-saccadic periods. Median (SE) latency of the disjunctive saccade was longer in children 429ms (18ms) than adults 324ms (14ms) (lmer: $p < 0.001$). **Conclusions:** Seven-year-olds produced comparable asymmetric vergence responses to adults for disparity only stimuli (except for the latency of the disjunctive saccade), consistent with a study of larger amplitude full cue stimuli (Yang et al, 2002). Our data suggest that both disjunctive saccades and slower vergence responses can be driven by disparity alone by 7 years. Given the separate neural pathways driving these components, the role of disjunctive saccades in human developmental disorders of eye alignment needs further investigation (Walton et al, 2017; Das, 2016).

16.327 THE ROLE OF DISPARITY AND VELOCITY DIFFERENCES IN DRIVING VERGENCE RESPONSES OF ADULTS

Fuseina Safianu¹, Bhagya L Marella¹, Stephanie A Biehn¹, Aishwarya Ravi¹, T Rowan Candy¹; ¹Indiana University

Objective Vergence eye movements are employed to track objects in depth. Change in disparity over time (CDOT) and intraocular velocity differences (IOVD) serve as cues to motion in depth (Giesel et al, 2019). Two studies in our lab have demonstrated vergence responses to (i) disparity alone and (ii) both cues. This study directly assessed the role of CDOT and IOVD in tracking vergence responses of adults.

Methods A 10° diameter, spatially 1/f noise texture was presented dichoptically against a mean luminance background at 100cm viewing distance using a PROPixx projector (RB3D mode, VPixx). This stimulus isolates the CDOT cue by dynamically changing both disparity (0.1Hz, 2° amplitude sinusoid) and texture at 120Hz over time. The amplitude of 30s of sinusoidal vergence response was compared to that to a 10° diameter Maltese cross, containing both CDOT and IOVD cues undergoing the same disparity temporal profile (combined cue stimulus). Data were recorded from 10 functionally emmetropic adults using an EyeLink 1000 Plus (SR Research). Results Fast Fourier transforms were performed on the horizontal vergence movements. Group mean vergence amplitude and latency at 0.1Hz for the CDOT stimulus were 1.86° (SD:±0.46) and 209ms (SD:±156), and for the combined cue stimulus were 1.93° (SD:±0.32) and 152ms (SD:±95), closely matching the amplitude of the sinusoidal stimulus. Increasing the stimulus amplitude to 4° revealed asymmetries between convergence and divergence responses in the CDOT condition for 3 participants while their combined cue responses remained symmetric. The asymmetries were consistent with their phoria and fixation disparity measurements. **Conclusion** Adults generated robust tracking vergence responses to these textures with naturalistic image statistics. The larger amplitude CDOT stimulus revealed the asymmetric nature of these vergence responses in some adults. Hence, studying infants' responses to these stimuli holds diagnostic potential for predicting binocular motor function during development.

NEI RO1 EY014460

16.328 BRIEF MONOCULAR STIMULUS FLASHES ARE SUFFICIENT TO SUSTAIN VERGENCE IN VIRTUAL REALITY

Reem Almagati¹ (ralmagati@berkeley.edu), Michelle Kelly Pham¹, Jorge Otero-Millan^{1,2}; ¹University of California, Berkeley, ²Johns Hopkins University

Heterophoria refers to the angular difference between ocular vergence and vergence demand of a given target when disparity signals are absent. A common clinical method of measuring heterophoria is the alternating cover test. We utilized virtual reality (VR) to test heterophoria, in three different versions of a virtual cover test: monocular continuous target, (dark) no target in either eye, and a monocular flashing target that was on only 50 milliseconds every 1 second. Eight subjects viewed stimuli using the Varjo XR-4 VR headset while we tracked their eye movements using Openlris. Participants were seated with their chin rested on a chinrest to minimize head movements. Monocular calibration was performed prior to each session. We tested the relaxation of vergence after a 5 second binocular period of forced convergence or divergence. We measured the average vergence angle reached between 8 and 10 seconds after the cessation of the binocular period. Our results show a slightly convergent (5.3 ± 2.2 degrees) vergence angle in the no target

condition (tonic vergence) consistent with findings in the literature. The vergence angle in the continuous target condition was 1.7 ± 1.9 degrees. Interestingly, the vergence angle in the flashing condition remained comparable to the non-flashing condition (2.4 ± 1.7 degrees) and far from the value of the no target condition. This suggests that the crosslinks between accommodation and vergence are sufficient to hold vergence during very brief stimuli that may not be enough to drive accommodation alone.

NEI training grant 5T32EY007043-43 NEI R00EY027846 Hellman Fellows Fund

Face and Body Perception: Social cognition 1

FRIDAY, MAY 15, 3:45 – 6:00 PM, BANYAN BREEZEWAY

16.329 PRETTY FACES LOOK LIKE US: IN-GROUP PROJECTION BIASES IN CATEGORIZATION OF EAST ASIAN FACES

Cansu Malak¹, Christian Wallraven¹; ¹Korea University

When we perceive a face, we rapidly judge both ethnicity and social traits like attractiveness and trustworthiness. But how do these two processes interact - especially for visually similar groups? We investigated whether judgments of attractiveness and trustworthiness enhance or distort the categorization of East Asian (Korean, Japanese, Chinese) faces by Korean observers. For this, we presented 600 full-face images of Korean, Japanese, and Chinese males (200ms exposure) to 104 Korean participants. The participants were assigned to two tasks: one group (N=52) rated each face for attractiveness and trustworthiness, while a second group (N=52) categorized the ethnicity of the same faces. We then correlated the social ratings from the first group with the categorization accuracy from the second. Results revealed a strong in-group bias: using the dataset labels, Korean faces were rated as significantly more attractive and trustworthy than Japanese or Chinese faces: for in-group Korean faces, higher attractiveness ratings ($r = .62$) and trustworthiness ratings ($r = .65$) strongly predicted greater categorization accuracy. For out-group Chinese faces, the opposite pattern emerged: higher ratings predicted lower accuracy (attractiveness $r = -.44$; trustworthiness $r = -.35$), revealing a misattribution bias where socially positive faces were incorrectly classified as Korean. No significant link was found for Japanese faces. Across all stimuli, more attractive or trustworthy faces were also more likely to be labeled as Korean, reflected in positive correlations with “Korean” responses (attractiveness $r = .61$; trustworthiness $r = .65$) and negative correlations with “Japanese” ($r = -.23$; $-.37$) and “Chinese” classifications ($r = -.45$; $-.32$). Overall, rapid social evaluations are not neutral; they are deeply intertwined with ethnic categorization. Our results are compatible with a strong in-group

projection bias, where socially positive faces are claimed as “one of us,” demonstrating that face categorization is a social-affective process, not just a visual one.

Supported by the National Research Foundation of Korea (BK21 FOUR, RS-2025-00555141), by IITP grants funded by the Korea government (No. RS-2019-II190079, Department of Artificial Intelligence, Korea University; No. RS-2021-II212068, Artificial Intelligence Innovation Hub).

16.330 INDEPENDENT CONTROL OF RACE AND IDENTITY INFORMATION REVEALS THE HIDDEN STRUCTURE OF OTHER-RACE BIASES

Fabian Soto¹, Emily Martin¹; ¹Florida International University, Miami, FL

Race influences both recognition and categorization of faces, yet most prior work relies on stimulus sets in which race and identity covary, preventing clear causal inferences about which facial properties drive these effects. We developed a controlled stimulus framework enabling independent manipulation of race and identity, allowing direct tests of how race-specific morphology shapes perception. Using anthropometrically validated 3D models of Caucasian, Asian, and African facial structure—together with an “alien” race matched in variability but unfamiliar to observers—we generated stimuli in which identity information was held constant across races while only race-specific shape varied. In Study 1, participants completed both a recognition-memory task assessing the other-race effect (ORE) and a race-categorization task assessing the other-race categorization advantage (ORCA). Because identity was matched across races, any differences reflect race morphology alone. Under these controlled conditions, we observed no ORE, but the ORCA remained robust. In Study 2, reverse correlation was used to test whether the features supporting identification of a familiar face differed across races. Despite clear visual differences among races, observers relied on the same identity-diagnostic information across race versions of the face, consistent with the absence of an ORE in Study 1. In Study 3, discrimination thresholds were collected across key regions of face space to compare population-encoding models. Encoding of familiar identities remained stable across races, but own-race faces structurally similar to the other race were easier to distinguish from the other race and harder to distinguish from nearby own-race identities. A model incorporating tuning shifts toward the other-race manifold best explained these effects. Together, these studies show that when race is isolated from identity, the ORCA persists while the ORE disappears, and representational changes are localized near the other race. These results suggest that identification and categorization biases arise from distinct underlying sources of information.

This research was funded by NSF grant 2319234 awarded to Fabian Soto.

16.331 VIRTUAL RACE EMBODIMENT MODULATES FACE RACE PERCEPTION

Pauline Schaller¹ (pauline.schaller@unifr.ch), Steve Katembu², Sotaro Taniguchi³, Junhao Jiang³, Akira Sarodo³, Charles Obadiah Kimamo², Katsumi Watanabe³, Anne-Raphaëlle Richoz¹, Roberto Caldara¹; ¹Eye and Brain Mapping Laboratory (iBMLab), Department of Psychology, University of Fribourg, Fribourg, Switzerland, ²Department of Psychology, University of Nairobi, Nairobi, Republic of Kenya, ³Faculty of Science and Engineering, Waseda University, Tokyo, Japan

Understanding how social identity influences face perception remains a fundamental challenge in vision science. Prior research reveals consistent perceptual advantages for categorizing other-race faces, the so-called Other-Race Categorization Advantage (ORCA). However, the mechanisms underlying this effect remain contested, particularly regarding the interplay between socio-cognitive factors and perceptual expertise. Virtual Reality (VR) technology offers a novel approach by temporarily manipulating perceived racial identity through avatar embodiment. We conducted a large-scale, cross-cultural study involving over 250 participants from Kenya, Japan, and Switzerland. Participants completed two face race categorization tasks: one using 100% race-typical faces and another using morphed stimuli varying continuously along a race composition continuum. Participants underwent embodiment in either same-race (SR) or other-race (OR) avatars matched to their sex, engaging in three interactive embodiment-enhancing activities involving visuo-auditory and visuo-tactile feedback, in a VR environment replicating the real laboratory. Following embodiment, participants repeated the categorization tasks with new face stimuli. Our data showed that OR embodiment did not reduce the ORCA in coarse categorization, suggesting the speed advantage for processing OR faces remains resilient to transient social identity shifts. However, OR embodiment reliably modulated fine-grained race perception: participants embodied in OR avatars were significantly more likely to classify ambiguous morphed faces as SR post-embodiment compared to pre-embodiment, indexed by a substantial positive shift in the Point of Subjective Equality relative to SR-embodied controls. These findings demonstrate that shifts in perceived social identity, can modulate perceptual categorization. These novel insights into the mechanisms underlying face race perception across diverse populations validate VR embodiment as a powerful tool for experimentally manipulating social affiliation and open new theoretical and methodological pathways for understanding the psychological bases of perception. By enabling controlled changes in perceived group membership, VR provides a promising foundation for intervention studies designed to overcome other-race effects and reshape face processing biases.

This work was supported by a grant from the Swiss National Science Foundation [n°100019_189018] awarded to Roberto Caldara.

16.332 TRAIT IMPRESSIONS SHAPE SOCIAL CATEGORIZATION: EVIDENCE FROM THE ISRAELI FACE DATABASE

Maayan Trzewik^{1,2} (maaiantk@gmail.com), Bat Sheva Hadad¹, Mayan Navon³, Tal Moran⁴, Hadas Wardi², Adi Langer², Carmel Sofer², Niv Reggev²; ¹University of Haifa, ²Ben Gurion University, ³UCLouvain, ⁴The Open University of Israel

Much research demonstrates that people tend to form more negative impressions of outgroup faces and more positive impressions of ingroup faces. Here we ask whether this motivated preference for the ingroup also operates in the opposite direction—namely, whether trait impressions themselves guide bottom-up social categorization, leading perceivers to classify faces as ingroup or outgroup members based on the positivity or negativity of the impressions they elicit. We tested this question using the Israeli Face Database (IFD), a novel database representing individuals from different social groups residing in Israel, including individuals of Jewish and Arab ancestries. A total of 1,336 Jewish participants residing in Israel and 1,199 American participants from diverse ethnic backgrounds rated IFD faces on a range of positive traits (e.g., trustworthiness, friendliness, intelligence) and negative traits (e.g., meanness, weirdness). In addition, Israeli participants rated each face's likelihood of belonging to their ingroup (Jewish Israelis) or outgroup (Arab Israelis). Importantly, participants were not informed of the faces' true group identities. Results show that faces rated higher on positive traits and lower on negative traits across both Israeli and American samples were more likely to be categorized by Israeli participants as belonging to their ingroup. Conversely, faces rated lower on positive traits and higher on negative traits by both samples were more likely to be categorized as outgroup members. Thus, social categorization judgments aligned with trait impressions independently formed by American observers, who were unaware of the ingroup–outgroup distinction and of any social categorization. These findings suggest that motivation to maintain a positive view of the ingroup and a negative view of the outgroup does not only bias evaluations of known group members but also shapes group categorization based on appearance. More broadly, the results highlight a bidirectional relationship between person perception and social motivation.

16.333 HYPER-REALISTIC REVERSE CORRELATION OF MENTAL ILLNESS REPRESENTATIONS: A ROLE FOR STIGMATIZING BELIEFS?

Siddharth Munupatrolu¹ (munupat2@illinois.edu), Daniel N. Albohn², Ignazio Ziano³, Yasin Koc⁴, Stefan Uddenberg¹; ¹University of Illinois Urbana-Champaign, ²University of Chicago Booth School of Business, ³University of Geneva, ⁴University of Groningen

Mental illness is stigmatized, leading to discrimination and social exclusion. Are there shared visual stereotypes of mental illness amongst the general population? If so, what is the visual content of such stereotypes? And are individual representations of mental illness influenced by the level of stigmatizing beliefs the perceiver might hold? Here we explored these issues using hyper-realistic generative reverse correlation (Albohn et al., 2022; 2025). Participants (N=420) read a description of a prevalent mental illness (i.e., bipolar disorder,

general anxiety disorder [GAD], major depressive disorder [MDD], or post-traumatic stress disorder [PTSD]) and judged whether synthetic faces (300 faces, presented serially) looked like they "had" or "did not have" the described condition (they could also respond "not sure" if uncertain about their choice). We extracted mental representations of these conditions by taking the difference between the averaged faces judged as "having" versus "not having" the condition, yielding a vector in face space distinguishing the two categories (Albohn et al., 2022; 2025). These representations were then compared to our pre-existing models of social attribute impressions (Peterson et al., 2022) using signed cosine similarity (s) where $s > 0$ indicates alignment and $s < 0$ anti-alignment with the attribute vector. Across all conditions, illness representations were visually interpretable and reliably anti-aligned with perceived health, trustworthiness, and attractiveness, while being reliably aligned with perceived masculinity (all conditions). In addition, some conditions were aligned with perceived age (GAD/PTSD), dominance (PTSD only) and darker skin tone (PTSD only). Contrary to our expectations, participants' Mental Illness Stigma Scale (Day et al., 2007) scores were not systematically associated with visual differences in the extracted representations of the illnesses, suggesting that one's visual stereotypes of mental illness may not be driven by stigmatizing beliefs.

16.334 SIMPLE 3D POSE FEATURES SUPPORT HUMAN AND MACHINE SOCIAL SCENE UNDERSTANDING

Wenshuo Qin¹ (wqin6@jh.edu), Leyla Isik¹; ¹Johns Hopkins University

Humans can quickly and effortlessly extract a variety of information about others' social interactions from visual input, ranging from visuospatial cues like whether two people are facing each other to higher-level information like whether people are communicating. Yet, the computations supporting these abilities remain poorly understood, and social interaction recognition continues to challenge even the most advanced AI vision systems. Here, we hypothesized that humans rely on 3D visuospatial pose information to make social interaction judgments, and this information is absent from most AI vision models. To test this, we designed a novel pipeline that leverages state-of-the-art pose and depth estimation to automatically extract 3D body joints from short video clips depicting everyday human actions. We compared the performance of these body joints against over 350 current AI vision models in predicting human social judgments on these videos. Strikingly, 3D body joints outperformed most current AI vision models, revealing that key social information is available in explicit and interpretable body joints but not in the high-dimensional learned features of most vision models. We next reduced the 3D body joints to an even more compact set that captured only the 3D positions and directions of people in the videos. We found that this minimal set of 3D features (but not their 2D counterparts) was both necessary and sufficient to explain the prediction performance of the full set of 3D body joints. Moreover, this minimal 3D set also predicted the extent to which AI models aligned with human social judgments and significantly improved their performance on these tasks. Together, our findings

provide strong evidence that human social scene understanding relies on explicit 3D pose representations and can be supported by simple, structured visuospatial primitives.

16.335 GENDER BIAS IN VISUAL CATEGORIZATION: WHEN DICAPRIO IS AN ACTOR BUT JOLIE IS A WOMAN

Rui Zhe Goh¹ (rgoh1@jhu.edu), Jiayi Li², Ian Phillips¹, Chaz Firestone¹; ¹Johns Hopkins University, ²Cambridge University

Objects can be categorized at differing levels of abstraction; for example, the same image may be classified as a blue jay, a bird, an animal, or a living thing. This pattern also arises for people; for example, the same image may be classified as Gordon Ramsay, a chef, a Scot, or a person. But which categories are most visually salient—and is such salience biased by gender? We investigated this question using a visual categorization task. Subjects saw images of objects and people, and were asked simply to name any category corresponding to the image. In addition to objects such as vegetables and buildings, each subject saw one woman and one man, drawn from one of the following classes: 1) famous people from various professions (e.g., Angelina Jolie and Leonardo DiCaprio, Serena Williams and LeBron James); 2) stock photographs of anonymous people from minimally-gendered categories (e.g., tourist, pedestrian); 3) AI-generated images of these same minimally-gendered categories, created to be identical except for gender. Every image was part of a matched pair of men and women from the same image class, but no subject saw a man and woman from the same matched pair. (For example, a subject could see Angelina Jolie and LeBron James, but never Angelina Jolie and Leonardo DiCaprio.) Remarkably, subjects were more likely to categorize women according to their gender than men. Put differently, they were biased to say "actor" for DiCaprio but "woman" for Jolie, and "pedestrian" for male pedestrians but "woman" for female pedestrians. This trend arose for all three image classes, suggesting that it was not driven by idiosyncratic image differences or knowledge of the people depicted. Our results reveal a gender bias in visual categorization that may have pervasive cognitive and social implications.

16.336 BEHAVIOR-GUIDED FINE-TUNING MAKES VISION MODELS MORE HUMAN-LIKE IN SOCIAL PERCEPTION

Kathy Garcia¹ (kgarci18@jh.edu), Leyla Isik¹; ¹Johns Hopkins University

Humans effortlessly perceive rich social information in visual scenes, yet it remains unclear whether current AI vision models organize these scenes the way people do. Matching this structure is important for using models as scientific accounts of human vision and for building systems that understand social scenes. Here, we introduce a new benchmark where we measured the similarity structure of human social perception of over 49,000 odd-one-out judgments on 250 three-second videos of everyday actions, defining which videos people treat alike. We then ask whether pretrained vision models capture this structure, and find that vision models struggle: their internal

embeddings do a poor job of predicting which clips people judge as the odd-one-out. To address this, we design a novel behavior-guided, hybrid-triplet-RSA objective, inspired by representational similarity analysis: videos judged as similar by humans are pulled together in the model's space and dissimilar ones are pushed apart. We then use this novel objective to fine-tune vision models (e.g. CLIP, VideoMae, Timesformer) directly on human judgments. Notably, this behavior-guided fine-tuning significantly increased model-human alignment, improving explained variance and odd-one-out triplet accuracy on unseen videos. Interestingly, qualitative analysis of attention rollouts show that while baseline vision models focus on low-level and background information, our novel objective encourages attention on participants and actions in social scenes. Finally, probing analysis demonstrates that our fine-tuned models show stronger linear readouts of social-affective attributes (e.g. intimacy, valence, communication) relative to pretrained baselines. Overall, our findings reveal a significant gap in pretrained AI vision models' ability to match video judgments and provide compelling evidence that human behavioral supervision can close this gap, making models organize social scenes more like humans do. Our benchmark and behavior-guided objective define a data-efficient, model-agnostic framework for building and rigorously evaluating biologically informed models of social vision.

This work was funded in part by NSF GRFP DGE-2139757 and NIMH R01MH132826.

16.337 MID-LEVEL MOTION STATISTICS AS A MODEL FOR SOCIAL PERCEPTION IN THE LATERAL STREAM

Ming Zhou¹ (mzhou52@jhu.edu), Leyla Isik¹; ¹Johns Hopkins University

The lateral visual stream supports social interaction perception, while there are few computational models to characterize the neural representations underlying this process. Although existing deep learning models can predict the neural responses in these regions, they lack interpretability and are computationally intensive. To address these gaps, we present an image-computable model, derived from first principles, to characterize social motion responses in higher-level regions across the lateral stream. We derive second-order features from the Portilla-Simoncelli texture model to construct mid-level motion statistics. Specifically, we take a standard motion energy model, consisting of spatiotemporal Gabors with different directions of motion, spatial frequencies and temporal frequencies as our first-order features. Analogous to the P-S model, we set up three kinds of second-order features based on correlation of distinct first-order features: the spatial-autocorrelation, cross-direction and cross-temporal-frequency respectively computed as the correlation across different spatial locations, directions and temporal frequencies. Visual inspection reveals that these features capture meaningful visual patterns in the stimulus, including periodicity, curved contours and biological motion. With these feature sets, we constructed encoding models to predict the human behavioral ratings and neural responses

in an fMRI dataset of 200 video clips involving two people engaged in everyday actions. We found that compared to first-order motion energy features, second-order motion features (cross-direction and cross-temporal-frequency) better predicted behavioral judgments of communicative interactions, and critically, explained significantly more variance in mid-level lateral regions (e.g., EBA) and high-level social interaction regions (e.g., pSTS). These findings suggest that second-order motion features can capture higher-level motion information, analogous to their static image counterparts, and may support social perception along the lateral stream. These hand-engineered mid-level motion statistics are also a promising, interpretable alternative to deep learning for understanding the neural representations of social interaction perception.

16.338 BOTH VISUAL AND THEORY-OF-MIND COMPUTATIONS EXPLAIN UNIQUE VARIANCE IN THE SOCIAL BRAIN

Manasi Malik¹ (mmalik16@jhu.edu), Minjae Kim¹, Tianmin Shu¹, Shari Liu¹, Leyla Isik¹; ¹Johns Hopkins University

Making social evaluations from visual input is a core human ability, yet the computational mechanisms that enable this remain unknown. Visual social recognition consistently engages regions involved in social perception, including regions along the superior temporal sulcus (STS), as well as higher-level theory-of-mind (ToM) areas such as the temporoparietal junction (TPJ). One common hypothesis is that these regions operate hierarchically, with social-perceptual regions carrying out bottom-up visual computations that serve as input to higher-level inference in the ToM network. This hypothesis, however, has never been computationally tested, in large part due to the lack of successful models of social processing. We recently developed computational models of both social interaction perception and theory-of-mind: a graph-neural-network (GNN) model that relies on relational visual information, and a generative inverse-planning model that recognizes social interactions by inverting a model of agents' goals and the physical world. In this preregistered study, we collected fMRI responses while participants ($n = 25$) watched 10s videos depicting social interactions and compared neural responses to both computational models using representational similarity analysis (RSA) in functional regions of interest and whole-brain searchlights. We found that both the GNN and the inverse-planning model significantly explained variance in both social-perceptual (e.g., pSTS) and ToM (e.g., TPJ) regions, even after controlling for variance explained by the other model. This suggests that both regions support a combination of relational visual processing and higher-level inferential reasoning. Exploratory analyses suggested differences in the time course of these different computations, indicating a shift from early visual processing towards later inferential evaluation in both regions. Overall, these findings suggest that both social-perceptual and ToM regions carry out a combination of relational visual and higher-level inferential computations, perhaps on distinct timescales with early relational visual processing followed by later inverse-planning-based inference.

16.339 CATEGORIZATION OF FACES IS DIFFERENT IN BINARY AND TERNARY TASKS

Émilie Désaulniers¹, Marie-Pier Plouffe-Demers^{1,2}, Marie-Claude Desjardins¹, Justin Duncan^{1,3}, Daniel Fiset¹, Caroline Blais¹;
¹Université du Québec en Outaouais, ²Université du Québec à Montréal, ³University of Ottawa

The other-race categorization advantage (ORCA) is a well-established effect where individuals are faster and better at categorizing other-race than own-race faces. This effect is typically found in binary racial categorization tasks, where participants categorize faces of two races, usually their own and another. However, recent work raised concerns about the robustness of ORCA, showing that relatively minor changes in task parameters can substantially alter the observed pattern. In a previous ternary categorization task including Black, White and Asian faces, we tested Chinese, South African and Canadian participants ($n = 648$) and observed a robust reverse-ORCA, meaning a better categorization of own-race than other-race faces, especially in South African participants. This unexpected reversal suggested that task structure may critically shape whether an ORCA emerges, possibly by shifting the categorization strategies used. The aim of the present study was to test whether the ORCA would re-emerge when using the same stimuli but a different task structure. We used the original 216 morphed stimuli (ranging from 2 percent to 86 percent racial prototypicality of White, Black and Asian faces) in a binary categorization task, in which participants judged whether each face belonged to their own racial group or not. 216 participants (108 Canadians and 108 South Africans) completed the task and reported their contact with people of same and other racial groups. Perceptual sensitivity (d') were calculated and analyzed in mixed 2 (groups: Canadians and South Africans) $\times$ 2 (condition: same vs other race) ANOVAs, revealing a significant interaction. Canadians showed a clear ORCA, while no significant differences were found in South Africans after Bonferroni correction. No significant correlations emerged between the contact measure and d' . These results indicate that the classic ORCA depends on a binary task context: introducing a third racial category abolishes the effect and can even invert it.

Temporal Processing: Duration and timing perception

FRIDAY, MAY 15, 3:45 – 6:00 PM, BANYAN BREEZEWAY

16.340 AUTOMATIC BIASES IN VISUAL AND AUDITORY PERCEPTION

Linda Garami¹, József Fiser^{1,2}; ¹Central European University, ²Center for Cognitive Computation

The brain relies on various internally generated categories and stereotypes, formed through biased processing of inputs, to make rapid predictions. We investigated the hypothesis that applying such biases -identifying and focusing on particular groups and patterns that prioritize fast, goal-relevant efficiency over veridical accuracy- is not limited to higher-order cognitive processes. Instead, this strategy is applied automatically and recursively from the earliest levels of sensory processing. We focused on identifying basic chunking principles in the visual and auditory modalities. Behaviorally, chunking consistently leads to decreased accuracy in detecting perceptual changes at perceived chunk boundaries compared to changes within a segment. Using this measure in stimulus sequences with increasing internal structure across different modalities, we tested whether automatically applied temporal grouping principles are shared across vision and audition. We implemented a stream-segregation paradigm in which we manipulated the duration structure of the elements: 200- and 600-ms simple tones or basic visual objects were presented in a continuously repeating three-element pattern, such as Short-Short-Long or Short-Long-Long. Our results ($N = 35$) show that (1) participants exhibit similar sensitivity patterns across modalities: as soon as a repeating pattern emerges, sensitivity to environmental changes becomes biased by the same chunking principles. Boundaries in both modalities are marked by longer-duration sensory objects that function as closing elements; (2) this bias is automatic and often consciously unavailable to participants, even when attention is directed toward the stream; and (3) this pattern representation is highly resilient to absolute changes in element duration, as long as the relative duration supports a consistent repeating pattern. These results suggest that the perceptual system organizes sequences of elements with varying absolute durations into chains of tokens based on rapidly formed categories that remain remarkably stable even in noisy environments.

HORIZON-MSCA-2023-PF-01 No.101155302

16.341 TEMPORAL COHERENCE SHAPES BEHAVIORAL AND NEURAL REPRESENTATIONS OF ACTION

Ali Feza Karahaliloglu¹, Burcu Ayşen Ürgen¹; ¹Bilkent University

Temporal receptive windows (TRWs) describe the timescales over which sensory systems integrate dynamic information. Although TRWs are typically studied using long naturalistic stimuli, it remains unclear whether similarly graded perceptual signatures can be elicited with short, well-controlled stimuli, and how attention modulates these dynamics. We developed a new temporal-scrambling paradigm using 1.5-s action videos in which an actor dragged or rotated a spherical object. Each video was divided into equal-length clips whose order was permuted pseudorandomly to generate eight scrambling levels. In alternating blocks, participants attended either the action or a central fixation target. During action-attention blocks, participants judged whether the object moved toward or away from the actor. Twelve participants completed the task behaviorally. Accuracy increased and response times decreased as videos became more intact (Accuracy:

$F(7,77)=59.43, p<10^{-12}$; RT: $F(7,77)=5.68, p<10^{-3}$). A psychometric fit revealed a robust positive slope (mean=0.507 log-odds/level), explaining 90% of group-level accuracy variance. A small “toward” response bias diminished with temporal coherence, consistent with reduced uncertainty on intact trials. These results demonstrate that short, controlled stimuli can elicit clear TRW-like behavioral signatures. Three participants were scanned during 3T fMRI while performing the same task. Intact versus scrambled contrasts produced stronger responses in higher-level regions of the action observation network (AON), whereas the opposite contrast yielded clusters in early visual cortex. We assessed neural sensitivity to temporal coherence using permutation-based decoding of attend-to-action data. Decoding accuracy was reliably above chance across subjects, with the strongest effects in hMT (Cohen’s $d \approx 5.0$) and PMd ($d \approx 4.2$), and additional robust decoding in AIPS, IPL, and mid-level visual cortex (V2/V3). Early visual cortex (V1) showed no decodable structure, indicating that decoding reflects action-relevant temporal integration rather than low-level visual features. These findings introduce a compact paradigm for probing perceptual TRWs and attentional influences on action processing.

This project is funded by TÜBİTAK (The Scientific and Technological Research Council of Türkiye) - ARDEB (Academic Research Funding Program Directorate) within the scope of the “1001- Scientific and Technological Research Projects Funding Program” Project Number: 124K961

16.342 DIFFERENTIAL EFFECTS OF SELF-CONSTRUAL AND STIMULUS SPEED ON DURATION ESTIMATION AND REPRODUCTION

Yasmin Abuwi¹, Cadence Stolzer², Jenna Pedrick³, Gabriel Conn⁴, Katherine S. Moore⁵; ¹Arcadia university, ²Arcadia university, ³Arcadia University, ⁴Arcadia University, ⁵Arcadia University

Time perception can be distorted by multiple factors. For example, a fast-moving visual stimulus is perceived to last longer than a slow-moving stimulus of the same duration. Moreover, one’s self-construal can distort time perception, with independent (“I” focused) individuals overestimating intervals of moderate (60s) length more often than interdependent (“we” focused) individuals. Here, we examined both effects in a single experiment to examine whether self-construal distorts the magnitude of the speed illusion and to examine whether self-construal has differential effects depending on short (~2s) versus moderately long (~120s) intervals. Participants were first induced to adopt an independent (I) or interdependent (we) self-construal. Next, they performed a time reproduction task involving short (2s) durations paired with fast- or slow-moving stimuli (a visual gabor, or an auditory pulse). Participants also estimated long intervals at checkpoints occurring every 2-3 minutes. We replicated the speed illusion, with fast stimuli appearing to last longer than slow ones in both the visual and auditory domains. Surprisingly, participants were more accurate overall in reproducing intervals in the visual than the auditory task. Participants primed into independent self-construal overestimated

short intervals more than did participants primed into an interdependent self-construal, but self-construal did not interact with the magnitude of the speed effect. Moreover, contrary to prior findings, self-construal did not have an impact on bias (over/underestimation) nor precision in time estimates for longer judgments. These findings provide evidence for both personal and stimulus factors that contribute to distortions of time perception, but that self-construal may play a smaller role than previously thought.

16.343 DISTINCT VISUOSPATIAL TUNING OF TEMPORAL INTEGRATION AND SEGREGATION

Ryo Kanechi^{1,2} (ryo.kanechi@nict.go.jp), Neil W. Roach³, Masamichi J. Hayashi^{1,2}; ¹Center for Information and Neural Networks (CiNet), Advanced ICT Research Institute, National Institute of Information and Communications Technology, ²Graduate School of Frontier Biosciences, The University of Osaka, ³School of Psychology, University of Nottingham

Two stimuli separated by a short temporal interval are perceived as integrated, whereas those separated by a long interval are perceived as segregated. Temporal integration and segregation are thought to play important roles in perceptual stability and change detection, respectively. However, it remains unclear whether these processes are regulated by a single or distinct neural mechanism(s). Here, we characterize and compare the spatial tuning of sensitivity for temporal segregation and integration. We presented a pair of 4 x 4 multi-element panels sequentially, with a variable temporal interval ranging from 30 to 150 ms. Participants performed either an odd-element task (OET) or a missing-element task (MET) to measure temporal segregation and integration performance, respectively. Participants were asked to maintain gaze at a fixation point and identify the location where either an incomplete element appeared (OET) or where no element appeared across the panels (MET). To dissociate the retinal co-ordinates of the critical element from its position in the stimulus array, the fixation point either remained fixed at the center (Experiment 1) or varied between four quadrant locations on a trial-by-trial basis (Experiment 2). The results consistently showed that sensitivity (d') for both tasks declined as the retinal eccentricity of the critical element increased, but the drop-off was significantly more pronounced for the integration task. These results may reflect a greater dependence of temporal integration on the parvocellular pathway, within which cell density decreases steeply with eccentricity. In contrast, temporal segregation may be more dependent on the magnocellular pathway, which is characterized by more stable cell density across eccentricity. Overall, these findings indicate that temporal integration and segregation may be governed, at least in part, by distinct neural mechanisms.

16.344 DOES EMOTIONAL SALIENCE ALTER TEMPORAL PROCESSING IN PERI-HAND SPACE?

ANKIT MAURYA¹, TONY THOMAS¹; ¹Indian Institute of Technology Roorkee, Roorkee, India

Previous studies have shown temporal dilation for emotionally salient stimuli. Also, processing of such stimuli is enhanced in the graspable space of the hands, known as peri-hand space (PHS). However, the spatio-temporal interaction for emotionally salient stimuli in peri-hand space remains unexamined. Therefore, the present study investigated the temporal factors underlying the processing of emotionally salient stimuli in PHS. Participants completed standard temporal bisection task (N=34), that used a short (200 ms) & long (800 ms) as the standard duration. Participants had to judge whether the duration ranging from 200 ms to 800 ms (in steps of 100 ms) were closer to 'Short' or 'Long', in two hand conditions. First condition required participants to place their hands on the sides of a computer monitor, making the stimuli appear in peri-hand space (Near-hand). Second condition required the hands to be placed on the lap, making the stimuli appear outside of peri-hand space (Far-hand). The stimuli were high arousal (positive and negative) and low arousal (neutral) images. Participant's responses were compared between 'Near-hand' and 'Far-hand' conditions. Results showed lower Point of Subjective Equality (PSE) for high arousal stimuli ($M = 486$, $SD = 55.41$) compared to low arousal stimuli ($M=508$, $SD=60.11$), [$t(33) = -3.53$, $p = 0.001$] for the near-hand condition. No such PSE differences were obtained in the far-hand condition, [$p = .08$]. The findings imply temporal dilation for highly arousing emotional stimuli in PHS and not the space beyond, perhaps to allow for in-depth and detailed cognitive evaluation of emotional salience. The temporal dilation seems to be driven by early anticipatory mechanisms associated with PHS. Further, emotional salience seems to have spatial characteristics; arousing stimuli get prioritized in PHS and not the space beyond, thus providing allowance for better planning and execution of defensive actions in PHS.

16.345 FEEDFORWARD AND FEEDBACK EFFECTS IN VISUALLY PERCEIVED DURATION: THE INFLUENCE OF SEMANTIC AND MID-LEVEL VISUAL FEATURES

Martin Wiener¹, Anisha Krishnan¹; ¹George Mason University

A plethora of experimental evidence exists demonstrating that the perception of duration can be influenced by stimulus features such as size, numerosity, brightness, motion, emotion, clutter, and memorability. Extant theories to explain this effect have postulated a domain-general magnitude system, increased neuronal firing, attentional allocation, or processing efficiency. One method for determining how these effects are driven is whether they arise from feedforward, bottom-up processes across the visual hierarchy, or feedback, top-down processes from higher-order regions of the cortex back to lower-level ones. To address this, we conducted an experiment ($n=50$) wherein subjects were required to rapidly classify the durations of visual stimuli consisting of real objects or texture-matched images controlled for mid-level visual features yet devoid of semantic identity, known as "texforms" (Long, et al. 2016). The images consisted of animate or inanimate objects that were either large or small in real-world size, offering a test of magnitude-based accounts.

Notably, previous work has shown that, despite only containing mid-level features, texforms can still activate ventral stream responses – to a lesser extent and intensity – and retain some identifiable information. Results demonstrated that, for regular images, larger real-world size induced time dilation, with no effect of animacy, and no effect observed within texforms. Surprisingly, we also observed that regular images were dilated compared to texforms, to a much larger degree. A follow-up experiment ($n=50$) with texforms and regular images inter-mixed replicated these effects, again with no effects within texform images. To understand this further, we compared texform and regular images in a recurrent convolutional neural network. Here, we observed that regular images were processed faster than texforms through the network, but this effect was only found at higher layers. The findings thus support the idea of recurrent or feedback effects to visual cortex as driving time dilation effects.

16.346 SAMPLING MOTION IN TIME: DOES ALPHA ENTRAINMENT SHIFT THE PERCEIVED ONSET OF MOTION IN THE FRÖHLICH EFFECT?

Andrew Zavala¹, Paul Dassonville¹; ¹University of Oregon

Neural oscillations in the alpha-band frequency have been purported to gate perception through alternating states of neural excitation and inhibition. These alternations are theorized to modulate perceptual outcomes such as target detection thresholds and perceived timing (Mathewson et al., 2012). Does this feature of the visual system extend to the ability to localize moving targets? And can it provide an explanation for the Fröhlich effect, where an observer mislocalizes the starting location of a moving object in the direction of motion? In an initial experiment, participants reported the location of the starting point of a moving target that was presented after an auditory entrainment of alpha at 8.5 or 11.5 Hz. The magnitude of the Fröhlich effect was significantly smaller with the 11.5 Hz entrainment, suggesting that the faster alpha rhythm accelerated perceptual processing and allowing target localization earlier in its trajectory. However, the illusion magnitude was not modulated by alpha phase at target onset for either frequency. To ensure that entrainment was successful, a second experiment used audio-visual entrainment to test whether the Fröhlich effect would be modulated by alpha phase and included an additional task that has been used to demonstrate a modulation of the two-flash fusion effect with alpha phase (Ronconi et al., 2017). A Fast Fourier analysis replicated the finding of an alpha modulation of the two-flash fusion effect but again found no evidence that the Fröhlich effect is modulated by alpha phase. Our results provide some evidence that the Fröhlich effect is influenced by entrainment frequency, but not in a way that is detectable by phase alignment. It is possible that the illusion is partly the result of visual sampling delays driven by the alpha rhythm, but that this effect is diluted by additional mechanisms that impact higher levels of perceptual processing (e.g., motion extrapolation).

16.347 SYNCHRONY PERCEPTION IS SHAPED BY THE SIMILARITY BETWEEN FOVEAL AND PERIPHERAL REPRESENTATIONS

Nihan Alp¹ (nihan.alp@gmail.com), Zeynep Karagül¹, Bilge Sayim²;
¹Sabancı University, Istanbul, ²École Normale Supérieure (ENS), PSL University, CNRS

Peripheral vision differs from foveal vision in several respects, including a reduction in spatial resolution with increasing eccentricity, and different temporal processing. Recently, we found that the ability to discriminate the synchrony of two dynamic faces—one presented at the fovea and the other in the periphery—declined with increasing eccentricity. Interestingly, this decline was nonlinear: Performance was relatively stable at small eccentricities, and declined sharply at eccentricities larger than 6°. However, it remains unclear whether this decline is driven primarily by characteristics of peripheral vision, such as reduced spatial resolution and increased crowding—or by an increasing difference between features extracted in the fovea versus the periphery. Here, we addressed this question by investigating how foveal blur affects the perceived temporal synchrony between two faces, one presented foveally and the other peripherally (at 6° or 12°). By systematically varying the blur of the foveal face, we aimed to manipulate the similarity between foveal and peripheral inputs. The peripheral face was always unblurred. The foveal face was presented with one of four Gaussian blur levels ($\sigma = 0, 20, 40, 60$; hereafter blur0–blur60). Participants indicated whether the two face videos were synchronous or asynchronous. Participants' sensitivity at 6° was higher than at 12° across all blur levels, with a gradual decrease of the difference with increasing blur. At 6°, there was a gradual decline with increasing blur, followed by a steep decline at the highest blur level. At 12°, by contrast, sensitivity did not differ between moderate blur levels (blur0 to blur40), and there was a steep decline at the highest blur level. Taken together, our results indicate that synchrony perception depends not only on stimulus properties, but also on the location of stimuli in the visual field which shapes their perceived similarity.

N.A and Z. K. are supported by the Scientific and Technological Research Council of Turkey (TUBITAK 1001) under the Grant Number 122K922.

16.348 SYNCHRONY PERCEPTION OF DYNAMIC FACES IS LIMITED BY THEIR ECCENTRICITY – NOT THEIR DISTANCE

Zeynep Karagül¹ (zeynep.karagul1@sabanciuniv.edu), Bilge Sayim², Nihan Alp¹; ¹Sabancı University, Istanbul, ²École Normale Supérieure (ENS), PSL University, CNRS

Visual perception does not operate uniformly across the visual field. Peripheral vision—compared to foveal vision—has lower spatial resolution, stronger crowding, and different temporal processing. Temporal processing differences across the visual field, such as the perceived timing of dynamic stimuli, could plausibly influence whether two events are perceived as synchronous. Here, we investigated

whether the perceived synchrony of two dynamic faces is influenced by their locations in the visual field. We conducted two experiments in which participants judged whether two simultaneously presented faces were synchronous or asynchronous. In Experiment 1, we manipulated face-to-face distance (6°, 8°, 12°) in two conditions. Participants either fixated on the centrally presented face, while the other face was presented in the periphery (Center condition), or they fixated between the two peripheral faces, with one face at the same eccentricity in each hemifield (Symmetric condition). Sensitivity analyses revealed an interaction between fixation location and face-to-face distance. In the Center condition, sensitivity decreased with increasing distance. By contrast, sensitivity remained constant across distances in the Symmetric condition. These results indicate that—for the tested face-to-face distances—increased distances impaired synchrony perception only when one face was presented at the fovea and the other in the periphery. In Experiment 2, we compared eccentricity instead of face-to-face distance. The peripheral faces in both the Center and Symmetric conditions were presented at the same three eccentricities (6°, 8°, and 12°). We observed a robust main effect of eccentricity: sensitivity decreased with increasing eccentricity. Unlike in Experiment 1, this decrease did not differ between the Center and Symmetric conditions. Our findings show that synchronization sensitivity is limited more strongly by eccentricity than absolute face-to-face distance or fixation location. We suggest that the visual system compensates for inherent differences between foveal and peripheral vision, enabling effective integration of information across the visual field.

Z.K. and N.A. are supported by the Scientific and Technological Research Council of Turkey (TUBITAK 1001) under the Grant Number 122K922.

16.349 TEMPORAL ERROR MONITORING IS ROBUST TO WORKING MEMORY LOADS

Taylor Gepes-Carroll¹, Martin Weiner¹, Tutku Öztel¹; ¹George Mason University

Error monitoring involves the ability to keep track of one's own errors without receiving external feedback. Recent studies demonstrated that this ability holistically captures metric errors, including temporal, spatial, and numerosity domains, abiding by A Theory of Magnitude (ATOM; Walsh, 2003). Previous findings demonstrated that increased WM load in a dual-task paradigm, where the secondary task involved mental alphabetization, impaired error monitoring in a perceptual decision-making task (Maniscalco & Lau, 2015). However, this effect did not extend to temporal error monitoring (TEM), which highlights the mechanistic differences associated with these two metacognitive processes (Öztel & Balci, 2023). This discrepancy may be due to domain-specific WM load effects, where the mental alphabetization task may not have engaged the metric characteristics required for a diminishment in TEM. We investigated this possibility in two separate dual-task experiments. In both experiments, participants (N=44) performed temporal reproduction (2500 milliseconds) as the primary task, where they evaluated their own performance in terms of

magnitude (via confidence ratings) and error directionality (via short-long judgments) relative to a target duration on a trial-by-trial basis. Depending on the experiment, the secondary task involved alphabetizing a letter sequence (Experiment 1- direct replication of Öztel & Balci, 2023) or calculating numerical equations (Experiment 2-original experiment). We investigated TEM abilities separately for magnitude and error directionality under three WM load indices, including WM difficulty, number of letters/arithmetic operations, and WM accuracy. In both experiments, mean confidence was higher for easy WM trials. Furthermore, larger letter sequences were associated with increased “long” responses for error directionality judgments in Experiment 1. However, both experiments demonstrated intact TEM abilities for magnitude and directionality monitoring across the levels of all WM indices. Together, these results indicate that the sensitivity of TEM is robust to WM load, regardless of the metric characteristics associated with WM load.

16.350 A MEASUREMENT OF FLICKER DISCRIMINATION SENSITIVITY IN MIGRAINE WITH PHOTOPHOBIA

Ruby S Bouhassira¹, Samantha A Montoya¹, Kristina Kitsul¹, Alan A Stocker¹, Geoffrey K Aguirre¹; ¹University of Pennsylvania

While people with migraine and photophobia experience discomfort from bright and flickering lights, there has been limited study of psychophysical discrimination sensitivity in this population. We designed a measure of wide-field flicker frequency discrimination and examined if people with migraine have altered sensitivity, especially for flicker presented at higher light levels. On each of 1600 trials over 4 sessions, participants with migraine with aura and photophobia, and headache-free controls (n=9 and 12, respectively, of 15 planned for each group) viewed simultaneous, achromatic, sinusoidal flicker presented separately to the left and right hemifield of a binocularly fused, dichoptic stimulus field (30° spatially uniform). Participants performed a same/different judgement (ignoring phase) and received auditory feedback. Sessions alternated between a high (~2100 cd/m²) and low (~10 cd/m²) luminance background. High (30%) and low (10%) contrast levels were intermixed. All trials presented one of five reference frequencies (10, 13, 17, 23, 30 Hz) and the test frequency was guided by an adaptive procedure. We estimated internal noise in frequency discrimination by fitting a probabilistic model to the psychophysical data, and then assessing sensitivity across group, reference frequency, contrast, and background luminance conditions. At this intermediate stage of data collection (~2/3 complete), we did not find significant interactions between group and reference frequency, contrast, or light level. Across groups, discrimination was uniformly better at high vs. low contrast ($F(1,19) = 43.76$, $p < 0.001$), as expected. Discrimination also improved with increasing reference frequencies, departing modestly from Weber's law ($F(4,76) = 4.00$, $p = 0.005$). Despite experiencing discomfort from flickering stimuli, we find that people with migraine and photophobia are able to discriminate the appearance of flickering stimuli in a manner similar to headache

free controls. It remains possible that subtle differences between the groups will be revealed in our final, pre-registered cohort.

1R01EY036255

Functional Organization of Visual Pathways: Cortical visual processing 1

FRIDAY, MAY 15, 3:45 – 6:00 PM, BANYAN BREEZEWAY

16.352 TEMPORAL CONTEXT OF VISUAL INPUT INFLUENCES CENTER-SURROUND INTERACTIONS IN PRIMARY VISUAL CORTEX

Joseph Emerson¹, Brock Carlson¹, Ryan Holland^{2,3}, Yifan Hu^{2,3}, Audrey Sederberg^{4,5}, Gordon Smith^{2,3}, Cheryl Olman^{1,2}; ¹Department of Psychology, University of Minnesota - Twin Cities, ²Department of Neuroscience, University of Minnesota - Twin Cities, ³Optical Imaging and Brain Sciences Medical Discovery Team, University of Minnesota - Twin Cities, ⁴School of Physics, Georgia Institute of Technology, ⁵School of Psychology, Georgia Institute of Technology

Surround suppression in primary visual cortex (V1) is a form of contextual modulation in which neurons are suppressed by the presence of stimuli outside their classical receptive fields. However, the effects of temporal context on center-surround interactions remains poorly understood. To probe how temporal context influences center-surround interactions, we used two-photon calcium imaging in anesthetized ferrets while presenting drifting grating stimuli in two paradigms: a heterogeneous condition in which the grating is stepped rapidly through eight directions over 5 seconds, and a homogeneous condition in which the grating drifted in one direction before reversing mid-trial during 3-second trials. Each paradigm was shown both in a 15° aperture and full-field configuration to measure classical and extra-classical receptive field contributions. To our surprise, we found that population responses revealed stronger and more consistent surround suppression under heterogeneous stimulation, whereas the homogeneous stimulus produced a broader distribution spanning suppression to facilitation. These results provide evidence that the temporal context of orientation heterogeneity strongly influences the degree of surround suppression in ferret V1. Ongoing work aims to explain the sensitivity of local recurrent interactions that facilitate surround suppression to temporally-dependent network states.

NIH R01 NS123482

16.353 SPATIAL FREQUENCY ENCODING AND DECODING IN MACAQUE V1

Xin Wang¹, Shu-Chen Guan¹, Shi-Ming Tang¹, Cong Yu²; ¹Peking University, ²Zhejiang University

The scarcity of V1 neurons selective for low spatial frequencies (<1 cpd) raises the question whether low spatial frequency information is encoded in population activity. We analyzed V1 population data collected using two-photon calcium imaging in multiple FOVs (each containing >1000 neurons) while presenting a parafoveal Gabor in twelve orientations (0°–180°), six SFs (0.25–8 cpd). PCA analysis of population responses revealed that, like high spatial frequencies, low spatial frequencies (0.25, 0.5, & 1 cpd) formed distinct manifolds in representational space, demonstrating robust encoding of low-frequency information in the geometry of V1 population responses. Moreover, as each frequency-specific manifold originate from trials at different orientations, we used manifold volume to estimate orientation discriminability at each frequency. Volumes were minimal at very low frequencies and increased until peaking at 4 cpd before declining at 8 cpd, indicating that orientation discriminability was poor at low frequencies and became greatest at intermediate frequencies. To compare low- and high-spatial-frequency decoding, we trained a transformer-based model to reconstruct the stimuli, achieving >90% accuracy. An ablation experiment examined the dynamics of stimulus reconstruction. Neurons were incrementally added according to their self-attention scores while others were nullified to track how stimulus features were recovered. For higher spatial frequencies (1–4 cpd), reconstruction first recovered orientation before frequency. However, for lower spatial frequencies (0.25–0.5 cpd), reconstruction either recovered frequency first or frequency and orientation simultaneously. These results indicate that spatial frequency and orientation decoding follows a coarse-to-fine principle, recovering coarser orientation first at higher spatial frequencies. However, as lower spatial frequency is a coarser feature, frequency may be recovered either first or simultaneously with orientation. We conclude V1 encodes low spatial frequencies through population-level geometry and supports a coarse-to-fine decoding process reflected in the dynamics of spatial frequency and orientation reconstruction.

STI2030-Major Projects grant (2022ZD0204600)

16.354 CONTRAST ADAPTATION IN THE RECEPTIVE FIELD STRUCTURE OF FOVEAL V1 NEURONS

Maxwell J. Greene¹, Ramon Bartolo Orozco¹, Felix Bartsch^{1,2}, Jasper C. Hood², Daniel A. Butts², Bevil R. Conway¹; ¹Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, ²Program in Neuroscience and Cognitive Science and Dept. of Biology, University of Maryland

Neurons in V1 optimize their ability to signal spatiotemporal patterns by bringing their dynamic range into register with the prevalent scene contrast. Prior studies have also shown that contrast can influence receptive field (RF) spatial structure in V1, though this work has been largely limited to the parafovea and periphery. Here, we describe the effect of varying luminance contrast on the space, time, and color RFs of V1 cells receiving input from the center-of-gaze. We measured the responses of foveal V1 units in awake fixating macaques to dynamic noise, which consisted of a luminance (L+M) component as well as L-

M and S-varying components. In one condition ("high-contrast"), the luminance contrast spanned the full range permitted by the display, which approximates the statistics of natural scenes in which luminance variation exceeds chromatic variation. To investigate the dependence of V1 color sensitivity on environmental color statistics, we included a second condition ("low-contrast") where the luminance contrast was restricted to ~10% of the maximum to match the L-M and S contrast of the high-contrast condition. Fine-grained spatiochromatic RFs were determined using linear and quadratic data-driven models of each neuron, which incorporated eye tracking to precisely estimate the retinal location of the stimulus. Under low contrast, the population mean firing rate was 12.4±9.0 Hz (s.d.), whereas under high contrast it was 16.7±11.0 Hz (N = 33). This firing rate increase of only ~0.13 log units over a log unit increase in luminance contrast is clear evidence of contrast adaptation. Across contrast levels, chromatic and achromatic RF structure remained largely unchanged, although the magnitude of the luminance filter increased in low contrast. Our results are consistent with early visual mechanisms that preserve foveal RF structure across large variations in illumination, which is likely important for the maintenance of high acuity vision.

BRC: NIH IRP 1ZIAEY000558; NSF 0918064; NIH R01 EY023322; DAB: NSF IIS-2113197; NIH R01 EY037347-01. Contributions of NIH authors are considered Works of the United States Government. The findings and conclusions are those of the authors and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

16.355 ORIENTATION AND VISUAL FIELD ASYMMETRIES FOR CONTRAST-MODULATION STIMULI: IMPLICATIONS FOR MAGNOCELLULAR PROCESSING

Ana Ramirez¹ (ana.ramirezherandez@mail.mcgill.ca), Curtis Baker¹, Ari Rosenberg²; ¹McGill University, ²University of Wisconsin-Madison

Parasol, Y-like retinal ganglion cells (RGCs) are functionally distinct in that they respond nonlinearly to high spatial frequencies (SFs) and temporal frequencies (TFs). Neurophysiology has indicated that cat Y-cells prefer orientations aligned with the area centralis, which is analogous to the fovea. Contrast-modulation (CM) stimuli (high-SF carrier gratings contrast modulated by a low-SF sinewave envelope) selectively engage the nonlinear mechanisms of Y-like RGCs which initiate the magnocellular pathway. Thus, CM stimuli offer a promising tool for probing magnocellular function in pathologies such as glaucoma. Human visual fields (VFs) measured with simple gratings typically show high foveal sensitivity and polar-angle asymmetries. Previous human psychophysics demonstrated that CM motion performance varies little with eccentricity. Here we investigate whether CM motion performance in humans shows fovea-aligned orientation biases for the carrier, and whether CM motion performance reveals VF asymmetries. Achromatic CM stimuli were presented on a CRT display and viewed monocularly. The stimuli consisted of a sinewave-grating envelope (0.50 cpd) drifting at 3 Hz, which modulated the contrast of a high-SF sinewave carrier (6 cpd) flickering at 15 Hz, with envelope

and carrier contrasts of 80-100%. The stimuli were presented at 8 polar locations, 4.3 deg eccentricity from fixation. Observers reported the direction of CM envelope motion on each trial (2afc button-press). Envelope motion direction discrimination performance was measured for 4 carrier orientations at each VF location. The results show CM envelope motion performance that varied systematically with carrier orientation, with the optimal orientation often aligned towards the fovea. Performance was better in the lower than upper VF. There was not a clear horizontal-vertical anisotropy. These findings indicate that CM perimetry may exhibit fovea-aligned orientation biases consistent with Y-like processing, while the VF dependence might exhibit some polar angle asymmetries as seen for simple sinewave gratings.

Canadian NSERC (RGPIN-2023-03559) to CB. National Institutes of Health Grants (EY029438 & EY035005) to AR.

16.356 RESTING-STATE APERIODIC EEG ACTIVITY DOES NOT PREDICT THE STRENGTH OF VISUAL SURROUND SUPPRESSION

Yiming Jin¹ (jinyiming1@zju.edu.cn), Nan Jin^{1,2}, Yongchun Cai¹; ¹Zhejiang University, ²Ludwig-Maximilians-Universität München

The aperiodic exponent (spectral slope) of the neural power spectrum has recently emerged as a putative non-invasive marker for cortical excitation/inhibition (E/I) balance. While this metric reflects the global state of cortical circuitry, it remains unclear whether intrinsic resting-state E/I balance constrains specific, localized sensory computations. Here, we investigated whether resting-state EEG 1/f slopes predict the strength of Visual Surround Suppression (VSS), a canonical inhibitory mechanism in the visual cortex wherein neural responses to stimuli within the classical receptive field (CRF) are suppressed by stimulation of the extra-classical surround. We recorded resting-state EEG from 25 healthy adults and extracted 1/f slopes from the occipital cortex using spectral parameterization (i.e., FOOOF). We then measured VSS strength for each participant using three independent paradigms to ensure construct validity: (1) a contrast matching task (measuring the suppression of apparent contrast by a surrounding grating); (2) a motion discrimination task (measuring the effect of stimulus size on duration thresholds, indexing spatial suppression); and (3) an SSVEP paradigm (measuring the modulation of signal-to-noise ratio in steady-state responses to a central flicker by a static surround). At the group level, all three paradigms yielded robust surround suppression effects, confirming the validity of our measurements. However, correlation analyses revealed no significant relationship between resting-state occipital 1/f slopes and any behavioral or neural measures of VSS strength across individuals. These findings suggest that while the 1/f slope may index global cortical E/I tone, visual surround suppression relies heavily on local circuit dynamics (such as divisive normalization) that operate independently of the baseline cortical state. Our results imply a mechanistic dissociation between global resting-state E/I balance and the local inhibitory computations governing sensory processing.

16.357 SSVEP REVEALS CONSISTENT INTEROCULAR SUPPRESSION AND SENSORY EYE DOMINANCE ACROSS BINOCULAR RIVALRY, MATCHING, AND STEREOPSIS PARADIGMS

Chao Han¹, Xiao Shang¹, Wanyi Huang²; ¹Shanghai University of Sport, ²Shanghai Normal University

The human visual system integrates inputs from both eyes to form a coherent and stable perceptual experience. Imbalanced interocular suppression can lead to sensory eye dominance (SED), potentially impairing binocular stereopsis. Conventional SED assessments typically rely on subjective reports during dichoptic presentation of differing stimuli, a condition that differs markedly from natural viewing. Moreover, it remains unclear whether interocular suppression mechanisms operate consistently across different binocular visual contexts, especially whether suppression is restricted to differences in orientation channels or extends more generally across stimulus features. Here, we objectively quantified eye dominance and interocular suppression using steady-state visual evoked potentials (SSVEP) in three binocular paradigms: (1) binocular rivalry, with orthogonally oriented gratings presented dichoptically; (2) binocular identical, with matching grating orientations in both eyes; and (3) binocular disparity, employing random-dot stereograms to promote the integration of input from both eyes. These paradigms systematically varied stimulus similarity and ecological validity. Results revealed that (1) psychophysically measured SED correlated strongly with SSVEP-based measures; (2) the magnitude of interocular suppression did not differ significantly across the three paradigms; and (3) the SSVEP approach remained robust when interocular stimulus contrast was deliberately unbalanced in order to quantify the interocular suppression. These findings demonstrate that SSVEP provides an objective, bias-free method for assessing interocular suppression, offering a promising avenue for evaluating SED in populations less able to perform behavioral tasks (e.g., infants and clinical groups). Moreover, early interocular suppression appears to operate broadly across stimulus features rather than being confined to orientation channels. The results enhance our understanding of binocular visual processing and support the use of binocular identical paradigms in SED assessment.

16.358 THE SECOND VISUAL AREA STRAIGHTENS NATURAL VIDEO TRAJECTORIES

Tien Dung Nguyen¹ (tien-dung.nguyen@utexas.edu), Yoon H. Bai², Corey M. Ziemba³, Olivier J. Hénaff⁴, Robbe L. T. Goris¹; ¹The University of Texas at Austin, ²Massachusetts Institute of Technology, ³National Eye Institute, ⁴Cursive, London, UK

Primates rely on their visual system to infer current and future states of the environment. Natural visual input typically evolves along complex temporal trajectories that are difficult to extrapolate. Recent perceptual experiments in humans and physiological experiments in monkeys suggest that the visual system transforms natural video

trajectories into straighter neural representations, thereby facilitating temporal prediction. While both types of experiments yielded qualitatively similar results, perceptual effects were substantially stronger. Which factors underlie this discrepancy? Here we tested two hypotheses. First, that the discrepancy is due to differences in stimulus location. Stimuli in the perceptual study overlapped with both the central and peripheral visual fields, whereas stimuli in the physiological study were limited to the periphery. If perceptual straightening varies with eccentricity, this could account for the discrepancy. We measured perceptual curvature of 6 video sequences presented in the fovea, parafovea, and periphery and found that natural videos elicit specific and robust straightening effects across the visual field (median change in curvature: fovea: -45.7° , parafovea: -58.8° , periphery: -40.4°). Alternatively, the discrepancy may be due to differences in neural processing stages. Perceptual experiments engage many stages of the visual hierarchy. In contrast, the physiological experiments were exclusively conducted in the primary visual cortex (V1). Perceptual straightening may arise from a cascaded computation that also involves later processing stages. To test this hypothesis, we conducted new electrophysiological experiments in two awake, fixating macaques. We measured the curvature of neural population trajectories elicited by 16 videos in the first and second visual area (V2). We found that natural videos elicit specific neural straightening effects that increase in strength between V1 and V2 (median change in curvature: V1: -2.7° , V2: -10.5°). Together, these results demonstrate that perceptual straightening arises from a cascaded computation that involves at least the early and mid-level processing stages.

16.359 TASK STRUCTURE AFFECTS EYE MOVEMENTS AND MODULATES HEMODYNAMIC ACTIVITY IN EARLY VISUAL CORTEX

Philipp Kreyenmeier¹, Corey M. Ziemba¹, Klaury Youchom-Tagheui¹, A. Tyler Morgan¹, Hendrikje Nienborg¹, Elisha P. Merriam¹; ¹National Institutes of Health

Even in the absence of visual stimulation, early visual cortex (EVC) exhibits hemodynamic responses related to task structure, task difficulty, and arousal state. The origins underlying this widespread response in EVC remain unknown. Here we test whether task predictability or external behaviors modulate the task-related response. Participants performed a rapid serial visual presentation task at the screen center, followed by a resting period during which participants saw the same visual stimuli, but did not perform a task. In Experiment 1, the duration of the rest period was fixed, making the onset of each trial predictable. In Experiment 2, we made the trial onsets unpredictable by varying the rest period from trial to trial. We measured eye movements, heart rate, and respiration, simultaneously with fMRI BOLD responses. Pupillometry, eye movements, and blinks showed a strong periodic modulation at the task frequency. Pupil size increased during the task phase and decreased during the resting phase. Saccades and blinks were suppressed during the task phase

and showed a pronounced release during the resting phase. Similarly, the BOLD response over EVC showed a strong periodic modulation at the task frequency. Whereas visually-stimulated foveal regions showed an increase in the BOLD response during the task phase and a decrease during the resting phase, unstimulated peripheral regions showed a later BOLD response peaking towards the end of the trial. This task-related BOLD response persisted when the timing of the task was unpredictable (Experiment 2), and when accounting for the systematic modulations in blink rates. Our findings are broadly consistent with previous imaging studies in monkeys and humans. Extending these earlier findings, we show that neither temporal predictability, nor task-related blinking can account for widespread task-related hemodynamic activity in EVC.

16.360 TOP-DOWN FOVEAL FEEDBACK CAN FOLLOW FEEDFORWARD VISUAL INPUT SIGNALS WITHIN 10 MS: EVIDENCE FROM BACKWARD MASKING

Li Zhaoping¹; ¹University of Tuebingen, Max Planck Institute for Biological Cybernetics

Core object recognition refers to the idea that primates can rapidly identify objects—largely independent of viewing distance or angle—through a single feedforward sweep of visual signals along the ventral visual pathway. This sweep, thought to take about 100–200 ms and to culminate in object representations in inferotemporal cortex before the first eye movement, has been considered largely free of top-down influence (DiCarlo et al., 2012). Recent experiments challenge this view by showing that perception of low-level features such as orientation and depth is strongly shaped by top-down feedback arriving within 50 ms after the first feedforward signals. In these studies, observers judged the depth of a surface in a random-dot stereogram (RDS) or the orientation of texture elements (Zhaoping, iScience 2025; VSS and ECVF abstracts 2025). Stimuli were presented briefly in the central visual field for a duration Δt , then replaced by a mask to disrupt feedback. Masks were either unrelated to the stimulus or were subsequent frames of a dynamic RDS, which preserved overall dichoptic statistics while the random dots were independently regenerated. On each trial, the stimulus was constructed to be either reliable—activating V1 neurons tuned to the true depth or orientation feature value—or misleading—activating neurons preferring the opposite feature value. Misleading inputs induce illusory percept of the opposite depth or orientation feature unless vetoed by the feedback process. When Δt was below ~50 ms, masking allowed these illusions to be visible. By mixing the reliable and misleading stimuli, the experiments revealed that the earliest effective feedback can act at Δt of roughly 10 ms or less. These results imply that, in the central visual field, visual inputs must be limited to ~10 ms and rapidly backward-masked if we wish to isolate purely feedforward processing in the visual pathway.

work supported in part by the University of Tübingen and the Max Planck Society

SATURDAY MORNING POSTERS IN BANYAN BREEZEWAY

Perceptual Training, Learning and Plasticity: Psychophysics

SATURDAY, MAY 16, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY

23.301 TOWARD A MORE ACCURATE ESTIMATION OF PERCEPTUAL LEARNING MAGNITUDE IN THE TEXTURE DISCRIMINATION TASK

Teruaki Kido¹, Yuka Sasaki¹, Nitzan Censor^{2,3}, Takeo Watanabe¹;
¹Brown University, Providence, USA, ²Sagol School of Neuroscience,
Tel Aviv University, ³School of Psychological Sciences, Tel Aviv
University

Perceptual learning has traditionally been induced and studied through repetitive practice. The texture-discrimination task (TDT; Karni & Sagi, 1991) is a standard training paradigm for perceptual learning research, yet several procedural and analytical aspects may limit the accuracy with which learning magnitude is estimated. Here, we examined methodological factors that may have influenced earlier findings and propose modifications to improve the reliability of TDT-based learning measures. Specifically, we evaluated three potential issues: (1) imperfect gaze control, (2) task complexity during the response phase, and (3) analyses potentially prone to bias, which could lead to undesirable gaze strategies, non-perceptual improvements, or estimation bias in task performance. First, we investigated how the lapse rate might affect performance measures. Specifically, we simulated data computationally while varying the underlying lapse rate, then compared the performance estimates from different lapse-rate modeling approaches. We found that conventional approaches could significantly overestimate performance at relatively high lapse rates, which may occur in the TDT due to its dual-task nature. Next, we conducted a single-session behavioral experiment with the TDT to test procedural and analytical modifications. We monitored gaze, preventing its deviations in real time, and simplified the response procedure. To reduce bias, we added catch trials with long stimulus-to-mask onset asynchronies and used a bias-robust model (Prins, 2012). When data and analysis aligned with those of previous studies, the estimated task performance was consistent with that in the preceding studies, confirming task compatibility despite the procedural modifications. However, with catch trials and the bias-robust method, performance appeared worse than in earlier findings, suggesting overestimation in those studies. Together, these findings underscore the importance of improved data modeling and procedural control. Implementing the proposed modifications may enable more accurate

estimation of perceptual learning magnitude in TDT-based experiments.

NIH R01EY027841, NIH R01EY019466, NIH R01EY031705, and BCS-2241417

23.302 POSITION-SPECIFIC LEARNING: EVIDENCE FROM A HIGH-LEVEL TEXTURE IDENTIFICATION TASK

Pranjal N. Patel¹ (patep180@mcmaster.ca), Jamie G.E. Cochrane¹, Allison B. Sekuler^{1,2,3}, Patrick J. Bennett¹; ¹Department Of Psychology, Neuroscience & Behaviour, McMaster University, ²Rotman Research Institute, Baycrest Academy for Research and Education, ³Department Of Psychology, University of Toronto

Visual Perceptual Learning (PL) typically exhibits stimulus and position specificity: learning does not fully generalize to new stimuli or to familiar stimuli presented at new positions. One explanation of specificity is that PL results from neural changes in cortical areas such as V1 and V2, and specificity occurs because neurons in those areas respond to specific stimulus characteristics (e.g., orientation) at particular locations in the visual field. An alternative explanation is that learning alters how higher-order mechanisms integrate responses from lower-level inputs. According to this explanation, specificity reflects what is learned rather than where it occurs in the visual pathway (Mollon & Danilova, Spatial Vision, 1996). As a continuation of Cochrane et al., (Journal of Vision, 2025) we investigated this alternative explanation by examining the position-specificity of PL in a 1-of-5 texture identification task requiring higher-order processing. The stimuli were band-limited noise textures that varied randomly on every trial. Because the textures varied randomly, observers could not learn idiosyncratic features of a small set of stimuli. Instead, they had to adopt a strategy that could be applied to a large population of bandpass textures. During training, all textures were presented slightly above or below a central fixation point, and the position was switched during the test phase. Preliminary results indicate that learning occurs in this task (i.e., identification accuracy improved during training) and that there is partial transfer of learning to a new position. These preliminary results suggest that position-specific learning can occur in a high-level task, and support the idea that specificity reflects what is learned rather than where it occurs in the visual pathway.

NSERC

23.303 THE EFFECT OF DOUBLE TRAINING ON POSITION TRANSFER OF LEARNING IN FACE AND TEXTURE IDENTIFICATION TASKS

Jamie G.E. Cochrane¹, Allison B. Sekuler^{1,2,3}, Patrick J. Bennett¹;
¹Department of Psychology, Neuroscience and Behaviour, McMaster University, ²Rotman Research Institute, Baycrest Academy for Research and Education, ³Department of Psychology, University of Toronto

In the double training paradigm, two tasks are trained at different visual field positions to promote transfer of perceptual learning between

them. Prior studies using this method have focused on learning in low-level feature detection and discrimination tasks. We investigated whether double training increases transfer in identification tasks that rely on inferotemporal processing. Previous studies found that perceptual learning in texture and face identification is stimulus- and position-specific. We tested whether double training increases transfer across positions in these tasks. Participants completed two interleaved 1-of-5 identification tasks with two sets of faces or textures. During training, one stimulus set was always shown above fixation, and the other was presented below. At test, positions of the two stimulus sets were reversed. For the texture task, double training successfully induced full transfer across positions: accuracy at the new location matched accuracy at the end of training. However, transfer was asymmetric for the face task. Full transfer occurred only for faces presented above fixation during training, perhaps because identity information is concentrated near the eyes and is therefore closer to fixation for faces tested below. We further investigated the effects of task relevancy by training a new set of participants with intermixed face and texture identification tasks. During training, the positions of the textures and faces were fixed (above or below fixation) and the positions were switched during testing. Only partial transfer was found in both tasks. Compared to the first experiment, learning in the texture identification task exhibited less transfer to a new stimulus location, but learning in the face identification task, when trained above, exhibited greater transfer to the test location. These results show a potential benefit of double training for inducing position transfer in higher-level tasks; however, factors such as task relevance appear to undermine its efficacy.

NSERC

23.304 TASK-IRRELEVANT DELAY FAILS TO DISRUPT SUBJECTIVE FEELINGS OF OWNERSHIP, AGENCY, AND PRESENCE IN VIRTUAL REALITY

Asaf Ali Altun¹ (altunalia@gmail.com), Loes CJ van Dam², William Harwin¹, Peter Scarfe¹; ¹University of Reading, ²Technical University of Darmstadt (TU Darmstadt)

Feelings of ownership, agency, and presence (OAP) are fundamental to virtual reality experiences, yet they are susceptible to disruption by factors such as temporal delay, which creates a mismatch between predicted and observed sensory consequences. Previous research by van Dam and Stephens (2018) has demonstrated that users can adapt to sensory feedback delays and that this results in concurrent improvements in OAP. However, it remains unclear whether any form of temporal delay in general disrupts OAP, or if the delay must be relevant to the user's immediate task. Here, we present data replicating and extending van Dam and Stephens (2018) using a virtual reality target-tracking task, wherein participants controlled a cursor to match the 3D position of a moving target. The novel aspect of our experiment involved introducing a temporal delay into either translation (task-relevant) and/or rotation (task-irrelevant) of the cursor. This allowed us to determine if the temporal delay needed to

be task-relevant to impact feelings of OAP. Our results confirm previous findings: Imposing a 200ms temporal delay in the 3D position of the cursor initially results in decreased feelings of OAP, but over time, as participants adapt to the temporal delay and get better at the tracking task, feelings of OAP increase. Critically, however, we found that temporal delay applied to rotation had no significant effect on either task performance or subjective measures. These findings demonstrate that one can experience significant temporal delays in a virtual environment without diminishing ownership, agency, or presence, provided those delays are irrelevant to the task. This suggests a close linkage between sensory adaptation driven by task-relevant error and the restoration of subjective feelings of OAP, rather than sensory (in)consistency in general.

This research was generously supported by the Republic of Türkiye Ministry of National Education.

23.305 LEARNING DYNAMICS REVEAL THE NATURAL BASIS OF VISUAL REPRESENTATION

Zirui Chen¹, Michael Bonner¹; ¹Johns Hopkins University

What are the axes along which visual systems learn and represent information? Studies using deep neural networks have debated whether sensory representations are best understood by focusing on individual neurons or latent dimensions of population activity. We approach this debate from the perspective of learning dynamics: which representational basis reveals the structure of how features develop during training? We trained ResNet-50 models on ImageNet and tracked the developmental trajectories of individual neurons versus principal components (PCs). We defined PCs from the fully trained network and projected activations from all training epochs onto this fixed PC basis to assess when each component emerges. We found clear hierarchical learning dynamics: lower-rank PCs with more variance are almost fully developed after just one epoch, while higher-rank PCs converge progressively throughout training, with some not stabilizing until the final epochs. This rank-dependent trajectory confirms theoretical predictions and reveals a natural learning process where the most important features emerge first. In contrast, individual neurons show no such differentiation in their developmental trajectories. There is no clear distinction between early-developed versus late-developed neurons. Critically, the hierarchical PC learning pattern holds robustly even when we incorporate spatial topographic constraints that induce distinct functional clustering in the fully trained network. These findings demonstrate that PCs, not individual neurons, constitute the natural basis for understanding visual feature learning and representation, suggesting that population-level latent dimensions should be the primary unit of analysis for interpreting neural codes.

23.306 VISUAL ADAPTATION SHIFTS ANIMACY JUDGEMENTS ACROSS PHOTOGRAPHIC AND CARTOON FACES

Logan Wu¹, Jessica Huang¹, Jason Haberman²; ¹Rhodes College

Our visual system recognizes faces across a wide range of depictions. We can readily recognize a familiar individual stylized to look like a cartoon or even a caricature. Although face recognition is invariant across various depictions, we nonetheless exhibit extreme sensitivity to whether a face is real. This raises the question of whether realness is explicitly coded by the visual system, which we can explore using adaptation procedures. In our experiments, we examined how adaptation to either a photographic or cartoonish depiction of a face influenced subsequent realness judgments using the method-of-constant-stimuli. In Experiment 1, we generated interpolated morphs from 'real' to 'cartoonish.' Participants (N=27) first adapted for 45s to a single identity in either photograph or cartoon form, and then categorized whether a subsequently presented morph (between 20%-80% real) appeared more like a photograph or a cartoon. Each trial was preceded by a 3s top-up stimulus. Participants completed four blocks in random order (two blocks with a photograph adaptor and two blocks with a cartoon adaptor). Results revealed a stark shift in 'realness' thresholds — participants were more likely to call a morph real following adaptation to a cartoon, and vice versa. We replicated this pattern in a second experiment (N=23), in which the morph identity participants had to judge was randomly replaced on every trial with one of seven possible different identities. This confirmed that whatever features confer a sense of realness from a face remain consistent across different identities, even though other physical features may differ. Overall, our results show that animacy or 'lifelikeness' is an adaptable feature explicitly coded by the visual system.

23.307 PERCEPTUAL LEARNING AND TRANSFER OF MOTION DIRECTION DISCRIMINATION IN CHILDREN AND ADULTS

Sarit F.A Szpiro^{1,2}, Maysa Agbaria¹; ¹University of Haifa, ²The Edmond J. Safra Brain Research Center for the Study of Learning Disabilities, University of Haifa, Israel

Background: Perceptual learning (PL) involves improvements in perceptual task performance through practice. Here, we examine visual PL for motion direction discrimination in children and adults. Based on developmental differences in neural plasticity, we explored whether children and adults would differ in the rate and magnitude of learning, as well as in transfer of learning to untrained directions. Methods: We examined children (ages 8-10) and adults (ages 18-40) who underwent the same procedure. Testing sessions (pretest and posttest) were conducted on in controlled laboratory settings, while the three training sessions (days 2-4) were delivered via web browser on participants' home computers. Participants viewed random dot kinematograms stimuli and performed a motion direction discrimination task. Stimuli consisted of near-oblique ($40^\circ \pm 10^\circ$) or orthogonal ($320^\circ \pm 10^\circ$) directions, that were counterbalanced to serve as trained or untrained conditions across participants. During testing, motion coherence thresholds were measured using a staircase procedure for both trained and untrained directions to assess learning specificity and transfer. During training, participants only viewed the

trained directions. Results: Both age groups showed significant perceptual learning, with no differences in learning magnitude between children and adults. Moreover, both groups demonstrated complete transfer to untrained directions. Conclusions: These findings add to the growing literature on perceptual learning across development, suggesting that transfer patterns in motion discrimination may depend on task-specific factors rather than age-related differences in neural plasticity alone. Further research is needed to understand the conditions under which developmental differences in learning and generalization emerge.

Israel Science Foundation (ISF) [grant number 1198/22]

23.308 PLFEST: A PLATFORM TO SUPPORT PERCEPTUAL LEARNING RESEARCH

Boris Penalzoza¹, Marcello Maniglia², Jaap Munneke¹, C. Shawn Green³, Aaron Seitz¹; ¹Northeastern University, ²Rochester Institute of Technology, ³University of Wisconsin-Madison

Perceptual learning (PL), defined as experience-based improvements on perceptual tasks, is an important field addressing both mechanisms of neural plasticity and methodologies to improve perceptual function. However, the field faces major limitations in terms of reproducibility and accessibility. Inconsistencies across studies may stem from methodological variations or differences between research participants. Moreover, despite its potential as a cost-effective intervention for visual disorders, PL research is often restricted to university laboratories or specialized clinics. To address these challenges, we developed PLFest, a Unity-powered application that enables data collection across different sites and platforms (e.g., computers, smartphones, and tablets). PLFest currently supports multiple vision training paradigms, as well as hearing, attentional control, and cognitive assessments. Here, we evaluate PLFest's reliability using data from a large group of participants (N=241). We demonstrate that PLFest is a reliable platform for a variety of assessments including visual acuity, contrast sensitivity functions, visual search, and reading, among others, with data replicating well-established findings in the field. Further we show that the platform is appropriate to collect high-quality perceptual learning data using a variety of approaches to train contrast sensitivity (e.g., standard perceptual learning with adaptive staircases, training with external noise, training with collinear flankers, stimulus variety across multiple spatial frequencies and locations, and multisensory facilitation with auditory cues). These preliminary data position PLFest as a shareable and reproducible research platform capable of reliably measuring and training visual functions in large groups of participants using accessible tools such as computer tablets. This suggests that PLFest can support big-data approaches to vision science, as well as providing promising results for increased outreach and translational opportunities.

Visual Working Memory: Interactions with long-term memory

SATURDAY, MAY 16, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY

23.309 BEHAVIORAL AND NEURAL CONSEQUENCES OF ORIENTING INTERNAL ATTENTION BETWEEN SHORT- AND LONG-TERM MEMORY

William Narhi-Martinez¹, Kaiki Chiu¹, Anna C. Nobre¹; ¹Yale University

We investigated how internal attention is directed to contents within or between short-term (STM) and long-term memory (LTM) across three experiments. Many contexts in daily life are associated with both STM and LTM content. We wanted to understand how attention operates when directed to one versus both of these internal domains. For LTM, participants learned the identities and locations of two objects within various scenes. The next day, the same scenes appeared with two new objects superimposed, which participants encoded into STM. At the end of the trial, participants made a decision about an item associated with the scene, either in STM or LTM. They selected which of two items was associated with the scene (behavioral Experiments 1 and 2) or decided whether a single item was associated with the scene (fMRI Experiment 3). Before the response array, a spatial retrocue could indicate which side the relevant item for the response had been located within the scene. The target item would come from either STM or LTM. Experiment 1 revealed significant performance benefits of internal spatial attention for valid vs. neutral retrocues for both STM and LTM targets. Experiment 2 showed that performance for LTM targets, but not for STM targets, was impaired when retrocues prioritized items from both STM and LTM domains compared to when contents from a single domain alone were prioritized. Initial results from Experiment 3 indicated dorsal frontoparietal regions exhibited significant clusters of activation specific to when participants were cued to combined STM and LTM contents, compared to when they were cued to contents purely in the STM or LTM domains. These findings challenge the standard view that once an LTM is recalled into working memory, it is maintained in an equivalent format to new STM information encoded from the external sensory stream.

23.310 DOES REACTIVATION OF LONG-TERM MEMORIES IN WORKING MEMORY INCREASE VULNERABILITY TO INTERFERENCE?

Ali Yilmaztekin¹ (ali.yilmaztkn@gmail.com), Eren Günseli¹; ¹Sabancı University

Working memory (WM) is vulnerable to interference, whereas repeated exposure often reduces this vulnerability by strengthening long-term memory (LTM) traces. Prior work has shown that context changes can trigger active representation of LTM items in WM. However, it remains unclear whether such reactivation reinstates

fragile WM representations or robust LTM-like states. Across two experiments, participants repeatedly studied and reported the color of one (Exp 1) or two (Exp 2) memory items over six consecutive trials. On half of the 5th color repetitions, the background context was changed at encoding to trigger reactivation in WM (Özdemir et al., 2024). A subset of these trials also featured a delay-period interference task. Participants either located an odd-colored item from a circle of images (Exp 1) or checked for a repeated color in a grid of colors (Exp 2). In both experiments, memory errors decreased across color repetitions. The interference task increased memory errors for the first color repetition, but this interference cost diminished with subsequent repetitions. Importantly, a context change on the 5th color repetition did not increase the interference cost, which would be expected if reactivation reinstated vulnerable WM representations. These findings suggest that LTM traces continue to buffer against disruption even when they are reactivated in WM. Alternatively, reactivation may reinstate a representational format that is less vulnerable to interference than newly encoded information.

23.311 HOW DOES WORKING MEMORY PREPARE FOR MENTAL OPERATIONS ON LONG-TERM MEMORIES?

İlayda Efsane Algin¹, Eren Günseli¹; ¹Sabancı University

The role of working memory (WM) for mental operations is well established by both behavioral and neural studies. While these studies used novel information on each trial, many everyday mental operations rely on information retrieved from long-term memory (LTM). Here, we investigate the role of WM for mental operations on LTM using EEG. In Phase 1, participants learned five color–position associations. In each Phase 2 trial, a color cue indicated a position to be retrieved from LTM, followed by a novel position. Across blocks, the positions were used either in a recognition task (matching the probe to a memory position) or a mental integration task (computing their spatial midpoint). We applied a spatial inverted encoding model (IEM) on alpha-band activity to reconstruct retrieved positions (Sutterer et al., 2019) before encoding the second position. Early in the delay, CTF slopes indexing location selectivity were higher for integration than recognition blocks, but this difference disappeared later in the delay, indicating comparable WM involvement. Additionally, an MVPA classifier distinguished the task types during the retention interval, regardless of position information, suggesting a persistent difference in preparatory format. These results indicate that spatial information retrieved from LTM is initially represented more strongly in WM when preparing for integration than for recognition. However, both tasks converge to similar WM strength over time while maintaining distinct task-set representations.

23.312 LONG-TERM MEMORY IS REACTIVATED IN WORKING MEMORY WHEN STABILITY IS VIOLATED

Eren Günseli¹; ¹Sabancı University, Istanbul

What determines when long-term memory (LTM) is reactivated into working memory (WM)? Across three EEG experiments, we examined

how WM–LTM dynamics respond to different forms of contextual structure. Experiment 1 involved long runs of trials with one task rule, followed by a switch to a different stable rule (e.g., A–A–A–A–B–B–B–B–B). The same task-relevant memory color repeated across the entire block, including across the rule change. CDA reduced with color repetitions but increased at task rule transitions, reflecting reactivation of the task-relevant color. In Experiments 2–3, memory colors and contextual features (task rule or background color) were manipulated in a 2×2 blocked design: in each block, they either repeated across trials or varied from trial to trial. CDA was reduced in color repetition blocks, consistent with a shift from WM to LTM. Critically, even when contextual features varied from trial to trial, CDA remained low—indicating no reactivation. Together, these findings show that WM is re-engaged when stability breaks and that the brain interprets consistent variability as a stable context. WM reactivation thus reflects the detection of contextual transitions.

23.313 EVIDENCE FOR A DOMAIN-GENERAL METACOGNITIVE ABILITY FOR VISUAL WORKING MEMORY AND VISUAL LONG-TERM MEMORY

Saud Altayar¹ (saud.altayar@mail.utoronto.ca), Keisuke Fukuda^{1,2}; ¹University of Toronto Mississauga, ²University of Toronto

Given that not all visual memories are accurate, we must distinguish accurate from inaccurate memories to guide our behaviour (Kozlova et al., 2025). To test whether such metacognitive judgments share a common mechanism between visual working memory (VWM) and long-term memory (LTM), participants completed a visual learning task in which they memorized either one, three, or five colored silhouettes of real-world objects at a time. Importantly, the color memory for half of the silhouettes was tested 1000 ms after encoding (assessing VWM) while the remainder was tested approximately 30 minutes after encoding (assessing VLTM). To quantify memory quality for each silhouette (i.e., target), participants used a continuous color wheel to indicate their best estimate of the target color (i.e., point estimate) and the range of all possible colors they thought the target could have been (i.e., certainty range). Participants' memory distribution was captured as a triangle using the point estimate as the peak and the certainty range as the base. The peak's height was set so that the triangle's area was one. The cumulative area of the memory triangle was then computed as a function of the distance from the target color, and the area under this cumulative distribution function quantified the memory quality. To measure subjective assessment of the memory quality, participants also indicated their confidence in the point estimate's accuracy. Finally, we built participant- and memory- specific regression models to explain the within-subject variability in confidence from the memory quality, and the resultant R-squared was used as a measure of metacognitive accuracy. A preliminary principal component analysis (n = 49/84) revealed a single latent factor that accounted for more than a half of the variance across set-sizes and memory types. This suggests a common mechanism used to assess the quality of visual memories across VWM and VLTM.

This research was supported by the Natural Sciences and Engineering Research Council (RGPIN-2024-05727).

23.314 FAMILIARITY REDUCES SENSORY DEPENDENCE IN WORKING MEMORY

Yiheng Qiu¹ (yihengqiu1224@gmail.com), Mowei Shen¹, Yingtao Fu¹, Hui Chen¹; ¹Zhejiang University, Hangzhou, China

How does long-term experience influence the representations formed from current sensory input? Evidence from expertise and long-term memory research suggests that extensive experience can yield representations that are more structured and place less weight on sensory detail. From this perspective, we predicted that familiarity would make working-memory (WM) representations less sensitive to transient, task-irrelevant sensory variation. To test this prediction, we conducted four behavioral experiments using faces as a controlled stimulus domain. In Experiments 1a–1c (two-session design), participants first learned name–face associations for six identities and then completed a change-detection WM task in which they judged identity changes while ignoring changes in facial expression. We systematically manipulated the degree of stimuli overlap between learning and test: Experiment 1a used the same six expressions at both stages; Experiment 1b used three expressions at learning and a non-overlapping set of three at test; Experiment 1c used neutral faces at learning and a full set of expressions at test. Across these variants, expression changes reliably slowed responses, indicating incidental encoding of expression. Crucially, the magnitude of the expression-induced interference was consistently reduced for learned (familiar) identities compared with unfamiliar faces. Experiment 2 replicated this attenuation using naturally familiar celebrity faces without laboratory training, demonstrating that real-world long-term knowledge produces the same pattern. Together, these findings converge to show that familiarity makes WM performance less affected by irrelevant sensory variation. The results provide behavioral evidence that long-term knowledge influences the degree to which WM relies on sensory-specific information, with implications for theories of representational abstraction and interactions between WM and long-term memory.

23.315 LEARNING SHAPES NEURAL SIMILARITY OF VISUAL WORKING MEMORY REPRESENTATIONS

Vivien Chopurian¹, Jacob Miller², Lana Gaspariani¹, Janna Wennberg¹, Arielle Tambini^{3,4}, Anastasia Kiyonaga¹; ¹UC San Diego, ²University of Miami, ³Nathan Kline Institute for Psychiatric Research, ⁴New York University Grossman School of Medicine

The transient maintenance of a detailed visual memory relies on representations in distributed brain regions. These representations can change over a range of time scales, within a single trial or across days or even months. For instance, representations in visual areas may be transformed across a trial from a sensory-like code into a more abstract goal-directed format. We have previously shown that repeated exposure to complex visual stimuli over months of learning increased the expression of item-specific and categorical neural

representations in prefrontal cortex. Here, we reanalyze this dataset to examine how the representation of different visual properties evolves across learning to support behavior. This study takes a dense sampling approach, where 3 human subjects underwent repeated fMRI scanning over 3 months. During each session, they completed a series of tasks that incorporated a set of 18 unique fractal stimuli. During each trial of the working memory task, subjects had to remember either one of these trained fractals or a similar novel fractal over a jittered delay. Some fractals were remembered well from Day 1, whereas others started out poor but were learned over time. How might these behavioral patterns track with neural differentiation in visual areas? We specified hypothetical models of visual similarity based on output from different layers of a neural network (VGG-16). Then we tested how well those models explain the neural patterns across visual and parietal regions over time. Across all sessions, earlier model layers better explained patterns in earlier visual regions (e.g., V1, V2), while later model layers better explained patterns in later visual regions (e.g., V4). These patterns also changed on different timescales, illustrating how learning unfolds for various levels of visual representation in distributed cortical areas. Evolving neural differentiation in visual areas may support discrimination between confusable working memory stimuli with learning.

23.316 FAMILIARITY ENHANCES BUT DOES NOT ELIMINATE CAPACITY LIMITS IN VISUAL WORKING MEMORY

Rai Bari¹ (rai.s.bari@durham.ac.uk), Linda H. Lidborg¹, Holger Wiese¹, Anna Grubert¹; ¹Durham University

Visual working memory (vWM) capacity has been shown to decrease with increasing stimulus complexity. Faces are among the most complex stimuli, and prior work suggested that only one unfamiliar face can be actively maintained in vWM (Towler, Kelly, & Eimer, 2016). Here, we examined whether this limit increases for famous faces, for which long-term memory (LTM) representations might support vWM maintenance. Participants performed a lateralised change detection task. A memory array (200ms) contained one versus two (Experiment 1) or two versus three (Experiment 2) face identities. After an 800ms retention interval, a test array appeared (200ms), and participants indicated whether all identities were repeated or whether one had changed. In addition to behavioural accuracy, we measured the contralateral delay activity (CDA) of the event-related potential during the retention period as an electrophysiological index of the number of items actively maintained in vWM. Behaviourally, accuracy declined with increasing set size, confirming higher memory demands in higher load conditions. In Experiment 1, CDA amplitudes were larger for load-2 than load-1 trials, indicating that participants maintained two identities when required. A control experiment confirmed that these load effects were independent of low-level pictorial cues. However, in Experiment 2, CDA amplitudes were virtually identical in load-2 and load-3 trials, suggesting that vWM capacity was limited to two face identities. Overall, our findings indicate that familiarity can boost the

number of complex objects maintained in vWM beyond the previously reported limit of one. Nonetheless, capacity remains sharply restricted and well below the typical 3-4-item limit observed for simple features or basic objects. Even for highly overlearned, meaningful stimuli such as famous faces, vWM capacity appears limited to two identities, highlighting a boundary on the extent to which LTM can support the active maintenance of complex visual information in vWM.

This work was funded by research grants of the Leverhulme Trust (RPG-2020-319) awarded to AG and the ESRC (ES/X002063/1) awarded to HW.

23.317 WHEN 120 REPETITIONS STILL TEACH NOTHING: EVIDENCE FOR A FAST VISUAL CACHE AND A SLOW EPISODIC MEMORY SYSTEM

Chaoxiong Ye^{1,2} (chaoxiong.c.ye@jyu.fi), Penglan Liu², Dan Nie², Robert H Logie³; ¹School of Education, Anyang Normal University, China, ²Department of Psychology, University of Jyväskylä, Finland, ³Department of Psychology, University of Edinburgh, UK

Two experiments explored a previous finding that 120 repetitions of the same six-item array for change detection resulted in no or very slow learning. This contrasts with learning from six repetitions tested by recall. Shimi and Logie (2019) proposed that array repetition for change detection is supported by a limited capacity, temporary visual cache memory, with contents replaced by the next study array, even if it is identical. This is coupled with a weak episodic trace that strengthens across repeated presentations but requires more than 60 repetitions for learning. Experiment 1 tested whether (a) a short study-test interval would result in reliance on the visual cache with no evidence of learning across 120 repetitions, and (b) a longer study-test interval would gradually strengthen the episodic trace but require many repetitions for learning. A 500 ms study-test interval showed no learning after 120 repetitions, and participants reported being unaware of the repetition. A 5000 ms study-test interval showed performance improvements but only after 40 repetitions and participants reported becoming aware of the repetition. In Experiment 2, different arrays on each of 120 trials with short and long study-test intervals showed the same lack of learning found for the 500 ms study-test interval in Experiment 1. Results appear consistent with a limited capacity visual cache memory for change detection that retains the array only for the current trial, working in tandem with a weak episodic trace that accumulates across trials but only supports performance after multiple repetitions and longer study-test intervals.

This work was supported by the Research Council of Finland (former Academy of Finland) Academy Research Fellow project (#355369 to Chaoxiong Ye) and Finnish Cultural Foundation (#231373 to Chaoxiong Ye).

23.318 REMEMBERING FOR ACTION: DO ACTION PLANS AMPLIFY MEMORY REACTIVATION?

Busenur akbas¹ (busenur.akbas@sabanciuniv.edu), Caterina Trentin², Christian Olivers², Eren Günseli¹; ¹Sabancı University, Turkey, ²Vrije Universiteit Amsterdam

Action planning is increasingly recognized as a key driver of working memory (WM) engagement, as WM items associated with an action plan are stored via enhanced perceptual representations. Here, we tested the effects of action plans on long-term memory (LTM) retrieval across two experiments (N = 68). In Phase 1, participants studied color-shape associations. In Phase 2, the color was provided as a retrieval cue for shape. After a retention interval, a shape and a square were presented concurrently. In Action blocks, participants reported memory matches by clicking the remembered shape; in Recognition blocks, they clicked the square, eliminating the link between action and the retrieved memory content. No response was required when the probe shape differed from the memory shape. Importantly, in half of the trials, a visual search display was presented instead of the memory test. The search target appeared either inside the memory shape (congruent) or a distractor shape (incongruent). Responses were faster on congruent than on incongruent trials, and this congruency benefit was larger in Action than in Recognition blocks. Experiment 2 replicated this pattern while minimizing reliance on familiarity by using other studied shapes as lures rather than novel shapes. These findings suggest that LTM representations are more strongly reactivated in WM when they are linked to an action plan. Thus, while recognition can rely more on passive LTM, action preparation triggers the reinstatement of active WM, highlighting a distinction between memory for knowing and memory for doing.

23.319 SCENE METAMER JUDGMENTS REVEAL ATTENTIONAL BIASES IN VISUAL WORKING MEMORY

Abe Leite¹, Ritik Raina¹, Seoyoung Ahn², Gregory J. Zelinsky¹; ¹Stony Brook University, ²University of California, Berkeley

Scene representations evolve across viewing, with each fixation revealing not only the directly viewed object but also constraining inferences about the rest of the scene. We propose a novel paradigm for probing representations in visual working memory [VWM]. Participants view an image for a controlled number of fixations (1, 3, 5, or 10), then (following an 8s delay) see a second image for 200ms and must judge whether the second image was identical to the first. We use the term 'scene metamers' to describe physically different images judged to be identical, and propose that these images are confused because they have similar VWM representations. In this study, we directly test two hypotheses: first, that VWM is biased towards representing information at fixated locations (gaze-contingent bias); and second, that among non-fixated locations, VWM is biased towards representing information at semantically-meaningful locations that could potentially have been fixated (meaning-contingent bias). To do so, we employ the highly-controllable Seen2Scene model (presented at VSS 2025) to generate potential scene metamers that match the scene's overall structure, but are biased to retain high-resolution visual information from small patches (size of human

parafovea) at condition-dependent locations. In the own-fixation condition (gaze-contingent), these are the points the viewer fixated. In the same-image condition (meaning-contingent), they are the fixations of a different viewer on the same image. And in the cross-image condition (control), they are the fixations of a different viewer on a different image. Data from over 50 participants viewing 300 images each clearly reflect both of the hypothesized biases: gaze-contingent generations are more often metameric than meaning-contingent generations, which are in turn more often metameric than control generations. These effects hold when controlling physical and semantic similarity metrics. We conclude that our Seen2Scene-based paradigm is a promising way to probe the information underpinning visual memory representations.

AL is supported by NSF-GRFP #2234683 and NIH-NEI R01EY030669 to GJZ. RR and GJZ are supported by NSF-CompCog #2444540 to GJZ.

Visual Working Memory: Interference, attention

SATURDAY, MAY 16, 8:30 AM – 12:30 PM,

BANYAN BREEZEWAY

23.320 CAUSAL EVIDENCE FOR INTERFERENCE BETWEEN VISUAL WORKING MEMORY AND MOTION PERCEPTION USING TMS

Sanika Paranjape¹ (sanikap@gmail.com), Sarah Shomstein¹, Dwight Kravitz^{1,2}; ¹The George Washington University, ²US National Science Foundation (SBE/BCS)

Behavioral evidence has shown that there is bidirectional interference between visual working memory (VWM) and perception for low-level features (e.g., color, orientation) and high-level features (e.g., motion direction, face identity) (Teng & Kravitz, 2019; Paranjape et al, 2024). The sensorimotor recruitment hypothesis proposes that the same areas involved in encoding a stimulus are involved in maintenance, leading to interference between VWM and perception. Here, using a within-subjects design, we provide causal evidence for the involvement of MT+ cortex in mediating this interference for motion direction. In a dual-task paradigm, transcranial magnetic stimulation (TMS) was applied during the delay period between the encoding of a motion stimulus and a subsequent perceptual discrimination. Two discrimination conditions were used: 1) middle, in which the two discrimination motion directions are on either side of the maintained motion direction (in a continuous space); and 2) side, in which both discrimination motion directions are off to one side of the maintained motion direction. We replicated a behavioral effect supporting sensory recruitment, finding that the maintained information interfered with discrimination, resulting in higher discrimination thresholds for the middle condition than the side condition. The discrimination motion

directions also introduced systematic biases in the continuous report of the maintained motion, with higher accuracy for the middle than the side condition. Critically, when TMS was applied to MT+ cortex, consistent with the prediction that maintenance occurs within MT+ cortex, we found a significant decrease in the interference compared to sham TMS applied at vertex. These results provide direct causal evidence that VWM maintenance occurs in the same cortices that are involved in perceptual processing of high-level stimuli, such as motion direction.

NSF BCS-2022572 to SSS and DJK; W911NF-23-2-0097 to DJK and SRM

23.321 DYNAMIC MASKS INTERFERE WITH ICONIC BUT NOT VISUAL WORKING MEMORY

Caroline Myers¹, Justin Halberda¹; ¹Johns Hopkins University

Which forms of memory truly depend on sensory codes, and which do not? Iconic memory depends on a sensory code - it preserves a sensory trace that is rapidly overwritten by subsequent visual input. Some contemporary sensory-recruitment accounts of visual working memory (VWM) suggest that VWM may depend on a sensory code as well - e.g., early perceptual areas may be recruited to maintain activity related to a remembered item. Here we investigated interference effects on Iconic and VWM using matched orientation-discrimination tasks with and without dynamic spatially overlapping visual masks. In both experiments, participants judged the direction of a small orientation change between a sample and test grating. Critically, the very same masks were used across experiments, allowing us to isolate the level at which interference occurs for each. Fitting psychometric functions to the data revealed that Iconic and VWM exhibit qualitatively different patterns of interference. Whereas observers' precision and guess rate in the VWM task were robust to the masking manipulation, performance in the Iconic memory task was substantially impacted. These divergent signatures demonstrate that the two forms of memory rely on distinct representational formats: one that is overwritten by incoming sensory input, and one that is not. Together, these results offer a direct behavioral dissociation between Iconic and VWM, advancing a representational account in which Iconic memory is genuinely sensory while VWM operates at a more insulated, non-sensory level of the visual system.

23.322 TO SHIELD OR NOT TO SHIELD: EFFECTS OF SALIENCE IN WHOLE-REPORT VISUAL WORKING MEMORY TASKS

Martin Constant¹ (martin.constant@unige.ch), Dirk Kerzel¹; ¹University of Geneva

We follow up our recent work that demonstrated the massive impact of salience on visual working memory performance (Constant & Liesefeld, 2021; <https://doi.org/gjk9jh>) and that this robust effect explains most of the performance differences, even when contradictory top-down information or cues are employed (Constant &

Liesefeld, 2023; <https://doi.org/gr6xzx>; Constant & Kerzel, 2024; <https://doi.org/mvnx>). In the present series of experiments, we re-employed this design, asking participants to encode three tilted bars (12°, 28° and 45°) presented among 33 vertical filler-bars for 350 ms. Following a 1000 ms retention interval, the response display appeared. Here, the key difference with the previous research is that participants had to recall the color of all three targets in a seemingly random order (i.e., whole-report procedure) instead of reporting the color of a single target. We wanted to examine whether salience could prevent or reduce the performance cost affecting targets that are recalled later. We hypothesized that this could occur through the storage of the most salient target in a special active state in visual working memory. Contrary to our expectations, the order effects' performance cost was comparable across all levels of salience, and therefore more salient targets were not shielded. We replicated this finding in a second experiment, and show that this is not prevented when targets of a given display share the same tilt. Finally, in a third experiment, we used a spatial pre-cue to indicate which of the three targets would be recalled last in the sequence. Preliminary results indicate that this advance information on the recall order does protect the cued target to some extent. However, this benefit appears to come with a performance cost for earlier targets, suggesting a performance trade-off.

23.323 DOES THE MAGNITUDE OF ATTRACTIVE BIAS IN VISUAL WORKING MEMORY DEPEND ON THE STRENGTH OF THE DISTRACTOR TRACE?

Ryan Miller¹ (ryanm@mun.ca), Blaire Dube¹; ¹Memorial University of Newfoundland

Representations in visual working memory (VWM) are susceptible to distortion from perceptual distractors. When a distractor shares a feature dimension with an item in VWM, the distortion is systematic: the remembered feature shifts toward the feature of the distractor. This attractive bias is often attributed to direct interference with the VWM representation at the moment the distractor is processed, but it has recently been suggested that the bias may instead reflect a lingering distractor trace available at the time of report. Notably, most experiments demonstrating attractive bias probe memory after a moderate, fixed post-distractor delay (typically ~1000-1500ms), leaving the temporal dynamics of the effect underexamined. Here, we tested whether the strength of the distractor trace impacts the magnitude of attractive bias by manipulating this post-distractor delay. Participants encoded the orientation of a randomly rotated bar. After a fixed initial memory delay, a to-be-ignored distractor bar appeared, rotated +/- 15 degrees relative to the target. The subsequent post-distractor delay was either 500ms, 1500ms, or 3000ms (randomly intermixed) before participants reported the orientation of the memory target. If attractive bias is driven by a lingering distractor trace at report, the longest delay should yield the weakest bias. While we observed robust attractive bias across all delay conditions, there was no effect of post-distractor delay. These results suggest that attractive bias

stabilizes by at least 500ms post-distractor and is not meaningfully influenced by residual distractor trace strength beyond this early period.

Natural Sciences and Engineering Research Council of Canada
Discovery Grant

23.324 POST-PERCEPTUAL DECISION DETERMINES THE DIRECTION OF SERIAL DEPENDENCE IN VISUAL PERCEPTION

Kuo-Wei Chen¹ (kchen172@asu.edu), Juhyun Ha, Ling-Jing Liu, Scott Janetsky, Gi-Yeul Bae; ¹Arizona State University

Achieving perceptual stability over time in the face of noisy visual inputs is a fundamental goal of visual information processing. Studies have proposed that the visual system achieves this goal by biasing perception of a new stimulus toward the past stimulus (i.e., attractive serial dependence). However, the visual system must also maintain perceptual sensitivity over time to detect differences in stimuli encountered at different time points (i.e., repulsive serial dependence). Here, we investigated how the visual system balances these competing demands in the context of motion perception. In two experiments, participants viewed a brief random-dot kinematogram with varying levels of motion coherence and reproduced the perceived motion direction using a computer mouse. We found that direction reports exhibited an attractive serial bias when the motion stimulus was noisier (i.e., lower coherence), and that this attractive bias decreased and eventually reversed into a repulsive bias as the motion stimulus became less noisy (i.e., higher coherence). To understand how these opposing biases emerged, we examined the temporal dynamics underlying the bias effects by analyzing mouse trajectories. We found that, for both attractive and repulsive biases, mouse reports began with a strong repulsive bias away from the prior stimulus, reflecting adaptation to the prior stimulus (or decision), and that this initial repulsion was substantially decreased over the course of the response. These results show that both the repulsive and attractive serial biases are driven by post-perceptual correction mechanism that reduces the strong repulsive bias driven by adaptation, with the direction of the serial bias effect determined by the quality of the stimulus information. Together, our findings suggest that the visual system optimizes perceptual behavior under stimulus uncertainty by balancing stability and sensitivity through a single quantitative decision mechanism, rather than through qualitatively distinct multiple mechanisms proposed by prevailing models of serial dependence.

23.325 REWARD REDUCES BUT DOES NOT ELIMINATE COSTS ASSOCIATED WITH UPDATING VWM

Eva Lout¹, Jarrod Lewis-Peacock¹; ¹University of Texas at Austin

Visual working memory (VWM) supports the storage, updating, and use of goal-relevant information for adaptive behavior. Prior work indicates that updating a memory item may not fully overwrite its original representation, allowing both the old and updated

representations to coexist. We had hypothesized that repeated updates may clutter VWM with old representations that degrade memory performance but instead found that increases in error are driven by the size rather than the number of updates. This suggests that memory degradation stems from imprecision in the updating process itself, rather than interference from older representations. Here, we asked whether incentivization could enhance the precision of updating. Rewards have been shown to improve VWM precision for stored items, but it remains unknown whether this benefit extends to transformed items. Participants completed a VWM task in which a remembered item needed to be mentally rotated a varying number of times (0, 1, or 2) and by different magnitudes (60°, 90°, or 120°). They earned reward points based on their memory precision on the first half (Experiment 1; N=17/30) or the second half (Experiment 2; N=16/30) of trials. We predicted that reward would enhance performance overall and may offset the costs associated with larger updates. Preliminary results show a statistical trend for improved performance with reward in Experiment 2, but not in Experiment 1. However, rewarded performance did not differ between experiments, suggesting the benefit of reward persisted even after it was removed in Experiment 1. Importantly, reward did not eliminate the magnitude-dependent cost of updating. These results suggest that reward can enhance the fidelity of transformation in VWM but cannot remove the costs imposed by larger changes. Thus, the mechanisms underlying VWM manipulation are precision-limited, and may underlie constraints on adaptive behavior in dynamic environments.

23.326 THAT'S OLD NEWS: TIME, NOT REPRESENTATIONAL FIDELITY, SHAPES THE EFFICACY OF ATTENTION IN WORKING MEMORY

Jamal Williams¹, William Narhi-Martinez², Anna Nobre²; ¹University of California, Santa Cruz, ²Yale University

As you enter the grocery store and check your list, you briefly hold several items in mind before deciding which one to grab first. In this instance, while much of the previous work would suggest that attention exerts the same influence over all items currently maintained in visual working memory (VWM), these findings are often based on simultaneous encoding. Yet, in the real world, representations are often encoded sequentially, thereby creating a "selection latency" between when an item is encoded and when it is prioritized by attention. Across several experiments, we demonstrate that this selection latency has a striking influence over the efficacy of attention, even when all representations are maintained with equal fidelity prior to being selected. In the current work, participants encode four real-world objects across two sequentially presented arrays: two objects appear in an initial "leading" array with the remaining two items presented in the "trailing" array. On half of all trials, a 100% valid retrocue indicates whether the tested item is drawn from the leading or trailing array; on the remaining "neutral" trials, the test item is selected at random. After a delay, a single test object is presented centrally and participants report its color from memory using a

continuous color wheel. On neutral trials, memory error is effectively identical for items from the leading and trailing arrays. Yet, once an item is cued, we find a clear temporal asymmetry: despite all cued items showing reduced error compared to neutral items, those encoded more recently (ie, trailing items) show a profound benefit of attention and exhibit significantly reduced error compared to items from the leading array. This asymmetry persists across multiple control conditions, indicating that selection latency plays a critical role in shaping the efficacy with which attention operates over VWM representations.

23.327 VISUAL COGNITIVE COST OF SIMPLE(ST) PHYSICAL ACTION

Yoojeong Choo¹, Weiwei Zhang², Weizhen Xie¹; ¹University of Maryland, College Park, ²University of California, Riverside

Simple physical actions, such as sustained handgrip, impose measurable cognitive costs on visual processing, impairing visual attention and working memory (VWM). Yet the mechanisms underlying this action-cognition trade-off are not well understood. Here, we used a change-detection task paired with concurrent handgrip, manipulating physical load (low vs. high) and the presence of task-irrelevant distractors, to examine two plausible impacts of physical exertion on VWM. One possibility is that action and cognition draw on shared resources, such that engaging the motor system reduces one's ability to actively retain visual information. Another possibility is that physical exertion places demands on executive control, particularly the ability to suppress distractors, without directly reducing the retention of task-relevant content. Across two experiments, concurrent physical load impaired VWM behavioral performance, particularly in the presence of distractors. In Experiment 1 (n = 20), EEG results showed that stronger concurrent handgrip force did not reliably reduce the contralateral delay activity (CDA), a neural marker associated with maintained VWM items, providing limited evidence for the retention cost account. Instead, higher physical load was associated with increased CDA amplitude when distractors were present, consistent with greater intrusion of task-irrelevant information into VWM during physical exertion. In Experiment 2 (n = 20), fMRI revealed engagement of bilateral inferior frontal and left posterior parietal cortices – regions previously implicated in both cognitive and motor inhibitory control processes – when physical load and visual distractors co-occurred. Taken together, these findings indicate that the cognitive costs of physical exertion arise not from depletion of a shared representational resource, but potentially from reduced inhibitory control that allows distractors to intrude into VWM. As inhibitory control emerges as a potential bottleneck for visual cognition under physical strain, future research should further elucidate how this bottleneck is implemented within, and potentially modulated through, frontoparietal network dynamics.

23.328 BEYOND ALL-OR-NONE: PROBABILISTIC VISUAL WORKING MEMORY PRIORITIZATION ENHANCES

MEMORY PERFORMANCE WITHOUT MODULATING N2PC OR ATTENTIONAL CAPTURE

Alexandre Fortuna¹ (alexandre.fortuna@unige.ch), Martin Constant¹, Dirk Kerzel¹; ¹University of Geneva

Attentional templates in visual working memory help guide selection during visual search, and retro-cues can change how much priority each stored item receives. This study used electroencephalography (EEG) to investigate how probabilistic retro-cues influence the dynamic interplay between visual working memory and attentional selection, hypothesizing that cue reliability induces graded, continuous adjustments in visual working memory prioritization rather than all-or-none shifts. Participants first memorized three colors. Following memorization, each color was associated with a retro-cue that signaled one of three probabilities (60%, 30%, or 10%) of that color being tested in the subsequent memory task. After a delay, participants completed a visual search task followed by a memory discrimination task. In the search task, participants located a target dot among two lateralized disks, where the distractor disk was sometimes colored to match one of the probabilistically cued VWM items, allowing measurement of attentional capture. The results did not reveal differences in N2pc amplitude across reliability conditions during visual search, and reaction times likewise did not vary as a function of cue reliability. However, memory discrimination performance improved gradually, with items linked to more reliable cues being judged more accurately and quickly. This demonstrates a graded distribution of memory priority rather than categorical differences. Overall, these findings indicate that probabilistic retro-cues adjust the prioritization of items in visual working memory, as reflected in graded differences in memory discrimination performance. However, this prioritization did not translate into measurable differences in attentional capture or N2pc activity during the search phase, suggesting that probabilistic weighting influenced memory retrieval more strongly than online attentional selection.

This work is supported by the Swiss National Science Foundation (SNSF) with the grand number 219517

23.329 DIRECTED FORGETTING AND INCIDENTAL ENCODING IN VISUAL WORKING MEMORY

Hana Yabuki¹, Susanne Ferber¹; ¹Carnegie Mellon University

Visual working memory (VWM) allows us to hold on to representations of items that are no longer available to view. However, VWM is limited in capacity, necessitating a selection process that favours task-relevant information. Directed forgetting - i.e., the volitional removal of items from working memory, is thought of as a process that allows us to delete outdated memoranda from working memory and frees up cognitive resources. To understand the nature of these newly-available resources, we examined if they can be reallocated to incidentally encode item-context after being released. We adapted Dames and Oberauer's (2022) WM removal paradigm. Here, participants (n = 63) were sequentially presented with a word

superimposed on a background image. In the removal condition, words were followed by a 'remember' or a 'forget' cue in a clockwise pattern, with each trial consisting of 6 words. At the end of a trial, participants recalled if the probe word at test was presented in the same location during encoding. We employed two baseline conditions, set size 6 and set size 3, where all presented words were followed by a 'remember' cue. Directed forgetting is observed by comparing performance for trials that tested words followed by a 'forget' cue with words followed by a 'remember' cue. We replicated Dames and Oberauer's (2022) finding that memory for 'remember' words was better than for 'forget' words, demonstrating directed forgetting in VWM. Performance for background image recognition, however, was poor, suggesting available resources do not retroactively enhance incidentally encoded context. To explore individual differences in VWM capacity and incidental encoding, further analyses will examine correlation between capacity from a change localization task and memory for background images.

Scene Perception: Models, natural image statistics

SATURDAY, MAY 16, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY

23.330 PROBING SCENE UNDERSTANDING IN LONG-TERM MEMORY USING GENERATIVE MODELS

Ritik Raina¹ (ritik.raina@stonybrook.edu), Abe Leite¹, Alexandros Graikos¹, Seoyoung Ahn², Gregory Zelinsky¹; ¹Stony Brook University, ²UC Berkeley

What information constitutes the long-term memory (LTM) representation of a viewed scene? We use our Seen2Scene framework to identify metamers in LTM—generated scenes that are confused with previously-viewed images. Seen2Scene is a latent diffusion model that generates scenes from sparse inputs: fixation tokens (DINOv3 patches at fixated locations), a gist-like peripheral representation based on low-resolution information, and optionally text descriptions. For combined visual-text generations, custom attention processors adaptively weight contributions based on visual feature confidence—text features fill in where visual information is ambiguous. Seen2Scene was integrated into a behavioral experiment spanning two sessions. In Session 1, participants freely-viewed scenes for a variable number of fixations (1, 5, or 10), and then immediately provided a detailed verbal description via a microphone. After a minimum 1-day interval, participants returned for an old/new recognition task where test images included: (1) original scenes, (2) novel scenes not viewed in Session 1, (3) generations based on visual information alone, (4) generations based on verbal descriptions alone, or (5) generations combining both visual information and verbal descriptions. Vision-based generations produced the highest LTM metamerism rates, with the effect increasing with the number of

fixations during Session 1 viewing. Verbal-only generations had low metamerism rates, performing only slightly above novel scenes. Adding verbal descriptions to vision-based generations did not increase metamerism rates. This implies that visual input during encoding, as captured by vision-based generations, primarily shapes our LTM scene representations, while verbal encoding contributes minimally. The retention interval revealed a striking crossover: metamerism from vision-based generations decreased with longer intervals, while metamerism from verbal-only generations increased. This pattern indicates that LTM representations initially reflect visual details, but may shift toward semantic content as perceptual details fade. Our findings reveal the multifaceted nature of LTM and demonstrate Seen2Scene as a powerful tool for probing scene memory representations.

RR and GJZ are supported by NSF-CompCog #2444540 to GJZ. AG is supported by NSF-IIS-2212046. AL is supported by NSF-GRFP #2234683 and NIH-NEI R01EY030669 to GJZ.

23.331 NEURAL RESPONSES TO NATURAL VERSUS AI-GENERATED AFFECTIVE IMAGES

Yujun Chen¹, Ruogu Fang¹, Andreas Keil², Mingzhou Ding¹; ¹Biomedical Engineering, University of Florida, ²Psychology, University of Florida

The International Affective Picture System (IAPS) contains 1,182 well-characterized photographs depicting natural scenes varying in affective content. These pictures are used extensively in affective neuroscience to investigate the neural correlates of emotional processing. Recently, in an effort to augment this dataset, we have begun to generate synthetic emotional images by combining IAPS pictures and diffusion-based AI models. The goal of this study is to compare the neural responses to IAPS pictures and matching AI-generated images. The stimulus set consisted of 60 IAPS pictures (20 pleasant, 20 neutral, 20 unpleasant) and 60 matching AI-generated images (20 pleasant, 20 neutral, 20 unpleasant). In a recording session, a total of 30 IAPS pictures and 30 matching AI-generated images were presented in random order, where each image was displayed for 3 seconds with neighboring images separated by an interval of 2.8 to 3.5 seconds. Each experiment consisted of 10 recording sessions spread over two days. The fMRI data was recorded on a 3T Siemens Prisma scanner. Pupil responses to image presentation were monitored using an MRI-compatible Eyelink1000 Eyetracker. Our preliminary analysis of the fMRI data (N=28) showed that (1) both the IAPS pictures and the matching AI-generated images activated higher-order visual areas such as OTJ and components of the anterior emotion network such as the insula and (2) in the retinotopic visual cortex including V1, MVP (Multivariate Pattern Analysis) classifiers built to decode emotional categories from voxel patterns activated by IAPS pictures could be used to decode emotional categories from voxel patterns activated by AI-generated images and vice versa, suggesting that the IAPS pictures and the matching AI images evoked similar neural responses in the brain. Efforts to confirm

these findings are underway by recruiting additional participants. Analysis is also being expanded to include the comparison of such measures as functional connectivity and pupillometry.

MH125615 and MH112558 and NSF grant BCS2318984

23.332 CONTEXT EFFECTS IN SIMILARITY JUDGMENTS IMPACT AI-HUMAN VISUAL REPRESENTATION ALIGNMENT

Eben Daggett¹ (edaggett@nmsu.edu), Giovanna Del Sordo¹, Jon Art¹, Michael Hout¹; ¹New Mexico State University

Recent studies have demonstrated the potential for improving Artificial Intelligence performance and alignment when utilizing human visual similarity judgements as training data. Computer vision models trained with human similarity data have exhibited visual representations of images that are better aligned with human visual representations than traditionally trained classifiers. Such models also show improved generalization abilities and, when errors are made, fail in a more predictable and less nonsensical manner. This finding has important implications in applications – e.g., the medical industry – where large, or nonsensical errors can be more deleterious than small errors. While such an approach to AI shows promise, it may be built on an unstable foundation, as the past 70 years of cognitive psychology research has demonstrated the construct of similarity to be highly flexible and difficult to quantify. In the present study, we trained CV models on human similarity judgments gathered for sets of medical pathology images and everyday household items, where the context of the similarity rating task was manipulated to induce context effects in the judgements. CV model performance in different conditions was analyzed to investigate how the flexible nature of visual similarity perception impacts an AI's ability to effectively model human similarity data and generalize to exemplars not previously encountered by the AI. Results showed wide differentials in CV model performance between context conditions, but also demonstrated that model performance – in isolation – may not be an appropriate metric for assessing representational alignment. The study was intended as an initial examination into how similarity may play a better role in the improvement of AI visual classifiers for high-stakes applications. Recommendations are provided for future research on human-AI representational alignment so that the field may better operationalize its study via the inclusion of further research into the dynamics of human similarity perception.

Institute for Applied Practice in AI and Machine Learning - New Mexico State University

23.333 VISUAL FEATURES PREDICT CONTINUOUS AESTHETIC PREFERENCES IN NATURALISTIC AND ARTISTIC MOVIES

Mustafa Alperen Ekinci^{1,2}, Daniel Kaiser^{1,2}; ¹Neural Computation Group, Justus Liebig University of Giessen, Germany, ²Center for

Mind, Brain and Behavior (CMBB), Universities of Marburg, Giessen, and Darmstadt, Germany

In real-world situations, our aesthetic experiences are shaped by exposure to constantly changing visual scenes. Yet, most research on visual aesthetics has relied on static or simplified displays, leaving it uncertain how aesthetic experiences form and evolve over time in naturalistic settings. Here, we examined how the visual characteristics of continuous naturalistic and artistic movies influence aesthetic experiences as they unfold. In two separate experiments, participants watched either the documentary “Home”, presenting a broad range of real-life visual scenes, or the animated drama “Loving Vincent”, based on Van Gogh’s paintings and his life story. The two movies differ vastly in terms of style and context: while “Home” features a range of photorealistic natural scenes, “Loving Vincent” features artistic animations akin to dynamic Van Gogh paintings. During the movies, participants continuously rated aesthetic appeal by adjusting a response slider. To predict these moment-to-moment ratings, we trained linear regression models on image-computable visual predictors extracted from every movie frame. Predictors included low-level and higher-order properties, including the efficiency of coding in a deep neural network (DNN) model, color statistics, symmetry, and motion energy. For both movies, the trained models successfully predicted aesthetic ratings for held-out movie parts. Predictions were robust within individuals and across individual observers, suggesting that perceptual features similarly shape aesthetic experience across people. Reduced model variants revealed that DNN coding efficiency and color statistics contributed most strongly to the successful predictions. Interestingly, color similarity also allowed for successful predictions across the naturalistic (“Home”) and artistic (“Loving Vincent”) movies, highlighting that color preferences prevail across vastly different visual inputs. Our findings demonstrate that visual features reliably predict aesthetic experiences even for complex and dynamic stimuli.

Supported by the Deutsche Forschungsgemeinschaft (DFG), KA4683/6-1 (project no.~536053998); and under Germany’s Excellence Strategy (EXC 3066/1, “The Adaptive Mind”, project no.~533717223).

23.334 EYE MOVEMENTS WHITEN TEMPORAL POWER SPECTRA DURING REAL-WORLD ACTIVITIES

Samantha A Montoya¹
(samantha.montoya@pennmedicine.upenn.edu), Zachary J Kelly¹, Sophia Mirabal¹, Ruby S Bouhassira¹, Carlyn Patterson Gentile², Alan A Stocker¹, Geoffrey K Aguirre¹; ¹University of Pennsylvania, ²Children’s Hospital of Philadelphia

In contrast to spatial vision, few studies have examined the temporal statistics of natural vision, especially having accounted for eye movements. We measured temporal variation in natural, gaze-centered vision, extending to the nearly full operational range of human temporal sensitivity. Five participants performed a set of 10

scripted, 4 minute long, real-world activities (ambulatory and seated, indoor and outdoor) while wearing custom-built spectacles equipped with a high-speed camera (120 FPS, 150° FOV) and synchronous eyetracking. Here, we discuss temporal statistics for walking indoors and working at a desktop computer. Videos were transformed to the perspective of the observer's visual field prior to calculating temporal spectral power densities (SPDs) of luminance contrast over 5s intervals with 3° spatial resolution. Analyses were conducted with and without correction for eye movements by performing virtual foveation. We summarized the data as the alpha exponent of a $1/f^\alpha$ model fit to the SPD, along with the total temporal contrast variance. In analyses omitting adjustment for eye movements, temporal SPDs were well described by $1/f^\alpha$, with exponents close to 2. With virtual foveation, we observed a whitening of the power spectrum at higher temporal frequencies. For walking indoors, during which tracking eye movements are prevalent, SPD whitening occurred above ~35 Hz. While working at a computer, the whitening began at ~20 Hz. Accounting for eye movements also introduced spatial structure to the pattern of temporal variation, with systematic variation between central and peripheral field locations. The particular properties of temporal variation in natural vision are important for efficient coding models, and studies of the features of stimuli that produce visual discomfort. We find that temporal variation systematically departs from an assumed $1/f^2$ distribution in the presence of eye movements, and that these effects vary with behavior and visual field location.

Funding was provided by NIH 1R01EY036255

23.335 INTERSUBJECT NEURAL SYNCHRONY DURING DYNAMIC EMOTION PERCEPTION PREDICTS INDIVIDUAL DIFFERENCES IN BEHAVIORAL MEASURES: AN fNIRS STUDY

Patrick Kelly^{1,2}, Jefferson Ortega¹, Keanan Joyner¹, Silvia Bunge¹, David Whitney¹; ¹UC Berkeley, ²UCLA

Naturalistic emotional videos elicit shared patterns of brain activity, but it is unclear what this synchrony reflects: merely shared attention to the same sensory input, or shared internal computations about a character's evolving emotional state. Here we test whether inter-subject synchronization in fronto-temporo-parietal networks indexes shared emotional meaning—whether people whose brains are more aligned with the group are also more accurate at tracking a character's moment-to-moment emotional state. We used functional near-infrared spectroscopy (fNIRS) during a dynamic emotion perception task, Inferential Emotion Tracking (Chen & Whitney, 2019). Participants (N = 33) continuously rated the valence and arousal of a target character in 28 clips while fNIRS measured oxygenated hemoglobin in bilateral prefrontal (DLPFC, VLPFC, OFC, RLPFC) and temporo-parietal regions (STS, TPJ). We computed region-of-interest (ROI) inter-subject correlations (ISCs) by correlating each participant's activation pattern with the leave-one-out group average. All twelve ROIs showed significant positive ISC (mean $r \approx .16-.26$, $p < 0.001$). Using inter-subject representational similarity analysis (IS-RSA), we then asked

whether ISC strength predicted emotion-tracking performance. Overall ISC robustly predicted Inferential Emotion Tracking accuracy: high performers' brain activity converged on similar patterns, whereas low performers' activity was more idiosyncratic (Spearman $r = .47$, 95% CI [.20, .60], permutation $p < 0.001$). Finally, penalized regression identified which regions carried these behaviorally relevant shared representations. ISC in bilateral VLPFC, left TPJ, left STS, and left OFC showed the largest weights, with additional contributions from right OFC and right TPJ. Effects were strongly lateralized: higher ISC in right frontal and left temporoparietal cortex was associated with better performance, whereas higher ISC in left frontal and right temporoparietal cortex was associated with worse performance. fNIRS ISC thus provides a sensitive, lateralized marker of shared emotional meaning during naturalistic emotion perception.

23.336 GAZE-CENTERED VS. HEAD-CENTERED IMAGE STATISTICS DURING NATURALISTIC HUMAN BEHAVIOR

Anna Shafer-Skelton¹, Mary Hayhoe¹; ¹University of Texas at Austin

To understand the visual system, it is critical to understand its inputs. Early visual regions are sensitive to orientation and spatial frequency information (crisp and blurry oriented edges), and classic work has quantified this information with Fourier spectra in both hand-selected natural-scene photographs and image sets structured by scene category. While some work has analyzed the information present at fixation during several naturalistic tasks, the release of large data sets has only recently made it possible to capture gaze-centered image statistics across the wider visual field. Here, we analyzed stills taken from the Visual Experience Dataset (VEDB; Greene et al., 2024), which recorded eye-tracking data alongside wide-angle egocentric video as participants engaged in a broad range of real-world activities. We captured 8,781 stills and analyzed the gaze-centered vs. head-centered Fourier spectra of each, across both a smaller and larger field of view. We summarized Fourier information using bins of four orientations and two spatial frequencies. Initial analyses show stronger cardinal information in head-centered vs. gaze-centered stills, especially at lower spatial frequencies and for the larger field of view. Ongoing analyses investigate how tasks, settings, and choice of image statistics influence gaze-centered vs. head-centered differences. This work demonstrates an important step towards understanding how our real-world visual inputs interact with the retinotopic (gaze-centered) visual system, as well as the impact of active sampling behavior on these inputs.

This work was supported by an NEI postdoctoral NRSA fellowship to AS (5F32EY036266-02)

Scene Perception: Intuitive physics

**SATURDAY, MAY 16, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY**

23.337 WHEN “IN” OVERRIDES “NEAR”: PERCEIVING CONTAINMENT INTERFERES WITH PROXIMITY JUDGMENTS

Alice Wang¹, Alon Hafri¹; ¹University of Delaware

When a ball is in a box, is it also near the box? Intuition says no: “in-ness” seems to override “near-ness,” as if certain spatial relations cannot coexist. What is the origin of such restrictions? Are they merely constraints on how scenes are described linguistically, or do they reflect a deeper property of how the mind represents spatial relations? As a case study, we examined proximity. Although seemingly a simple metric computation, proximity can also be construed in terms of the categorical near relation, and may therefore be subject to similar constraints as the linguistic use of “near.” We predicted that perceiving containment would interfere with proximity judgments, even when language was task-irrelevant. Across three experiments, participants saw displays with two blue circles and one or more edges and judged which circle was nearer to a blue target edge. In Experiment 1, the circles appeared on opposite sides of a single edge. On containment (IN) trials, the blue edge formed the side of either one or two gray squares: on one-IN trials, one circle was contained by a square, while the other remained outside; on two-IN trials, each circle was contained by its own square. Some trials involved no containment. Despite being directed to ignore non-blue elements, participants were slower and less accurate whenever containment was present. Crucially, one-IN and two-IN trials were equally difficult, ruling out effects of stimulus complexity and object-based warping. Experiment 2 replicated the IN disadvantage using two side-by-side displays, each with its own square and target edge. Experiment 3 addressed an “edge-selection” confound by placing non-IN circles between two edges (one being the target); the IN disadvantage, though reduced, persisted. These results suggest that perceiving containment interferes with proximity computations and reveal that perceptual and linguistic construals of space may be governed by similar constraints.

23.338 FROM EGOCENTRIC VIDEO TO PATCHWORK PHYSICS: BUILDING EMPIRICAL PRIORS FROM EVERYDAY EXPERIENCE

Abdul-Rahim Deeb¹, Jason Fischer², Leyla Isik¹; ¹Johns Hopkins University, ²University of Coimbra

Humans can anticipate how objects and materials will behave as they move, collide, or change state, and use these expectations to guide behavior. This capacity of intuitive physics is embedded in perception and action: recent lab-based studies show that physical regularities can act as predictive cues that are combined with incoming sensory information to support successful interaction with the environment. This raises a central question: how does perception internalize these regularities? We recently proposed a computational framework, the Patchwork Approach, that answers this by treating intuitive physics as a multivariate perceptual manifold shaped by environmental statistics, rather than as explicit simulation over symbolic physical states.

Dynamic events are located in this space by behaviorally relevant dimensions (e.g., object velocity, path geometry, apparent volume), and predictions are implemented as local computations over this structure. A remaining open question however, is what experiences shape this manifold and how can we measure them? Here we treat large-scale egocentric video as a proxy for lived experience. Datasets such as HOI4D and EGO4D use head-mounted cameras to capture thousands of everyday actions from the observer’s viewpoint. We developed a novel pipeline combining hand-object segmentation, temporal tracking, and monocular depth estimation, to extract physical variables from any egocentric video, including 3D trajectories, instantaneous velocities, path curvature, contact onset and duration, and object masks and volume. These features define the Patchwork manifold and allow us to estimate local physical predictions. We find that video-derived slopes relating apparent mass to speed, and those linking angular launch kinematics to object trajectories, reproduce observers’ biases from our prior lab-based experiments, despite using no fitted free parameters. This methodology provides a concrete, image-computable route to estimating perceptual priors from natural interaction and opens the door to probing how such priors emerge across datasets, domains, and developmental stages in vision science.

NIH R01MH132826

23.339 RATING OR COMPARING? DO PAIRWISE METHODS OFFER ADVANTAGES OVER LIKERT SCALES FOR MEASURING LATENT PERCEPTUAL DIMENSIONS?

William Gaudreau¹ (williamj.gaudreau@gmail.com), Kimele Persaud², Omer Daglar Tanrikulu¹; ¹University of New Hampshire, ²Rutgers University - Newark

Likert scales are the most common format for rating data, but they have limitations: over many trials, participants may rescale their internal standard to the stimulus set, recall earlier responses to stay consistent, and cluster responses on a few integers. Two-alternative forced-choice (2AFC) tasks sidestep these issues because observers compare two stimuli at a time and need not maintain a global rating scale. Building on this idea, Clark et al. (2018) proposed using mElo scores to place stimuli on a latent scale derived from 2AFC data. We asked whether these advantages extend to a latent dimension without ground truth (perceptual surprise) and whether mElo scores are more sensitive than mean Likert ratings. Thirty-six participants viewed 58 videos of everyday objects falling; some outcomes violated material expectations, others behaved normally. In one block, participants rated how expected each outcome was on a Likert scale; in another, they chose the more expected outcome in 2AFC trials. We then computed mean Likert ratings and mElo scores for each video. Contrary to our predictions, mElo values were not more uniformly spread than mean Likert ratings, and split-sample correlations (25%, 50%, 75% of observers) indicated similar reliability. The two measures were strongly correlated across videos and both distinguished expected from unexpected outcomes, with mElo yielding only a

modestly larger effect size for this contrast. Together, these results indicate that observers' visual expectations for dynamic material events can be captured reliably with both Likert ratings and 2AFC-based mElo scores. For the kinds of naturalistic stimuli commonly used in vision science, pairwise methods offer only modest gains in sensitivity, but they provide a viable alternative for building continuous perceptual scales on latent dimensions such as surprise, animacy, or realism, which are central to contemporary theories of predictive vision but lack an objective ground truth.

23.340 DOES PHYSICAL STABILITY FACILITATE VISUAL DETECTION?

Harini Sankar¹ (hsankar3@illinois.edu), Diane M. Beck¹, Stefan Uddenberg¹; ¹University of Illinois Urbana-Champaign

A central challenge for vision is not just to recover what is where, but to anticipate what is about to happen next, such as whether a tower of blocks will remain standing or collapse. Studies of intuitive physics have shown that observers can rapidly and accurately judge the stability of such towers from a single glance. But this work typically begins only after the scene has already been perceived. Here we asked whether the physical structure of a scene can shape this earliest stage of processing: how quickly the scene becomes visible at all. We first created random block towers that, as a whole, could either be stable or unstable, as determined by a physics engine. We also created images of scrambled blocks, where the blocks' positions were allowed to vary freely, up to and including floating in free space and clipping into one another. On each trial of the experiment, observers were briefly shown a single image, which could either be an intact tower (stable/unstable), or scrambled blocks, followed by a mask. Observers were tasked with discriminating between the intact and scrambled stimuli. Stimulus duration varied adaptively using a one-up two-down staircase procedure. Prior work on visual detection suggests that images which are more statistically regular, and which better match the visual system's expectations, reach awareness more readily than those that violate those expectations. Thus, if the visual system encodes physical stability and embodies a prior favoring stable configuration, then stable towers should be detectable at shorter durations than otherwise similar unstable towers. Preliminary results indicate that the stable towers were not detected at shorter presentation durations than unstable towers, indicating that a scene's physical stability may not affect the very earliest stages of scene processing. However, more data using other masks is needed to verify this finding.

23.341 A BENCHMARK DATASET FOR PERCEIVING OBJECT SPATIAL RELATIONSHIPS IN HUMAN AND MACHINE VISION

Georgina Woo¹, Nancy Kanwisher¹, RT Pramod¹; ¹MIT

Consistent with their importance for scene understanding and prediction, humans perceive physical relationships between objects like support, containment, and attachment, automatically and

abstractly across changes in object shape, viewpoints and surface properties (Hafri et al., 2024). How do we do this? Here, we introduce a new benchmark dataset designed to test various computational models as hypotheses for how humans perceive object relationships. The dataset contains images representing five spatial relationships – support, contain, attach, occlude, and cover; each instantiated across variations in object identities, materials, viewpoints, and backgrounds. We measured human behavior (N = 40) on a matching task wherein participants chose between two alternatives ('target' and 'lure') depicting the same object spatial relationship as the 'sample' image. In all, we collected behavioral data on 320 trials of this matching task where the sample image additionally differed from the target and lure images with respect to object shapes, viewpoints, and backgrounds. As expected, humans are highly reliable and accurate (accuracy: mean $\pm$ std = 97% $\pm$ 1.3%) in their judgments. We then evaluated various pre-trained models on the same task, including CNNs, vision transformers, and next-frame-prediction systems such as V-JEPA as candidate computational hypotheses for human perception of object spatial relationships. All tested models perform the task poorly, with the best model (ViT-B, trained on ImageNet) reaching an accuracy of only 67.5% (where chance is 50%). In ongoing work, we are: i) expanding the benchmark to include dynamic video stimuli and ii) testing whether end-to-end training specifically on object relationships will suffice, or whether these tasks can only be performed via explicit modeling of the 3D structure of the scene.

NSF STC Award Number 2124136 to NK

23.342 COMPUTATIONAL MODELS REVEAL INTUITIVE PHYSICS AND STATISTICAL CUES SEPARATELY CONTRIBUTE TO THE VISUAL PERCEPTION OF LIQUIDS

Yuting Zhang¹ (yuting.zhang@yale.edu), Wenyan Bi¹, Yuyang Miao², Ilker Yildirim¹; ¹Yale University, ²Columbia University

We are intimately familiar with liquids in our visual experience, yet the computational basis of liquid perception remains underexplored. This is an important knowledge gap because liquids, with their mutable shapes and complex intrinsic dynamics, differ remarkably from the commonly studied categories in computational vision, such as rigid objects or non-rigid solids. To understand the computational basis of liquid perception, we implemented different models of this ability and tested them in a new behavioral study. The models realize two distinct theoretical possibilities for the visual perception of liquid viscosity. The first possibility, and the focus of most existing work, explains the representation of liquid viscosity as a consequence of high-level image and motion statistics discriminative of the gradations of this physical property. A second, much different possibility is that the perceptual representations of liquids functionally map the physical processes of how viscosity and external forces (e.g., gravity, rigid surfaces) shape the way liquids move. We task these models and humans in a new behavioral task: making similarity judgments of liquid viscosity across pairs of animations depicting qualitatively different scenarios --- e.g., a metal ball falling into a liquid container vs. liquid pouring over a non-

flat surface. We find that a new model called Ripple, which builds and manipulates physics-based representations of liquid viscosity from sensory inputs, explains substantial variance in human judgments, beyond powerful, previously behaviorally validated, statistical representations in a deep neural network trained to regress viscosity from animations. Surprisingly, these statistical representations of viscosity remain alike across vastly different deep neural architectures and training datasets, but different from Ripple's physics-based representations. These results suggest that liquid perception extends beyond image statistics to also involve simulation-based intuitive physics.

23.343 CRITICALLY EVALUATING PHYSICAL REASONING IN COMPUTATIONAL MODELS

Grace Hu¹ (hugrace@mit.edu), Nancy Kanwisher¹, RT Pramod¹; ¹MIT

Humans perceive the physical structure of the world, predict what will happen next, and plan successful interactions. What computations underlie such intuitive physical reasoning, and can physical reasoning be learned from the statistics of visual input alone? Recent self-supervised video models such as V-JEPA have claimed to understand physics by merely learning to predict the future in naturalistic videos, suggesting that this is in principle possible for humans. Like humans, V-JEPA is 'surprised' by videos showing physically implausible events. However, previously tested violations – for example, object 'teleportation' – are also deviations from natural low-level visual motion statistics, providing an alternative explanation to V-JEPA's surprise. Thus, to unconfound low-level visual motion statistics from physical implausibility, we generated transformed versions of two synthetic physical scene understanding video benchmarks, GRASP (Jassim et al., 2023) and IntPhys (Riochet et al., 2020), by applying spatial (left-right and up-down flips) and temporal (time-reversal) transforms. These transforms all preserve motion statistics but differentially affect the underlying physics: left-right flips always maintain plausibility, while up-down flips and temporal-order reversals remove plausibility in videos involving gravity and/or momentum. On the transformed GRASP dataset, V-JEPA is not surprised by left-right flipped videos but is surprised by up-down flipped and temporal-order reversed videos, as expected if V-JEPA understands gravity and inertia. However, on the transformed IntPhys dataset, V-JEPA is surprised by left-right flipped instead of up-down flipped or temporal-order reversed videos, challenging earlier claims of the model's physical scene understanding. In ongoing work, we are expanding our benchmark to make fine-grained comparisons between models and humans and testing it on other computational models. Overall, our results temper previously made claims that fundamental physical scene properties can be learned just from observing natural statistics, and leave open the question of what computations underlie flexible physical reasoning in humans.

NSF STC Award Number 2124136 to NK

Undergraduate Just-In-Time 1

SATURDAY, MAY 16, 8:30 AM – 12:30 PM,

BANYAN BREEZEWAY

23.344 WHEN THERE'S NO WAY OUT: PERCEIVING 'NO ESCAPE' TRIGGERS MORE CAUTIOUS BEHAVIOUR

Alexander Guardado¹, Joan Danielle K. Ongchoco¹; ¹University of British Columbia

Feeling like we cannot escape a situation can lead us to freeze or make decisions we would not otherwise make. Where does this feeling come from? Sometimes, we can pin it down, perhaps from beliefs of helplessness and loss of control. But other times, we can feel it—the way visual illusions might be difficult to resist, despite all beliefs to the contrary. Here, we explored how subtle visual features of our immediate environment might contribute to a sense of "no escape". Across multiple trials, subjects moved through unique virtual rooms, each time also completing a risk-taking task (i.e., the Balloon Analogue Risk Task). Subjects were incentivized to accumulate as much "money" as possible by inflating a virtual balloon, with each pump corresponding to a monetary increment. Subjects could pump the balloon as many times as they wanted, at the risk of it popping, in which case they would lose all the money earned in that trial. Critically, and in a way that was completely irrelevant to the task, rooms were rendered in pairs, where each pair's spatial configuration was identical, except for the location of one spatial element that either created or blocked a path to the exit. Across experiments, despite subjects reporting no awareness of the key manipulation, results revealed a higher explosion rate when the rooms afforded an exit, suggesting that seeing an escape route encourages risk-taking. This influence can also be appreciated in subjects' moment-to-moment decisions, with longer inter-pump latencies in no-escape rooms, suggesting that entrapment instead fuels more cautious behaviour. These effects held even when subjects were exposed to the room layout for only 300ms. These findings suggest that escape routes may be spontaneously and quickly extracted in visual processing—and they demonstrate how visual entrapment can shape risk-taking decisions and behaviour in a surprisingly direct way.

23.345 JUST IN TIME! TRANSLATING TEMPORAL INTERVALS ACROSS MODALITIES

Lily Hallenbeck¹, Jamal Williams²; ¹UCSC

Perceiving and utilizing temporal intervals is critical for everyday life. Crucially, these abilities rely on multiple senses working together. Despite the multimodal nature of real-world time perception, we have a limited understanding of how we maintain and transfer time between modalities. This project addresses a fundamental and unresolved question: Can a duration learned through one sense (e.g., audition) be used to guide behavior in another sense (e.g., vision)? To answer this, participants encode a short, target-duration in working memory for a

later test. Target-durations are filled intervals that are presented as either a colored rectangle on a black background (visual), a continuous, complex tone (auditory), or both (audiovisual). While maintaining this duration, participants detect small bursts of coherent motion in a visual field of randomly moving dots. Importantly, the timing of these coherent motion bursts occasionally coincides with the interval in memory: dots jitter for the target-duration, followed immediately by coherent motion. In this task, performance is worse when motion bursts do not align with visually encoded target-durations, but not via auditory nor audiovisual intervals. These results suggest that temporal intervals guide behavior within, but perhaps not across, modalities. To further compare the quality of temporal-durations, the recognition memory task can be presented in the original modality (e.g., audition-audition) or across-modalities (e.g., audition-vision). Within modality, we find that memory performance is best for auditory intervals, worst for visual, and suffers only limited costs when transferring from vision to audition. In contrast, when an auditory interval is tested using a visual probe, performance is near chance. In tandem, these findings suggest that auditory memories are best remembered but that visual intervals are best for guiding behavior. These findings help inform competing models of temporal perception, including central clock theories and distributed, modality-specific timing models.

23.346 DOES THE FRONTO-PARIETAL PHYSICS NETWORK RESPOND TO AUDITORY STIMULI?

Thomas Brewitt¹ (twbrew@mit.edu), RT Pramod¹, Vivian Paulun², Nancy Kanwisher¹; ¹MIT, ²University of Wisconsin-Madison

Previous studies have identified a fronto-parietal Physics Network (PN) that is preferentially activated when people view or engage in physical reasoning. These studies have used only visual stimuli, but our intuitive understanding of physics is not limited to vision alone: think glass clinking or water splashing. We previously showed that the PN doesn't respond when subjects passively listen to physics-related sounds. Here we tested this finding in PN further when subjects are encouraged to attend to physical properties of our stimuli through an "odd-one-out" task. We gathered short (3-sec) naturalistic audio and video clips of physical and social conditions, each condition having 3 subconditions: collisions, flowing, and spinning for the physical condition, and speech, social interactions, and nonverbal sounds (crying, laughing, etc.) for the social condition. Stimuli were presented in a 2 (modality) x 6 (subcondition) blocked fMRI design, with interleaved fixation baseline blocks, while subjects (N=5) performed an "odd-one-out" task requiring them to identify a non-matching stimulus. The same participants also performed two additional experiments (one visual, one auditory) in which the same stimuli were presented along with blocks of their scrambled versions that were not recognizable but retained low-level features. The independently localized PN showed a stronger response to physical videos than social videos ($p < 0.005$). All auditory conditions produced a PN response near baseline. The PN showed a stronger response to intact

than scrambled physical videos ($p < 0.05$). Remaining conditions responded near baseline. These findings show that the PN doesn't respond preferentially to auditory physical events (compared to auditory social or scrambled physical events, or to baseline) even when the task requires discriminating event types. These findings suggest that PN doesn't engage in multimodal processing of physical information, and that intuitive physical information from auditory cues may be processed elsewhere in the brain.

23.347 LEARNED REGULARITIES OF EMOTIONAL FACE DISTRACTORS SCULPT ATTENTIONAL PRIORITY

Claire Choi¹, Daniel Thayer², Tommy Sprague³; ¹UCSB

Attention is often directed toward salient stimuli because they are behaviorally important. However, salient distractors can be suppressed when they occur predictably (e.g., Luck et al. 2020; Thayer et al. 2022). Notably, the more salient a stimulus is, the more effectively it is suppressed (Stilwell et al., 2023; Stilwell & Anderson, 2025). While this relationship has been demonstrated for basic visual features, it remains unclear whether similar learning mechanisms extend to complex information such as emotion. The present study investigated whether observers can suppress emotional distractors based on learned probabilities and whether this differs across emotions. Participants completed an additional singleton search task where they identified a target face and reported the orientation of a line superimposed on the target. The target was defined as a unique face identity among homogeneous faces. On some trials, an emotional face distractor was present, sharing the same identity but differing in expression (happy, angry, or fearful), which has been shown to effectively capture attention (Yiend, 2010). Three distractor probability conditions were implemented. In the fixed probability condition, the distractor emotion was always the same (e.g., always angry). In the random condition, the distractor had an equal probability for any emotion. In the mixed probability condition, one emotion appeared more frequently as the distractor (65% of trials). Participants were faster to identify the target when a high-probability emotional distractor was present compared to low-probability distractors, indicating learned suppression of emotional regularities. Suppression was emotion-specific: happy faces produced the strongest attentional capture in the absence of regularities and were also most effectively suppressed when regularities were present; fearful faces showed moderate effects; angry faces consistently captured attention regardless of regularity. These results suggest that suppression mechanisms extend to complex emotional stimuli and are influenced by the strength of attentional priority associated with specific emotions.

23.348 SERIAL DEPENDENCE IS A DECISION-LEVEL PROCESS, NOT A PERCEPTUAL ONE

Charles T. Demczuk¹ (demczuk.5@osu.edu), Sam Clarke², Sami R. Yousif¹; ¹The Ohio State University, ²University Of Southern California

The visual input that you receive changes constantly, yet your experience of the world feels relatively stable. How? Some have

suggested that the visual system benefits from stabilization mechanisms like serial dependence, which adaptively biases all input towards the previous input. But at what stage does this supposed stabilization occur? Are these processes truly perceptual in nature? Here, we provide evidence that serial dependence may not be. We tested serial dependence across two domains — number, which has clear perceptual grounding, and coin value, argued to be paradigmatically imperceptible — using four distinct paradigms: sequential forced-choice, simultaneous forced-choice, estimation with and without feedback. Critically, the latter two paradigms inherently involve more complex cognitive processes at the decision stage, whereas the former two may not. For number, we found significant serial dependence across paradigms, with one exception: the simultaneous presentation condition, in which the reference and probe are compared directly. Effect sizes were influenced by task type, with estimation tasks producing (much) stronger effects than forced-choice tasks. For value, the story was strikingly similar — neither forced-choice task produced a significant effect, yet both estimation tasks did. Collectively, these findings provide two reasons to believe that serial dependence is not a perceptual phenomenon. First, if serial dependence were perceptual, it is unclear why the effects would vary so dramatically across paradigms; comparison tasks should yield reliable effects. Second, if serial dependence were perceptual, it is unclear why the effects would occur for a feature that is paradigmatically imperceptible (value). That these effects emerged so consistently across (substantively distinct) domains — and varied so predictably across paradigms — casts doubt on the canonical view that serial dependence is a perceptual phenomenon. This opens the door to redefining what serial dependence is and what it is not.

23.349 LINKING NEURAL REPRESENTATIONS TO BEHAVIOR WITH A UNIFIED MODEL OF HUMAN VISION

Kushal Dudipala¹, Mayukh Deb¹, Ben Lahner², Ratan Murty¹; ¹Georgia Institute of Technology, ²Massachusetts Institute of Technology

A complete account of human vision must explain how visual inputs are transformed into neural population activity and, in turn, into behavior. To do this, we need models that operate on images (like us), build internal representations that correspond directly to measured brain responses (neurons/voxels) which are then linked to produce outputs that match human behavior. Progress has been limited by a prevailing one-level-at-a-time approach, in which models are optimized for neural alignment or behavioral prediction, but rarely both. Here, we introduce NeuroMap, a unified model of the human visual system trained to satisfy two constraints: its internal representations must align with activity in the visual cortex, and its outputs must predict human perceptual decisions. NeuroMap uses a Vision Transformer backbone with internal tokens mapped to visual areas V1, V2, V4, and IT using MOSAIC fMRI data, and an output layer trained on two-alternative forced-choice judgments from DreamSim. This joint objective directly couples neural representation and perceptual behavior within a single model while minimally increasing parameter

count. Unlike standard encoding models, NeuroMap does not need a separate mapper to turn model features into brain predictions. Its internal units are trained from the outset to align to fMRI responses (noise-corrected Pearson ρ on NSD held-out: V1 = 0.59, V2 = 0.66, V4 = 0.68, IT = 0.57). As a result, the model not only recapitulates key aspects of visual cortical organization, but also remains accurate in predicting human behavior, achieving 92.4% accuracy at predicting human 2AFC choices. Because brain-like internal units and behavioral outputs are linked within the same system, NeuroMap opens the avenue for lesioning and micro-stimulating units corresponding to different visual areas and measuring downstream consequences on behavior. These results make NeuroMap a promising prototype for a unified model of vision connecting neural activity to perception.

This work was funded by the NIH Pathway to Independence Award from the NEI (R00EY032603) and a startup grant from Georgia Tech (to NARM)

23.350 RETHINKING VISUAL WORKING MEMORY CAPACITY: PROACTIVE INTERFERENCE CANNOT ACCOUNT FOR REAL-WORLD OBJECT ADVANTAGE

Anantya Engineer¹ (aaengine@ucsc.edu), Yong Hoon Chung², Jamal Williams¹; ¹University of California Santa Cruz, ²Dartmouth College

Traditional theories of visual working memory (VWM) suggest that we have a capacity limit of around 3-4 items. However, people have been found to remember more real-world (RW) objects compared to simple stimuli like colors (Brady et al., 2016). Recent findings challenge the cause of this advantage by showing it can be eliminated when RW stimuli are repeated over multiple trials, suggesting that the RW benefit may instead be due to an imbalance of proactive interference (PI). This suggests that simple stimuli such as colors are subject to more interference from previous trials since many studies showing the RW benefit use unique objects on every trial. Under this account, once PI is matched across colors and RW objects, the RW object advantage should disappear. In the present study, we explore this hypothesis through direct and conceptual replications of recent work. Surprisingly, we find no evidence of a RW object benefit and instead find a robust “meaningless” benefit, where colors are remembered better than RW stimuli (exp 1). This effect appears to be driven by the distinctiveness of each color since replacing the colors used in previous studies with CEILAB colors this “meaningless” benefit disappears (exp 2). Across several studies we systematically manipulate (1) spatial cues, (2) encoding time, (3) stimulus sets, and (4) probe types (e.g., recognition memory vs change detection) to eliminate or exaggerate the effects that PI might have on VWM. Together, these findings draw attention to the fragility of any PI that is claimed to be the driving force of performance differences across conditions, and demonstrate the effect that stimulus distinctiveness can have in a design. Importantly, they also highlight the importance of properly curated stimulus sets when evaluating theoretical claims regarding WM capacity.

"This work was supported by the laboratory of Jamal Williams, University of California, Santa Cruz"

23.351 EYE METRICS DURING FIXATION NOT MODULATED BY SPOKEN LANGUAGE COMPREHENSION

Will Epstein¹, Grace Edwards¹, Sophia Lipetzky², Anna Seydell-Greenwald³, Elisha Merriam¹, Laurentius Huber⁴, Ella Striem-Amit³, Christopher Baker¹; ¹National Institute of Mental Health, ²University of New Mexico, ³Georgetown University Medical Center, ⁴Massachusetts General Hospital

When listening to spoken language, congenitally blind populations show increased activity in primary visual cortex (V1), potentially indicative of cross-modal plasticity. However, recent evidence shows that V1 is also modulated by spoken language in sighted individuals (Seydell-Greenwald et al., 2023), suggesting innate connections between V1 and areas associated with language processing. We replicated the activation of V1 during spoken language comprehension in 7 Tesla (7T) MRI, to examine the depth-dependent profile of feedback in V1 from non-visual stimuli. In order to isolate non-visual activity in V1, we examined fixational eye-movements to ensure that differential V1 activation could not be explained by condition dependent eye-movements. Following the methods of Seydell-Greenwald and colleagues, we presented sighted participants (n=10) with 3 runs of audio recordings of forward and reversed speech in 30-second counterbalanced blocks, both within the 7T and in a purely behavioral setting. In the forward speech condition, participants heard short sentences (e.g., "Something that keeps food cold is a refrigerator", "A three-sided shape is a globe") and were tasked to report semantically incorrect statements. In the reversed speech condition, participants heard the same sentences played backwards and were asked to report a beep at the end of some statements. Participants were asked to maintain fixation at the center of the screen while listening to the speech, and their eye movements were recorded using an Eyelink 1000 in both the 7T and the behavioral suite. Between the forward and reverse conditions, we found no difference in pupil size, gaze position density, and fixation drift across time. These results suggest that the V1 activation difference is not driven by differences in eye metrics across conditions, and strengthens the interpretation that the depth dependent profiles for forward and reverse spoken language in V1 can be restricted to feedback modulation from non-visual areas.

23.352 ARE VISUAL SEARCH EXPECTANCY EFFECTS STABLE ACROSS STIMULUS TYPES AND TIME?

Dana M. Farmer¹, Natalie A. Paquette¹, Joseph Schmidt¹; ¹University of Central Florida

Expectations often shape how individuals allocate attention during visual search (Schmidt & Zelinsky, 2017). When task-specific expectations are violated, search behavior shifts toward what was expected rather than what is required, producing an expectancy effect. Even subtle changes to task instructions can bias expectations (Cox, et al., 2021), leading behavior to align with expected rather than true

task demands. Although expectancy effects have been observed across different stimulus types (Paquette & Schmidt, 2025), it remains unclear whether the magnitude of these effects reflects stable individual differences that generalize across stimulus types and if these effects remain stable over time. The present study employed a within-subjects visual search task to assess if the expectancy effect remains stable across stimulus types and block, with participants searching for Landolt-C's and real-world-objects while eye movements were recorded. A blocked design was used to reinforce expected search difficulty and expectations were violated 25% of the time (e.g. easy trials in a difficult block, easier-than-expected; difficult trials in an easy block, harder-than-expected). Five performance and oculomotor-based measures were evaluated by block and violation type, resulting in a total of 20 individual correlations. Correlational analyses across stimulus types revealed a significant inverse relationship for target dwell time during block one easier-than-expected trials ($r = -0.38$, $p = .013$), suggesting that individual differences in expectancy effects show some consistency, despite a negative relationship. This relationship weakened in the second consecutive block ($r = -0.26$, $p = .095$). No significant relationships were observed for harder-than-expected trials in either block. These findings do not provide strong evidence for stable individual differences in performance following expectation violations. However, when stability is present, it weakens with task exposure. The attenuation of this effect over time may reflect an accommodation of existing task-oriented expectations following repeated expectation violations (Pinquart et al., 2021).

23.353 FACE PREFERENCE IN EARLY VISUAL CORTEX IS DRIVEN BY IMAGE STATISTICS

Emily Granger^{1,2}, Corey M Ziemba¹, Philipp Kreyenmeier², Hendrikje Neinborg¹, Elisha P Merriam²; ¹Laboratory of Sensorimotor Research, National Eye Institute, Bethesda, MD, ²Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, MD

Visual processing in the brain is organized hierarchically, with early areas selective for basic visual information and later areas becoming selective for complex, categorical information. Areas of the early visual cortex (V1-V3, EVC) are believed to mainly process low-level features such as position and orientation. The fusiform face area (FFA), in contrast, exhibits a strong preference for faces that relies on higher-level processing. Despite this, some studies have suggested that EVC contains category selective responses, particularly in the foveal representation, a property attributed to feedback from higher visual areas. Here we investigate whether this preference can be explained by low-level image statistics that are overrepresented in face stimuli compared with other image categories. Subjects viewed alternating blocks of rapidly presented images of isolated faces or objects while fixating on the center of the screen. Participants performed a 1-back task on the images while we measured BOLD fMRI responses with a 7T scanner. We observed strong face preference in FFA, as expected, and a robust face preference in the foveal representation of EVC. To better understand the origin of this foveal face preference, we used a

texture synthesis algorithm to scramble the more complex features of the face and object images, rendering image category unrecognizable, while maintaining the low- and mid-level features. We found that FFA did not exhibit significant selectivity to face textures relative to object textures, consistent with the putative role of FFA in face perception. However, we observed robust face preference in foveal EVC that was similar for both intact and scrambled images. These results suggest that feedback from FFA is unlikely the source of category selectivity in EVC. Rather, our results demonstrate that category selective response in EVC reflect a preference for low-level natural image statistics.

We acknowledge funding from the National Eye Institute and National Institute of Mental Health Intramural Research Program at the National Institutes of Health (NIH, grant no. 1ZIAEY000570-01 to H.N. and NIH, grant no. ZIAMH002966 to E.M.)

23.354 DOES EARLY SPATIAL BIAS IN LINGUISTIC VERSUS NON-LINGUISTIC SEARCH TASKS DRIVE SEARCH GUIDANCE?

Skylar Greenberg¹, Grace Sinclair¹, Joseph Schmidt¹; ¹University of Central Florida

Early spatial biases, indicated by initial saccades, have been observed in both linguistic and non-linguistic tasks. Studies have demonstrated an early leftward spatial bias in non-linguistic tasks like driving, memory, free-viewing, and visual search (Durgin et al., 2008; Staugaard et al., 2016; Dickinson & Intraub, 2009; Hartmann et al., 2019); whereas a rightward attentional spread occurs in left-to-right English reading, reflected in horizontal linguistic search tasks (Ojanpää, 2002; Rayner & Raney, 1996). We expect to see stronger search guidance (indicated by the proportion of initial search saccades directed toward the target) on the left for object search tasks due to the early leftward bias and stronger guidance on the right for word search due to the rightward attentional span in English reading. We conducted two experiments examining the proportion of initial saccades to the left side, and search guidance in Visual Word Search (Experiment 1) and Object Search (Experiment 2). In both experiments, participants searched for a target word or object among a circular array of distractor words/objects while eye movements were recorded. Targets were present equally often on the left and right side of the array. Both conditions exhibited an early leftward spatial bias. Search guidance was stronger on the right for the object search but stronger on the left for word search. Object guidance results are likely linked to the left hemisphere's efficiency in recognizing physical features of symbols in the right visual field (Polich, 1982), whereas the word search guidance effect is likely because English reading requires a leftward shift to start reading. Collectively, this demonstrates that an early spatial bias does not lead stronger search guidance in that region, and bias and guidance are non-uniform across the visual field (Kosovicheva & Whitney, 2017) and vary by stimulus type.

23.355 COMPARING VISUAL RESPONSES IN FRONTAL VERSUS TEMPORAL CORTEX MACAQUE FACE PATCH NEURONS

Carly Gregg¹ (carly.gregg@nih.gov), Kenji Koyano², Will Robison¹, Jess Taubert³, David Leopold¹; ¹Section on Cognitive Neurophysiology and Imaging, NIMH, ²Advanced Neuroimaging Center, QST, ³School of Psychology, The University of Queensland

Facial recognition, a critical skill in primate social behavior, is supported by specialized brain areas in the temporal and frontal cortices. We previously reported that neural responses in face-selective regions of the inferior temporal cortex, including the anterior fundus (AF) and anterior medial (AM) face patches, are shaped by both average facial structure and local feature details. The same neural populations exhibited minimal selectivity for pareidolia stimuli, namely objects whose internal configural arrangements cause them to resemble faces. The aim of the present study is to replicate and extend these findings in additional face patches in the temporal and frontal cortex. In two male Rhesus macaques, we used fMRI to localize the AM and perirhinal (PRh) face patches in the temporal cortex and the orbitofrontal cortex (PO) and frontal pole (FP) face patches in the prefrontal cortex. We implanted one 64-channel chronic brush electrode into each patch in both animals. We simultaneously recorded single-unit responses from the four areas to (1) identity morphed faces, (2) feature-recombined faces, and (3) pareidolia images. For the identity morphing experiment, animals viewed macaque and human faces across identity levels including caricatures (exaggerated identity), antifaces (negative identity), and near-average faces. Responses in frontal vs. temporal lobe face patches were generally similar for these stimuli, exhibiting V-shaped tuning around the average face. For the local features experiment, a set of 10 images was broken into component parts (eyes, nose-mouth, outer face, etc.) and recombined to create a set of 450 images which the animals viewed. Preliminary analysis shows that AM and PRh neurons are more strongly driven by local features than the frontal cortex face patches. Lastly, all cells responded only weakly to pareidolia stimuli. Together, these results highlight a significant overlap in response selectivity among fMRI-identified face patches in the temporal and frontal cortices.

23.356 DYNAMICS OF FACIAL EMOTION PERCEPTION ACROSS HUMANS, PRIMATE IT CORTEX, AND ARTIFICIAL NEURAL NETWORKS

Stephanie Habach¹, Marcus Wong^{1,2}, Sabine Muzellec¹, Maren Wehrheim^{1,3}, Kohitij Kar¹; ¹York University, Toronto, ²University of Toronto, ³Mila - Quebec AI institute, Montréal, Canada

Facial expressions unfold over time through rapidly changing visual cues. Humans can readily report when one emotion transitions to another, but how this ability relates to dynamic neural encoding and its underlying computations remains unclear. To operationalize this behavior, we developed a change-detection task quantifying perceived

transition timing in dynamic facial stimuli. We generated 1260 videos (500ms) in which one expression gradually morphed into another (6 expressions, 12 identities) at varying transition times. 344 human participants reported perceived change moments using a continuous slider. To assess accuracy, we compared behavioral reports to ground-truth transition times. Although performance was reliable, correlations with ground truth (Spearman's $R=0.56$) fell below the noise ceiling ($R=0.72$), indicating systematic deviations. We hypothesized that these deviations reflect stimulus-dependent biases. To test this, we focused on a subset of videos with identical midpoint transitions, holding ground truth constant. Despite this, observers showed reliable early-late biases across stimuli (reliability = 0.48), demonstrating that perceptual timing depends on stimulus-specific dynamics rather than objective timing. We next asked whether current vision model computations capture these biases. Mid-transition videos ($N=360$) were presented to artificial neural networks, and time-resolved features were decoded using cross-validated linear classifiers across early and late frame segments. Although models accurately classified emotions, their inferred change times did not correlate with human behavior and showed no systematic early-late differences. We then tested neural representations by presenting the same stimuli to two passively fixating monkeys while recording IT activity (114 neurons). However, unlike ANNs, IT dynamics exhibited a systematic bias: relative emotional strength was smaller for early-biased videos and larger for late-biased videos ($p = 0.031$), indicating that perceptual timing reflects time-integrated differences in competing representations. Together, these findings reveal species-conserved temporal dynamics in high-level visual representations that predict human perception of expression transitions that ANNs do not yet predict.

KK is supported by Canada Research Chair (CRC-2021-00326), Brain-Canada Foundation (2023-0259), Canada First Research Excellence Funds (VISTA Program), and NSERC (RGPIN-2024-06223). KK, SH, MW are supported by SURFIN (Simons Foundation). SM, MW are funded by Connected Minds (supported by CFREF).

23.357 SUSTAINED ATTENTION AND CHANGES IN FLANKER TASK PERFORMANCE ACROSS TIME

Zoe E. G. Greenwood¹, Paul Brancaleone², Bruce D. Bartholow³; ¹The University of Iowa 1, 2, 3

The present analyses aimed to characterize individual differences in sustained attention in adolescents via changes in behavioral performance and event-related potentials during a flanker task. Participants (Ps) were 313 adolescents and emerging adults (ages 14-19; $M_{age} = 16.7$; 66% female) recruited for an ongoing longitudinal study of alcohol use risk. Ps completed 300 trials of a standard arrowhead flanker task across three equal blocks while EEG was recorded. Ps were assigned to one of three post-hoc groups ($ns = 104$ -105 each) based on change in reaction time (RT) across the task (Block 3 - Block 1): RT improvers, RT stable, and RT worseners. Analyses tested for group differences in changes in P3 amplitude,

indexing attentional and motivational processing, and N2 amplitude, indexing conflict monitoring, across blocks. Mixed factorial ANOVAs and linear regressions were conducted in Jamovi and R, with Ps' age and sex included as covariates. Flanker congruency effects in accuracy did not differ across the post-hoc groups, $F(2, 298) = 0.197$, $p = .821$. There was a significant Block x Group interaction for P3 amplitude, $F(4, 556) = 2.43$, $p = .047$, indicating P3 change varied across groups. Post hoc Tukey tests showed that from Block 1 to Block 3, RT worseners showed the largest decrease in P3 amplitude (mean difference = 1.74, $p < .001$), RT stable participants showed a smaller decrease (mean difference = 0.64, $p = .038$), and RT improvers showed no significant change (mean difference = 1.65, $p = .207$). There were no P3 latency effects between groups and no significant difference in N2 amplitude. These findings suggest that worsening performance over time on a cognitive control task reflects reduced sustained attention and task motivation. Larger decreases in P3 amplitude were observed among participants whose performance worsened, whereas conflict monitoring did not differ across groups.

SATURDAY MORNING POSTERS IN PAVILION

Object Recognition: Reading

SATURDAY, MAY 16, 8:30 AM – 12:30 PM, PAVILION

23.401 LETTER RECOGNITION ACCURACY PREDICTS WORD RECOGNITION ACCURACY IN DISPLAYS WITH UNUSUAL SPATIAL CONFIGURATIONS

Arda Fidanci¹ (fidan005@umn.edu), Colin S. Flowers¹, Sophia Osovsky¹, Gordon E. Legge¹, Stephen Engel¹; ¹University of Minnesota

Remapping is an assistive approach for people with central vision loss (CVL) that works by relocating letters hidden by a scotoma to functioning retina outside the scotoma. How letters are relocated—for instance, by warping them around the scotoma versus moving entire lines up or down—can differentially affect letter recognition (Fidanci et al., 2024) and reading (Flowers et al., 2024). Here, we tested whether the differences in word reading due to various remappings can be predicted from letter recognition performance. We measured letter and word recognition in 10 young controls with three simulated scotoma shapes (circular, horizontally elongated, and vertically elongated). We compared five remapping strategies from prior work : (1) Maximum Row, moving letters onto the single row with highest overall letter-recognition accuracy; (2) Horizontal Gap, shifting letters horizontally to jump over the scotoma; (3) Vertical Gap, shifting letters up or down to bypass the scotoma; (4) Diagonals, positioning letters on diagonal paths around the scotoma; (5) MaxAcc, relocating each letter to the

most accurate location within its column. In the letter recognition task, three letters were flashed for 200 ms at successive positions along the remapping path while maintaining central fixation. In the word recognition task, words, ranging in length from 3 to 10 letters, were presented for 200 ms along the same remapping paths. Letters and words were presented in a total of 15 conditions (strategy x scotoma shape). The letter and word recognition accuracies varied widely across conditions and were best overall in the Horizontal Gap condition (letter = 83.42%; word = 77.98%) and worst in the MaxAcc condition (62.18%; 62.65%). Across all conditions, letter and word recognition correlated highly with each other ($r = 0.76$, $p = 0.001$). The spatial configurations in remapping strategies impact letter recognition differentially, which is what likely produces differential word-reading performance.

NIH Grant EY030890

23.402 THE EFFECT OF CASE AND FONT STYLE ON ENGLISH LETTER REPRESENTATION IN THE HUMAN OCCIPITOTEMPORAL CORTEX

Roxana Ismail-Beigi¹, Marvin Chun¹, Yaoda Xu¹; ¹Yale University

English letter recognition is easier with uppercase than lowercase letters, and easier with sans-serif than serif fonts. To understand the neural bases of these effects, in Experiment 1, we showed human participants individual letters in either uppercase or lowercase sans-serif (Arial) or serif (Times New Roman) font while they performed an orthogonal image jitter detection task. The average letter size between the two fonts was equated, while within each font, letter size difference between uppercase and lowercase was unaltered to preserve how these letters are normally perceived. We examined fMRI responses in retinotopic areas V1 to V4, object-selective areas in the occipitotemporal cortex, and the visual word form area (VWFA). Response amplitude was higher to uppercase than lowercase letters in V1 and V2, and higher to serif than to sans-serif letters in object-selective areas. Meanwhile, response patterns were more distinctive for sans-serif than serif letters in object-selective areas, and more distinctive for uppercase than lowercase letters across several areas. To replicate these results while controlling for typeface, in Experiment 2, we used the serif and the sans-serif fonts in the Lucida family. We additionally equated the average size of the uppercase and lowercase letters within each font style. With size equated, response amplitude measures no longer showed an effect of letter case in V1 and V2, but higher responses to serif fonts persisted in object-selective areas. Response patterns were again more distinctive for sans-serif letters in object-selective areas, and more distinctive for uppercase letters in several areas. Across both experiments, case and font style effects were largely absent in VWFA. Overall, these results showed that while letter size may contribute to the letter case effect by modulating response amplitudes, representational distinctiveness likely plays a greater role in mediating the relative ease of recognizing uppercase letters and sans-serif fonts.

Supported by NIH Grant R01EY030854 to YX.

23.403 ARTIFICIAL FIXATION POINTS USING LETTER HIGHLIGHTING IMPROVES READING SPEED AND COMPREHENSION

Ali Bhimani¹, A. Caglar Tas¹; ¹The University of Tennessee, Knoxville

Previously, we have shown that using artificial fixation points (AFPs) to cue participants while reading can potentially help readers by increasing reading speed and improving reader comprehension (Bhimani & Tas, VSS 2024). Further, we have also demonstrated that the type of AFP used has a significant impact on measures of efficient reading (Bhimani, Parker, & Tas, VSS 2025). We created AFPs by altering the first half of either highly fixated words or words which were pseudo-randomly selected based on a theoretical framework. In our prior work, AFPs were implemented by altering the morphological characteristics of manipulated letters through the addition of boldness (space and contrast) or color (contrast and saliency). In the present study, we aimed to test the generalizability of AFPs through the manipulation of features outside of individual letter morphology by highlighting the first half of selected words. Our results showed that highlight AFPs were similarly effective. For difficult texts, reading comprehension and reading speed (words per minute) were both significantly better for texts that used highlighted AFPs compared to texts that did not include any AFPs ($ps < .001$). For medium-difficulty texts, AFPs only improved reading speed ($ps < .001$), without affecting reading comprehension. AFPs did not have an effect for easy texts on either measure, suggesting that AFPs are utilized when texts are challenging. These findings support the generalizability of AFPs: Improved reading speed and comprehension can be achieved by manipulating both lexical properties (i.e., saliency of letters), but also saliency of the surrounding space.

23.404 CROWDING EXPLAINS READING SPEED AND COMFORT

Anna Bruns¹ (anna.bruns@nyu.edu), Najib J. Majaj¹, Maria Pombo¹, Minjung Kim², Denis G. Pelli¹; ¹New York University, ²Independent research scientist (Work completed while at Meta Platforms Inc.)

People read at widely different speeds, but we don't know why. Type designers emphasize reading comfort and create "Text" fonts that are comfortable to read, but we still lack a scientific account of how letter shape affects reading. One known constraint on reading speed is crowding – the failure to recognize a letter because its neighbors interfere. Here, we report the first joint measurements of reading speed, reading comfort, and crowding. Participants completed two reading tasks: ordinary reading of a multi-page story and Rapid Serial Visual Presentation (RSVP) reading of nouns flashed serially in one location. After each story, they rated reading comfort (1 to 7). Each of 24 fonts was tested with 26 to 49 participants. Crowding distance – the minimum spacing needed to identify a letter – varied across both fonts and individuals, and explained more than half the variance in reading speed and comfort ($R^2 = 54\%$ for RSVP reading, 55% for ordinary

reading, and 73% for comfort). Higher-crowding fonts were read more slowly, with less comfort. Higher-crowding participants also read more slowly but with the same comfort. Crowding varied sevenfold across participants, and fourteenfold across fonts. The product of font and participant crowding factors predicts 27% of the variance in individual reading speeds. These results suggest that reading is constrained by a crowding bottleneck, and that reading comfort is a perceptual marker of low crowding. The letters in our high-crowding fonts are connected, broken, or compressed, making them harder to individuate. Attempting a neuroscientific account of these behavioral results, V4 cortical crowding distance scales with perceptual crowding distance, suggesting that more V4 neurons are required to individuate letters in high-crowding fonts.

Supported by Meta, Inc., contract to D. Pelli at NYU.

23.405 THE RELATIONSHIP BETWEEN FONT BEAUTY AND READING COMFORT

Denis Pelli¹, Minjung Kim², Maria Pombo¹; ¹New York University, ²Independent research scientist (Work completed while at Meta Platforms Inc.)

Is comfort beautiful? Can beauty be uncomfortable? We tested 24 fonts to explore the relationship between the beauty and the comfort of reading. Participants viewed a paragraph of nonsensical Latin text and rated how much beauty they felt from the letters. They also rated reading comfort twice: after an ordinary reading task and after Rapid Serial Visual Presentation (RSVP), one word after another in the same location. After averaging across participants for each font, we find that the two comfort ratings are strongly correlated ($R^2 = 0.88$), and that comfort predicts beauty ($R^2 = 0.96$). Beauty and comfort together reveal three distinct font clusters that correspond to the traditional font categories: Text, Display, and Script, though neither measure alone can make this distinction. Text fonts (serif and sans serif) are average in beauty but high in comfort; Display fonts are low in beauty and comfort; and Script fonts are high in beauty and average in comfort. However, within Text fonts, beauty and comfort ratings do not distinguish serif from sans serif. We also find striking individual differences. Beauty and comfort are positively correlated for most participants, but strongly negatively correlated for others (6%). To our surprise, a participant's beauty-comfort correlation is predicted by their preference for minimalist style ($p < 0.001$) and by the difference in their ratings across font categories ($R^2 = 0.60$). In product design, form and function are often assumed to be independent, but our results reveal a near-perfect correlation between font beauty and reading comfort.

Supported by: Meta Inc. contract to D. Pelli.

23.406 INDIVIDUAL DIFFERENCES IN CROWDING PREDICT READING SPEED IN RAPID SERIAL VISUAL PRESENTATION

Fengping Hu¹ (fh986@nyu.edu), Anna Bruns¹, Najib J. Majaj¹, Jonathan Winawer¹, Denis G. Pelli¹; ¹New York University

Visual crowding—defined as failure to recognize in clutter—reduces the visual span, which reduces reading speed. Here, complementing the reported stimulus effects that relate crowding to reading, we ask whether there are also individual difference effects: Do individual differences in crowding predict differences in reading speed across a large online sample? 168 online participants completed three tasks: (1) Visual crowding: identifying the middle letter of a trigram at $\pm 8^\circ$ eccentricity (QUEST threshold spacing); (2) RSVP reading: identifying three sequentially presented nouns at fixation (QUEST threshold word duration); (3) Ordinary reading: speed in word/min based on time taken to read a short IReST story (Trauzettel-Klosinski et al., 2012), with Flesch-Kincaid grade level 7.2 ± 1.2 , with subsequent questions to verify comprehension. Results confirmed large individual differences: all three measures varied several-fold, with individual differences much larger than within-subject measurement noise. Combining the measures revealed a moderate correlation ($r = -0.38$) between crowding distance and RSVP reading thresholds, indicating that individuals with more crowding (larger spacing thresholds) read more slowly. In contrast, crowding did not correlate with ordinary reading speed. We speculate that for some participants, letter spacing is the bottleneck that limits the rate of RSVP reading, whereas there are other, narrower bottlenecks that limit the rate of reading the 7th-grade-level passages presented to our adult participants.

Supported by: Meta Inc. contract to D. Pelli.

23.407 EXAMINING THE BLUR PARADOX IN WORD RECOGNITION

Simrat Binning¹, Ipek Oruc¹; ¹University of British Columbia

Blur disrupts visual recognition by reducing high spatial-frequency information needed for fine-detail discrimination. Because distant objects project blurrier retinal images, recognition is expected to worsen with distance. However, our previous work on faces revealed a different pattern: blurry faces are recognized more accurately when viewed small or afar than when large or near - an effect we termed the blur paradox. Whether this phenomenon is face-specific or a general visual mechanism remains unclear. Here, we tested whether it extends to word recognition. We selected 100 frequency-matched words from the Corpus of Contemporary American English. Intact and blurry versions were generated at four sizes (81, 270.5, 460, 650 pt), using a Butterworth lowpass filter (cut-off: 1.1 cycles/letter). Stimuli were presented at a distance of 114 cm. Participants (Expt. 1: $n=12$; Expt. 2: $n=30$) completed 100 randomly ordered trials, viewing each word once in one of four conditions. Experiment 1 used a 2×2 within-subjects design: size (650, 81 pt) $\times$ blur (intact, blurry). Experiment 2 presented only blurry words at one of four sizes. Percent correct recognition was recorded. Recognition was at ceiling in both intact conditions (Expt. 1). Blur significantly reduced accuracy, demonstrated by a main effect ($F(1,11)=156.8$, $p<0.0001$). A main effect of size ($F(1,11)=170.7$, $p<0.0001$) was also qualified by a size $\times$ blur interaction ($F(1,11)=170.7$, $p<0.0001$), with lower accuracy for blurry words at the large size (36.7%, $SD=16.3$) compared to small (95.7%, $SD=5.5$). In

Experiment 2, accuracy declined with increasing size ($F(3,87)=84.63$, $p<0.0001$), ranging from 88.9% (81pt) to 41.7% (650 pt). These results show that the blur paradox extends beyond faces, also occurring for words. Unlike face perception, small blurry words showed a near-complete recovery of recognition, indicating stimulus-specific processing strategies. These findings have implications for reading and everyday perception, where the visual system compensates for optical limitations.

This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant RGPIN-2025-05239 (IO) and an NSERC USRA award (SB).

23.408 INDIVIDUAL DIFFERENCES IN IDENTIFYING DEGRADED STIMULI: LANGUAGE-SPECIFIC AND GENERAL MECHANISMS

Silvia Guidi¹, Benjamin Wolfe^{1,2}, Anna Kosovicheva^{1,2}; ¹University of Toronto, ²University of Toronto Mississauga

Despite our extensive reading experience, people vary greatly in how well they read. These differences in reading performance can partly arise from differences in how a reader copes with the limitations of their visual system. We investigated readers' ability to identify visually degraded stimuli in order to estimate individual susceptibilities to these manipulations. We tested whether there are common mechanisms for identifying degraded text and, if not, what forms of degradation relate to each other. Participants ($N=11$) completed two tasks: a lexical decision task and a non-lexical control task with six-character strings. We visually degraded the stimuli for both tasks with one of four methods by blurring, spatially distorting, reducing the contrast, or reducing the size of the stimuli, and varied the strength of the manipulation on each trial. We determined the stimulus level that corresponded to the 80% threshold for each condition in each task. In the lexical decision task, thresholds were moderately correlated with each other across manipulations (mean pairwise Fisher's $z = 0.41$). Blur and acuity were the most strongly correlated, such that participants who were more susceptible to blur also showed the largest performance decrement with small text sizes. Correlations were similarly strong in the non-lexical task (Fisher's $z = 0.49$), with the strongest correlation between size and contrast; participants who were better able to identify low contrast stimuli also had better acuity. These correlations show that susceptibility to different visual degradations is partially degradation-specific. Moreover, shared patterns of correlations across the tasks point to non-language specific mechanisms in play. We also correlated the difference in thresholds between tasks and found an average Fisher's z of 0.66, suggesting that language-specific mechanisms contribute to performance with degraded stimuli. Taken together, our results point to both language-specific and non-language specific mechanisms impacting the identification of degraded stimuli.

This work was supported by a SSHRC Insight Grant to AK and BW.

23.409 GLOBAL PRECEDENCE PERSISTS IN DYSLEXIA: A NAVON STUDY WITH CHINESE CHARACTERS

Shao-Chin Hung¹, Hsiao-Lan Sharon Wang², Chien-Chung Chen¹; ¹National Taiwan University, ²National Taiwan Normal University

Humans identify global figures faster than local ones. This global precedence is challenged in dyslexics, who show a local-before-global perception (Franceschini et al., 2017). Here we investigated the stimulus dependence of the global and local perception in three groups: Mandarin-speaking juveniles with developmental dyslexia, age-matched neurotypical juveniles (12-17 years), and neurotypical non-Chinese-reading adults (23-27 years). Participants completed a novel Navon task requiring judgments of whether the global or local elements of stimuli presented bilaterally were identical. Experiment 1 used large geometric shapes (circles, diamonds, or triangles) composed of smaller ones, and Experiment 2 used analogous stimuli constructed from Chinese characters. Across both experiments, all groups showed a global precedence effect, with increased accuracy in global tasks compared to local tasks. In Experiment 1, both juvenile groups showed a congruency effect for both task types, responding more accurately while global and local geometric shapes matched, indicating mutual interference across levels. In Experiment 2, the congruency effect appeared only in local tasks, indicating interference from the global to the local level but not vice versa. The dyslexics showed inferior local perception relative to the neurotypical juvenile, particularly when viewing Chinese characters. The non-readers exhibited global-local performance patterns similar to neurotypical juveniles even when processing Chinese characters. These findings suggest that the impaired reading performance in dyslexics may arise from deficits in visual word form analysis rather than from language-specific knowledge. Overall, we found no evidence of local precedence in dyslexia for either geometric shapes or Chinese characters. The impaired local perception of Chinese characters observed in dyslexics may suggest specialized channels for visual word form processing.

NSTC 114-2410-H-003-137

23.410 VISUAL SPAN CONSTRAINTS IN DYSLEXIA: FROM LAB MEASUREMENT TO SCALABLE ONLINE ASSESSMENT

Nilsu Atilgan^{1,2}, Ben D. Sawyer²; ¹Bilkent University, ²University of Central Florida

Visual span, the number of letters that can be recognized reliably without eye movement, is considered a key sensory bottleneck for reading efficiency. Prior work shows that visual span is associated with reading speed in both typical and low-vision populations (Legge et al., 2007). In a preliminary laboratory study, we investigated visual span and reading performance in individuals with dyslexia. Relative to typical readers, the dyslexia group exhibited significantly smaller visual spans ($p = 0.022$) and slower reading rates ($p < 0.001$). Crucially, visual span size showed a strong positive correlation with reading speed ($r(19) = 0.45$). We adapted this paradigm for online testing to

enable larger-scale assessment. A validation study confirmed the reliability of the online paradigm in a small sample, showing a strong positive correlation between visual span size and reading speed ($r(12) = 0.63$). We then collected data from 45 self-reported individuals with dyslexia (age range: 20–59; $M = 34.13$), using the online visual span test and an online MNREAD test. Consistent with our laboratory findings, this group demonstrated slower reading performance (Words per minute [WPM]: 123 ± 43) compared to typical readers (162 ± 21 WPM). Most importantly, visual span size again showed a strong correlation with reading speed across the heterogeneous sample ($r(45) = 0.60$). Collectively, these findings demonstrate two key implications. First, they confirm that, despite considerable heterogeneity, individuals with dyslexia show a reliable group-level pattern where a smaller visual span predicts slower reading. Second, the successful replication across laboratory and online settings demonstrates that a brief, psychophysically grounded visual span test can be administered reliably online, enabling scalable and valid assessment in larger populations. Our ongoing work examines whether asymmetries in visual span between the left and right visual fields further contribute to performance variability in dyslexia.

3D Shape and Space Perception: Cues, integration

SATURDAY, MAY 16, 8:30 AM – 12:30 PM,
PAVILION

23.411 EXPOSURE TO COVARYING VISUAL DEPTH CUES ACROSS GAZE DISTANCES PROMOTES GENERALIZED STEREOSCOPIC PROCESSING IN DEEP NEURAL NETWORKS

Joonsik Moon¹ (moon.joo@northeastern.edu), Peter Bex¹;
¹Northeastern University

When shifting gaze between far and near distances, vergence eye movements rotate the eyes to fixate in 3D space, inducing changes in low-level image disparity statistics across the retina. Crucially, near and far fixation alters the distributions of binocular crossed and uncrossed disparities along with covariation with other spatial information (e.g., spatial frequency). We hypothesized that exposure to these naturally covarying statistics may be essential for the visual system to learn generalized and robust stereoscopic processing, an aspect often overlooked in standard computational models. We trained Deep Neural Networks on stereoscopic image datasets simulating "Near" and "Far" fixations designed to capture the distance-dependent covariance of disparity range and spatial information. We systematically manipulated training composition by varying the proportion of Far-to-Near fixation ratios and evaluated depth discrimination performance on independent test sets. Networks trained on mixed fixation distance datasets achieved better depth

estimation accuracy compared to homogeneous models (trained on 100% Near or Far fixation), even when tested on their trained domain. We also found that mixed training induced a significantly higher standard deviation in the first convolutional layer's filter weights. This result suggests that exposure to broader visual cue statistics prevents overfitting to the intrinsic statistics of specific fixation distance ranges. Instead, the system appears to evolve efficient coding mechanisms to extract inherent cues that remain consistent across variations in gaze geometry. Visual processing mechanisms are shaped by the statistical structure of the natural environment and by binocular viewing geometry. The results suggest that the joint statistics of disparity and spatial cues across varying gaze distances are critical for constructing an efficient and robust stereoscopic system.

This research was supported by NIH R01EY032162.

23.412 HUMAN DORSAL CORTEX IS TUNED TO REAL-WORLD CUES THAT SUPPORT SHAPE PERCEPTION

Qihan Wu¹, Grant T. Fairchild^{2,3}, Lars Strother², Kallie E. McDonald², Erez Freud⁴, Jacqueline C. Walsh-Snow^{1,2}; ¹Department of Psychology, University of Utah, ²Department of Psychology, University of Nevada, Reno, ³Department of Psychology, Emory University, ⁴Department of Psychology and the Centre for Vision Research, York University

Although the neural basis of object perception has been studied extensively, most research has relied on artificial stimuli in the form of pictures or computer images, rather than tangible, real-world objects. This approach overlooks the possibility that the brain's processes and organization reflect the important features of naturalistic objects and environments, including tangibility and actability. Here, we used fMRI to compare shape selectivity for tangible objects versus matched planar photographs of objects across the ventral and dorsal visual pathways. We were particularly interested in whether shape-processing areas of dorsal cortex would show enhanced responses to tangible objects compared to pictures, given the pathway's role in spatial processing and action guidance. We used a parametric shape scrambling paradigm in which human participants directly viewed 3D-printed everyday objects and matched photos of those objects, each presented with different levels of scrambling. The results revealed distinct format-dependent patterns of shape selectivity: dorsal regions showed stronger shape selectivity for tangible objects than for pictures, whereas ventral regions responded similarly across formats. Our results suggest that dorsal object areas are preferentially tuned to the properties of real-world, manipulable objects, while ventral areas encode object shape independently of physical format. This work provides new insights into the functional role of dorsal shape areas and highlights the value of using real objects to study the integrated functioning of perceptual and motor systems.

National Eye Institute of the National Institutes of Health (R01EY026701)

23.413 LINEAR PERSPECTIVE AS A SOURCE OF INFORMATION ABOUT ENVIRONMENTAL SLANT

William Marcum¹, Maria Carmichael¹, J. Farley Norman¹; ¹Western Kentucky University

An experiment evaluated the importance of linear perspective for the visual perception of environmental slant. Eighteen observers (ages ranged from 19 to 80 years) judged the slant-in-depth of a rectangular (98 cm wide x 120 cm long) physical surface, where the actual slants varied from 20 to 65 degrees. The rectangular surface possessed six narrow (3.7 cm) wooden slats oriented along the longer axis; a column of 11 LED's (light emitting diodes) ran along the center of each lengthwise-oriented slat. When viewed by an observer in lighted visual conditions, the physically parallel slats of the slanted surface appeared to converge (in a projected image, such as a retinal image) due to the effects of linear perspective. Likewise, when viewed by an observer in the dark, the six columns of LED's also converged in the observers' retinal image due to the effects of linear perspective. The observers' task was to adjust the orientation of a line on a computer monitor (relative to vertical & horizontal axes) until its orientation matched the perceived slant of the rectangular surface in depth. The observers viewed the experimental stimuli monocularly (to prevent the use of binocular disparity) and used a chin rest to similarly prevent the potential use of motion parallax as a source of information about slant. Multiple judgments were obtained for each physical stimulus slant. Linear regressions obtained for each observer in both full-cue and dark conditions showed that the observers (both younger and older) could perceive environmental slant almost as well in the dark (when only linear perspective information was available) as in fully lighted conditions (average Pearson R correlation coefficients were 0.951 and 0.935 for lighted and dark conditions, respectively). The results clearly demonstrate that linear perspective is a powerful and effective source of information about environmental slant.

23.414 CUE-INVARIANT 3D SHAPE RECONSTRUCTION FROM HUMAN BRAIN ACTIVITY

Shiyan Yang^{1,2} (yang.shiyan.32r@st.kyoto-u.ac.jp), Shuntaro Aoki^{1,2}, Reo Tsukasa^{1,2}, Misato Tanaka^{1,2}, Yukiyasu Kamitani^{1,2}; ¹Graduate School of Informatics, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan., ²ATR Computational Neuroscience Laboratories, Soraku, Kyoto 619-0288, Japan.

The human brain effortlessly constructs three-dimensional (3D) percepts from depth cues within 2D retinal images. While the neural processing of individual depth cues is well-documented, how these signals are integrated to form a coherent, cue-invariant 3D perception remains elusive. Here, we present a method to reconstruct 3D object appearance directly from human brain activity. Adopting a translator-generator framework, we decoded fMRI responses elicited by rendered 2D images into the latent space of a deep autoencoder trained on 3D point clouds, and subsequently transformed these latent features into explicit 3D shapes using a generative model. This approach accurately reconstructed 3D geometries of single objects

from fMRI responses to their rendered 2D images, capturing the global shape and orientation for both familiar object categories and entirely novel objects unseen during training. Crucially, the same decoder, trained exclusively on 2D pictorial depth cues, successfully generalized to reconstruct 3D shapes from random-dot stereograms (RDSs), stimuli defined solely by binocular disparity. Reconstruction performance was generally high across the visual cortex for both stimulus types. However, in the RDS condition, while the dorsal visual cortex and MT-neighboring regions remained robust, performance in early visual areas tended to decline. Control experiments using "contour-matched" stereograms of slanted bars confirmed that the dorsal stream could recover the slanted 3D structures even though 2D contour information was experimentally silenced. These findings provide a tangible externalization of the brain's internal 3D world, offering direct evidence for a robust, cue-invariant representation of 3D appearance.

Supported by JSPS KAKENHI (JP25H00450, JP20H05705, JP20H05954), NEDO (JPNP20006), and JST CREST (JPMJCR22P3), all awarded to YK.

23.415 TESTING MODELS OF SENSORY CUE INTEGRATION WITH PHYSIOLOGICAL MEASUREMENTS

Peter Scarfe¹, Anna Horwood², Patricia Riddell³; ¹School of Psychology and Clinical Language Sciences, ²University of Reading, ³Reading, UK

Contemporary models of sensory cue integration propose that by combining sensory cues we gain a more precise sensory estimate compared to that available from the cues in isolation. This is achieved by weighting cues in proportion to their relative reliability, and their reliability relative to a priori information about the perceptual property being estimated. Cue weighting is typically measured behaviourally, e.g. via a two alternative forced choice experiment. However, theoretically, one could determine cue weighting using data from any level of the perceptual hierarchy, e.g. from physiological response through to perceptual decision. Here, we derive equations allowing one to test cue integration using cue weights inferred from the gain function relating physiological responses to a varying driving stimulus. We apply this model to an extensive preexisting data set of vergence and accommodative responses (measured with a Remote Haploscopic Photorefractor) to distance targets (0.5, 1, 2 and 3 dioptres) varied to isolate individual cues of disparity, blur and angular subtense (Horwood and Riddell, 2008). We find that combined cues physiological gain functions are well predicted by that of individual cues. This being clearer for vergence compared to accommodative responses. Our results offer a novel way in which to test models of sensory cue integration, which complements current behavioural methods. We highlight the assumptions and limitations of this approach, as well as its ability to disambiguate between competing models of cue integration.

23.416 CONTEXTUAL SHAPE CUES ENHANCE SHAPE CONSTANCY FOR SLANTED SURFACES

Iroshini Gunasekera^{1,2} (iroshini@yorku.ca), Domenic Au^{1,3}, Abtin Zaker³, Robert S. Allison^{1,4}, Laurie M. Wilcox^{1,2}; ¹Centre for Vision Research, York University, ²Department of Psychology, York University, ³Department of Biology, York University, ⁴Department of Electrical Engineering and Computer Science, York University

Shape constancy allows the brain to interpret the 3D structure of an object despite changes in its retinal projection. For example, when a planar surface is slanted in depth, its retinal projection becomes foreshortened, and the visual system must account for slant to recover its shape. If depth and slant are estimated accurately, then the shape should be accurately perceived. If not, it would be distorted. While such distortions have been documented, it is not clear which sources of depth information best support shape constancy. Here, we systematically assessed the impact of monocular and binocular depth cues on shape constancy for planar surfaces rendered in perspective. Renderings of rectangular table-top surfaces were presented sequentially using a mirror stereoscope; one was positioned so that it receded in depth along its long x-axis (reference surface), and the other was aligned horizontally. Using a method of adjustment, participants (N = 18) modified the length of the horizontal surface to match the perceived length of the reference surface. Four conditions with varying amounts of depth information were tested in random order in both perspective (2D) and stereoscopic 3D: untextured rectangles, dropped shadows, texture, and contextual shape cues (table legs and plates). In all viewing conditions the stimuli were rendered from a 25-degree angle at 0.45 m. Results showed that shape estimation improved most under stereoscopic 3D viewing, with the contextual cue condition producing the most accurate estimation in both 2D and 3D. Although shape estimation did not significantly differ between the untextured rectangles, dropped shadow and texture conditions, participants showed consistently higher shape estimates when stereopsis was available. These findings highlight the strong impact of contextual cues on shape constancy and demonstrate that even small reference shapes (table legs and plates) greatly improve the visual system's ability to recover shape from slanted projections.

This research was supported by grants RGPIN-2019-06694 to LW and ALLRP 570802-21 to RA and LW from the Natural Science and Engineering Council (Canada), and CF-REF Connected Minds funding to I. Gunasekera.

23.417 COOPERATIVE INTEGRATION OF SHADING AND TEXTURE IN 3D SHAPE PERCEPTION

Celine Aubuchon¹, Roland W Fleming^{1,2,3}; ¹Justus Liebig University Giessen, ²Center for Mind, Brain and Behavior, Universities of Marburg, Giessen & TU Darmstadt, Germany, ³Excellence Cluster "The Adaptive Mind", Universities of Marburg, Giessen and TU Darmstadt, Germany

While image patterns that elicit 3D percepts (e.g., shading, texture) are typically described as "depth cues", they often do not actually specify depths. Instead, different cues are systematically related to different 3D properties. For example, shading relates primarily to an object's second-order surface structure, but does not capture how stretched or sheared it is (first-order structure). Conversely, texture gradients vary directly with first-order structure, but are often too coarse to capture subtle curvatures. This means that in typical conditions, when both cues are present, they actually provide complementary information about shape. How are these naturally occurring "conflicts" about depth structure resolved? Put another way: how are the complementary signals integrated? To investigate this, we designed a class of 3D objects that maximize perceptual shape differences between shading and texture. Smooth, thin blobs were first perturbed by shallow second-order corrugations along one-direction and then strongly sheared in the orthogonal direction. This leads to shapes that look quite different depending on whether they are shaded or textured. Observers adjusted 3D probes to report the perceived shape across different surfaces in shading-only, texture-only, and combined-cue conditions. As expected, in the shading conditions observers reliably perceived the corrugations, but saw the object as largely fronto-parallel rather than sheared. Conversely, in the texture conditions, shear was apparent, but the subtle corrugations appeared flatter. Interestingly, when both cues were available, observers' perceived shapes were not a compromise between the two single-cue estimates, but rather retained the more informative aspects of both cues. Indeed, depth estimates were closer to shading in one direction and closer to texture in the orthogonal direction within the same stimulus. This has important implications for cue-integration models, because it suggests that cues do not necessarily supply "redundant" estimates, but instead may be assigned specialized roles in estimating particular aspects of shape.

This study was funded by the European Research Council (project number 101098225 - ERC-2022-AdG "STUFF")

3D Shape and Space Perception: Surfaces, objects

**SATURDAY, MAY 16, 8:30 AM – 12:30 PM,
PAVILION**

23.418 WHY DO HUMANS UNDERSTAND PICTURES?

Aaron Hertzmann¹ (hertzman@dgp.toronto.edu); ¹Adobe Research

I outline possible answers to the long-standing question of why pictures work: why can people look at a painting or photograph, and see a depicted subject, rather than just marks on a page or lights on a display? Observers with no prior experience with pictures can understand some kinds of pictures, indicating that picture understanding is not solely a product of experience or culture. I argue

that picture perception can be explained as a product of several properties of real-world vision. First, the fact that humans can understand certain real-world phenomena—refraction, reflection, cast shadows—as simultaneously surface phenomena but also images of an underlying cause explains why we can see pictures as depictions and not just markings. Second, the fact that viewers can understand real-world scenes with unfamiliar combinations of objects explains our ability to understand many different styles of depiction. For example, we can understand black-and-white photos of people because, in real-world vision, we could recognize a familiar person who had been painted gray. Third, our robustness to visual defects and other difficult viewing conditions explains our ability to understand styles of pictorial textures, like paint strokes. Extensions of these basic ideas can explain depiction in many different visual styles, including photographic tone reproduction, line drawings, silhouettes, cartoons, painterly styles, and more. The proposed models of picture understanding could significantly inform future analysis of perceptual mechanisms, picture aesthetics, and the nature of different styles of depiction.

23.419 MONOCULAR ESTIMATION OF GROUND SURFACE ORIENTATION IN NATURAL SCENES

David N. White¹ (davey@autistici.org), Seyed Hosseini¹, Richard Murray¹, James H. Elder¹; ¹York University

Accurate estimation of 3D scene structure is essential for effective interaction with the environment. 3D reconstruction of the ground surface is particularly important for navigation and for anchoring the locations of objects resting on the ground. In the near field, stereo and motion cues provide rich depth information but these cues become weaker in the far field, suggesting a role for monocular cues to surface orientation. To better understand this monocular contribution, we measured human slant discrimination performance for natural ground surfaces and compared human performance with two computational models tested with the same stimuli and task. In a two-interval task, observers judged whether a monocular test stimulus was more or less slanted than a monocular reference stimulus. Stimuli were presented monocularly within a 3D virtual environment. The test stimuli were natural image patches sampled from roughly planar ground-surface regions in scenes from the SYNS image database. Reference stimuli were planar surfaces painted with a high-contrast checkerboard pattern known to elicit relatively accurate percepts of surface orientation. The tilt of the reference stimulus was matched to the tilt of the test stimulus and the slant was systematically varied to sweep out a psychometric function for slant discrimination, providing estimates of both bias and sensitivity relative to the reference. Observers repeated each stimulus condition twice, allowing for analysis of both intra- and inter-observer variability. Results were compared with two-alternative image-computable surface orientation models. The first relies upon an analysis of changes to the shape of the spectral density as a function of depth. The second relies upon perspective distortions in orientation cues. Our results provide a human performance benchmark for the

important visual task of monocular ground surface orientation estimation and offer insight into the computational mechanisms underlying this performance.

Supported by the VISTA postdoctoral fellowship at York University

23.420 A DIRECT VISUALIZATION OF THE SURFACE BOUNDARY OF SUBJECTIVE VISUAL SPACE, THE INTRINSIC BIAS

Lizhu Yan¹ (lizhu.yan@louisville.edu), Teng Leng Ooi², Zijiang He¹; ¹University of Louisville, ²The Ohio State University

A stationary observer perceives a single monocular target as if located on an implicit quarter-elliptical surface in the dark, which we coin the visual system's intrinsic bias (Ooi et al., 2001; Bai et al., VSS 2024). We hypothesized the intrinsic bias, serving as a priori, represents the farthest surface boundary of visual space. Should this be true, multiple monocular targets, even when arranged linearly, would be perceived along the curved intrinsic bias. Five LED targets were presented as: (1) vertical line LED-display ~1.5m (eye-height dependent) from the observer, or (2) horizontal LED-display about the floor 1.5–7.0m from the observer. The angular declination of both the vertical and horizontal targets was 5–45 deg. The 5 LEDs (0.23 or 0.16 cd/m²) had the same angular size (0.23 deg) and inter-target angular separation (10 deg). Observers were tested both with monocular and binocular viewing and responded using drawing-on-paper and blind-walking-gesturing tasks. Confirming predictions, observers with monocular viewing drew the perceived linearly arranged vertical and horizontal LED-displays with curved profiles. This was corroborated with the blind-walking-gesturing task that measured the judged locations of single targets within the displays, and without the displays. The average (n=8) judged target location profiles with the vertical and horizontal displays overlapped (ps>0.05) and were well fitted by the quarter-elliptical function, reflecting the intrinsic bias' shape. We further found with binocular viewing, judged target locations in the two displays deviated away from the intrinsic bias (ps<0.05) and were more veridical. And with the drawing task, observers drew the vertical and horizontal LED-displays as curves that were respectively, steeper and flatter than the intrinsic bias. Overall, the vertical and horizontal LED-displays provided direct visualization of the quarter-elliptical shape of the intrinsic bias. With binocular viewing, the visual system represented target locations by integrating the binocular depth cue with the intrinsic bias.

NIH R01EY033190

23.421 THE VISUAL PERCEPTION OF ORDINAL DEPTH RELATIONSHIPS ON SOLID OBJECT SURFACES

J. Farley Norman¹, William B. Marcum¹, Maria Camichael¹, Evan Hagan¹, Cain B. Del Savio¹, Ash Sheehan², Avery Sutton¹; ¹Western Kentucky University, ²Carol Martin Gatton Academy of Mathematics and Science

Thirty younger and older adults, ranging in age from 18 to 82 years, made ordinal depth judgments for locations on the surfaces of smoothly curved solid objects. The objects' 3-D structure was defined by motion, image shading, and specular highlights; depth differences ranging from 0.4 to 1.2 cm were evaluated. The separation (in the projected images) between the surface regions to be judged ranged from near separations (4.0 cm or less) to far separations (7.0 cm or greater). The task was to indicate which of the highlighted surface regions was closer (to the observers) in depth. The observers' d' values ranged from near chance levels to 3.6, depending upon the image separation, the depth difference magnitude, and the individual observer. There were significant effects of image separation, depth difference magnitude, and age. Overall, the younger observers' performance was 25.2 percent higher than that exhibited by the older observers (average d' values for the younger and older observers were 1.643 and 1.312, respectively). The image separation had large effects: performance for near separations was 113 percent higher than for far separations when the depth difference was identical (average d' values for near and far separations in this case were 2.031 and 0.954, respectively). Larger depth differences produced higher performance than smaller depth differences, with 0.4 cm increases in depth difference magnitude producing 64 percent increases in d' for near separations and 73.6 percent increases in d' for far separations. The current results both confirm (effects of image separation and depth difference magnitude) and extend previous findings (significant negative effect of age). While an age effect did occur, it was small (partial eta squared was 0.144). Older adults (at least up to the age of 82 years) therefore maintain an effective ability to judge ordinal depth relationships on solid object surfaces.

23.422 GEOMETRY REPRESENTATIONS ALONG VISUAL PATHWAYS IN HUMAN SPATIAL NAVIGATION

Taiping Zeng^{1,3} (zengtaiping@fudan.edu.cn), Ming Bo Cai^{2,3}; ¹Fudan University, ²University of Miami, ³University of Tokyo

The representation of geometric structures of one's surroundings is key to self-localization during human spatial navigation. However, the spatial organization of geometry representation in the visual system has not been fully characterized. We displayed navigational videos from identical realistic virtual environments under different weather and lighting conditions to different participants while their brain activities are recorded by fMRI. By modeling the synchronized brain activity from participants watching videos of navigating through the same trajectories, we found that a compact representation of egocentric 3D scene geometric structures is present in a widespread network of brain regions. These regions carrying geometry-related signals form three parallel pathways, which we collectively refer to as "geometry visual pathways", starting from the primary visual cortex: the dorsal and medial pathways end in the intraparietal areas, while the ventral pathway arrives at the hippocampus via the parahippocampal gyrus. The synchronized neural activity in the identified geometry visual pathways allows for reliable decoding of the

3D scene geometric structures. Furthermore, road types, a more abstract representation of the route geometry, are encoded in overlapping pathways, distinctly absent from early visual cortex (V1, V2, V3). In addition to complementing the classical "what" and "where" dichotomy, the identified brain-wide geometry visual pathways narrow a critical gap in understanding how the brain constructs cognitive maps from visual inputs.

The study was funded by WPI-IRCN startup budgets (M.B. Cai and T. Zeng) and JSPS KAKENHI Grant Number JP21K20679 (T. Zeng). The research does not reflect the funders' opinions.

Attention: Neural

**SATURDAY, MAY 16, 8:30 AM – 12:30 PM,
PAVILION**

23.423 DISENTANGLING INHIBITORY AND EXCITATORY EFFECTS OF ANTICIPATORY SPATIAL ATTENTION

Lena L. Kemmelmeier¹, John T. Serences¹; ¹University of California, San Diego

Anticipating a relevant stimulus leads to spatially selective modulations of neural activity in visual cortex, producing a "baseline shift" even before a stimulus appears. These baseline shifts are often interpreted as increased gain in the neurons selective for the spatial position of the attended location. Here, we explore the alternative hypothesis that baseline shifts arise from top-down feedback targeting inhibitory neurons. Targeting inhibitory neurons could stabilize sensory populations and increase their dynamic range, aligning with observed reductions in neuronal variability at stimulus onset. Related fMRI work also shows that the magnitude of baseline shifts scales with anticipated noise in stimulus displays, suggesting the visual system may prepare by suppressing irrelevant inputs. To investigate, we used biologically inspired continuous-time recurrent neural networks (ctRNNs) containing "sensory" layers with tuning properties similar to those found in early visual cortex and a higher-order, PFC-like control layer. We trained the ctRNNs on a selective attention task requiring discrimination at one of two possible spatial locations, and we examined how feedback from the control layer targeted excitatory and inhibitory sensory units. We also manipulated the signal-to-noise ratio (SNR) of the stimulus inputs, and imposed sparsity constraints on the control layer weight matrix to limit processing capacity (i.e. to constrain 'attentional resources'). Pre-stimulus feedback preferentially targeted inhibitory sensory units, especially at the noncued sensory location. Moreover, lower SNR elicited stronger preparatory feedback, paralleling prior fMRI findings. These effects held across regularization strengths, suggesting that the ctRNNs naturally converged on an inhibition-focused strategy across stimulus noise levels and resource constraints. Overall, anticipatory attention may be a regulatory process in which pre-stimulus inhibition increases the dynamic range of neural

responses so that evoked responses related to relevant stimuli stand out against background noise.

23.424 NEURAL CONTROL OF SHIFTING BETWEEN EXTERNAL AND INTERNAL ATTENTION

Daniela Gresch^{1,2}, Dengxinyi Wei^{2,3}, Anna C. Nobre^{2,4}; ¹Ludwig Maximilian University of Munich, ²Yale University, ³University of Oxford, ⁴Wu Tsai Institute

During natural behavior, attention shifts seamlessly between information from the external sensory stream and internal representations held in memory. Only recently has research begun to investigate these dynamic shifts between external and internal attention using behavioral, eye-tracking, and electrophysiological measures. These studies have revealed behavioral differences accompanied by distinct neural patterns: differentiations between within- and between-domain shifts emerged early, evolved dynamically over time, and were broadly distributed across the scalp, suggesting the involvement of multiple processing stages within distributed networks as attention moves between perceptual and mnemonic domains. To determine which neural systems are recruited during such between-domain shifts, we conducted an event-related functional magnetic resonance imaging study using a recently developed task that combines perceptual and working-memory components. In this task, participants were presented with two consecutive cues, sequentially guiding attention to encoded contents, upcoming sensory input, or both. Crucially, the second cue could either shift spatial attention within the same domain (internal-to-internal or external-to-external) or shift spatial attention between domains (external-to-internal or internal-to-external). This design allowed for the systematic investigation of attentional shifts both within and between perception and working memory, as well as direct comparisons of the neural systems engaged in orienting external and internal attention. Consistent with previous work, both external and internal shifts of spatial attention activated dorsal frontoparietal regions, with internal shifts recruiting additional posterior parietal as well as lateral and medial prefrontal areas. Shifting attention between the perceptual and working-memory domains modulated brain activity more subtly, warranting further analysis to elucidate the specific contributions of individual brain regions during between-domain shifts.

23.425 INCIDENTAL LEARNING ABOUT RELEVANT AND IRRELEVANT VISUAL FEATURES IS DRIVEN BY EARLY ATTENTIONAL CHANGES, NOT LATER DECISION-MAKING EFFICIENCY

Kevin Ortego¹, Douglas Addleman², Viola Stoermer¹; ¹Dartmouth College, ²Gonzaga University

Statistical regularities in the environment, such as target or distractor items appearing more frequently in particular locations or colors, influence the brain's processing and prioritization of information, even when we are not explicitly aware of these regularities. Previous research has suggested that the behavioral benefits from incidental

learning rely on modulations of both early attentional processes and later decision-related processes (Ortego et al., 2025; Wang et al., 2023). In two experiments, we recorded EEG while participants performed a visual search task in which color probabilities were manipulated such that either the target (a left/right gap Landolt C among top/bottom gap distractors) or distractors occurred more frequently in a particular color (Exp 1 target learning, N=32; Exp 2 distractor learning, N=54). In both experiments, trials in which the target/distractors matched the most likely color (valid) showed faster response times than trials where this association was reversed (invalid). These behavioral benefits were accompanied by latency shifts of the N2pc ERP component, a neural index of attentional selection, as well as amplitude modulations of the LPC, an ERP component often interpreted as reflecting later decision-making stages, suggesting that the color regularities impacted both early selection and later decision-making processes. However, when controlling for response time differences across conditions (both by subsampling trials to equate RT distributions before computing stimulus-locked ERPs, and by analyzing response-locked rather than stimulus-locked ERPs) we found that the amplitude differences in the LPC disappeared, suggesting these differences are a downstream consequence of earlier processing changes and not a unique contributor to the behavioral benefits. We further demonstrate that latency between attentional selection (as indexed by the N2pc) and response execution was similar for valid and invalid trials, suggesting that modulations early in the processing chain, rather than at later decision-making stages, are the key driver of incidental learning benefits.

23.426 DISASSOCIATE ATTENTION MODULATION BY OBJECT AND LOCATION VALUE IN THE PRIMATE FRONTAL-COLLICULAR PATHWAY

Xuefei Yu¹, Atul Gopal¹, Okihide Hikosaka¹, Ken-ichi Inoue², Masahiko Takada²; ¹Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Bethesda, Maryland 20892, United States of America, ²Center for the Evolutionary Origins of Human Behavior, Kyoto University, Inuyama, Aichi 484-8506, Japan

Attention can be guided by multiple sources, including the value of object features (feature-based attention) and the value of spatial locations (spatial attention). Although the frontal eye field (FEF) and superior colliculus (SC) are known hubs for attention modulation, it remains unclear whether feature-based and spatial signals are processed independently or integrated as they propagate through the frontal-collicular pathway. Here, we addressed this question by associating value with either object features or spatial locations and examining their influence on neural activity using retrograde optogenetics. Two macaque monkeys were trained on an object-value task, in which specific objects predicted high or low reward, and a location-value task, in which reward was tied to stimulus location. Both monkeys acquired these associations, exhibiting faster saccades and preferential choices toward high-value objects and locations. Neural

recordings from FEF and SC revealed robust modulation by both object and location value. These effects were largely dissociable: neurons could encode object value, location value, or both, and among neurons showing dual modulation, ROC-based measures of the two effects were uncorrelated. This indicates that feature-based and spatial attention rely on distinct mechanisms within both structures. To test whether these signals remain disassociated along the pathway, we identified SC-projecting FEF neurons with retrograde optogenetics. These projection neurons also showed dissociable modulation by object and location value, with integration present only in a subset. Conversely, SC neurons receiving FEF input exhibited a slightly higher prevalence of dual modulation, suggesting limited integration may occur at the FEF–SC synapse. Together, our results demonstrate that feature-based and spatial attention modulate neural responses largely independently in FEF and that this dissociability is largely preserved during transmission to SC. This organization supports a flexible coding scheme in the frontal–collicular pathway, enabling both parallel and partially integrated processing of attentional signals.

23.427 ATTENTION-RELATED LFP POWER IN MACAQUE FEF AND FSTS DURING INACTIVATION OF THE SC

Gwen Folkert^{1,2}, Leor Katz¹, Richard Krauzlis¹; ¹Laboratory of Sensorimotor Research, National Eye Institute, NIH, ²Brown University

The midbrain superior colliculus (SC) is crucial for visual attention and plays a causal role in shaping attention-related activity across several cortical areas. We recently reported that inactivating SC reduces attention-related modulations of spiking activity in frontal eye fields (FEF) and the floor of the superior temporal sulcus (fSTS). The contribution of SC to mechanisms of rhythmicity and synchrony, measured in local field potentials (LFPs), is unknown. To address this gap, we investigated the effects of inactivating SC on LFP activity in FEF and fSTS. Two monkeys performed a spatial attention task before and during pharmacological inactivation of SC while electrophysiological activity was recorded from a multichannel probe in FEF or fSTS. Monkeys maintained central fixation while patches of moving dots were presented in each hemifield, such that one of the patches was within the response fields of recorded neurons. In Attend blocks, monkeys were rewarded for releasing a joystick in response to a change of motion direction in either of the two patches, requiring attention to both peripheral locations. In Ignore blocks, monkeys were instead rewarded for reporting a dimming of the fixation point. Before SC inactivation, LFP power in FEF was higher during Attend than Ignore trials in the 500 ms epoch preceding the change, prominently in theta (4–8 Hz) and gamma (>30 Hz) frequency bands. This attention-related LFP modulation was weaker and less frequency-specific in fSTS, but still present. During SC inactivation, deficits in task performance were accompanied by ~50% reductions of attention-related LFP modulation in FEF, consistent with the large impact of SC inactivation on spiking. Attention-related LFP modulation in fSTS was relatively unaffected, despite robust effects on spiking activity. This dissociation suggests that the SC contributes to distinct mechanisms

of attentional modulation in frontal and temporal cortices, rather than exerting stereotyped effects on neuronal synchronization.

23.428 INTERHEMISPHERIC SYNCHRONY CAUSALLY SHAPES RHYTHMIC ATTENTIONAL SAMPLING

Shrey Grover¹, Feiyang Cao¹, Phillip (Xin) Cheng², Robert M. G. Reinhart¹; ¹Boston University, ²Princeton University

Spatial attention samples the environment rhythmically, waxing and waning at a theta frequency (2–8 Hz). However, the neural mechanisms giving rise to this phenomenon remain under investigation. One possibility is that interhemispheric synchrony drives attentional sampling, given that bilateral, particularly parietal, networks coordinate dynamically during attentional allocation. Here, we tested this hypothesis by manipulating synchronization between bilateral parietal cortices using inphase or antiphase high-definition transcranial alternating current stimulation (HD-tACS). Prior work shows that inphase HD-tACS enhances interregional synchrony, whereas antiphase stimulation disrupts it. We designed optimized montages for inphase and antiphase stimulation that targeted bilateral parietal regions at comparable voltage gradients (0.19 V/m). Attentional sampling was measured with a probabilistic cueing task in which participants detected targets embedded in dynamic distractors across varying stimulus-onset asynchronies (SOAs), with blocks before, during, and after stimulation. We delivered HD-tACS at 4 Hz, matching the intrinsic frequency of attentional sampling reported in such tasks, and quantified changes in rhythmicity as decibel-normalized spectral-power changes during and after stimulation relative to baseline using the reaction time-SOA function. We found that inphase HD-tACS (N = 28) modulated rhythmicity as a function of the difference between the stimulated frequency (4 Hz) and participants' intrinsic sampling frequency. Individuals with larger mismatches showed increased power at the stimulated frequency and decreased power at their intrinsic frequency during stimulation, and the increased contribution of the stimulated frequency persisted after stimulation. By contrast, antiphase HD-tACS (N = 29) did not show graded effects but produced a general reduction in low-frequency rhythmicity post-stimulation. These results demonstrate that noninvasive stimulation can causally modify attentional sampling, suggest the possibility of multiple rhythmic oscillators which can be independently entrained through extrinsic stimulation, and show that bilateral parietal dynamics directly contribute to the emergence and strength of behavioral rhythmicity.

23.429 ATTENTION ENHANCES FIGURE-GROUND SEGMENTATION SIGNALS IN VISUAL CORTEX

Michael Epstein^{1,2} (mlepstein88@gmail.com), Emil Olsson³, Juneau Wang², Karen Tian², Brian Maniscalco³, Jennifer Motzer², Diana Gamboa², Angela Shen³, Minh Nguyen³, Meera Sriram³, Sydney Liu³, Olenka Graham Castaneda³, Maggie Zhang³, Vanessa Wu³, Richard Brown^{4,5}, Victor Lamme⁶, Hakwan Lau⁷, Biyu He¹, Jan Brascamp⁸, Ned Block¹, David Chalmers¹, Megan Peters³, Rachel Denison²; ¹New York University, ²Boston University, ³University of California, Irvine, ⁴LaGuardia Community College, ⁵The Graduate Center, CUNY,

Background: Figure-ground processing is a critical early step of object and scene perception. Although attention has been suggested to enhance figure-ground processing, there is limited neuroimaging work testing how attention influences visual cortical activity during figure-ground texture segmentation. In an fMRI experiment (27 participants with four sessions each), we tested whether spatial attention cues boost figure-ground processing in early visual cortex. **Methods:** Participants viewed four texture-defined figure-ground oval stimuli in each quadrant of the screen (3x4°, 5° from fixation, 250 ms duration). Textures were composed of parallel line elements tilted $\pm 45^\circ$. Oval figures within the texture were defined by orthogonal line orientations relative to the background. Stimuli were independently balanced across trials to be figure-present or figure-absent, have vertical or horizontal oval orientation, and display at one of five thresholded strength levels, based on the length of the texture line elements. Participants were pre-cued on each trial to attend a given quadrant and report the orientation of the oval in that quadrant, along with their subjective visibility of the oval (behavioral results to be reported elsewhere). **Results:** Given findings that figure-ground processing shows a center-surround structure, we grouped voxels based on their mean signal across conditions—either responding more to the figure than the background (figure-positive voxels) or more to the background than the figure (figure-negative voxels). We evaluated how attention affected BOLD responses within these two groups of voxels, comparing trials in which the oval figure was present vs. absent. In figure-positive voxels, attention increased responses to the figure relative to the background, whereas in figure-negative voxels, attention reduced responses to the figure relative to the background. **Conclusions:** Our results suggest that attention facilitates figure-ground processing by both enhancing the figure and suppressing the ground within visual cortex.

Templeton World Charity Foundation 0567 to BH, JB, NB, DC, MP, and RD, startup funding from the University of California Irvine and support from the Canadian Institute of Advanced Research to MP, startup funding from Boston University to RD, and BU UROP funding to DG.

23.430 CONTROL OVER CONSCIOUS PERCEPTION THROUGH MEDITATION: EVIDENCE FROM BINOCULAR RIVALRY

Heleen A Slagter¹ (h.a.slagter@vu.nl), Surya Selvam², Brendan T Hutchinson³, Daphne Witmer⁴, Tess Roder⁵, Johannes J Fahrenfort⁶, Ruben E Laukkonen⁷; ¹Vrije Universiteit Amsterdam

Predictive processing theories propose that perceptual experience is strongly shaped by prior beliefs. Some priors are particularly resistant to change, as illustrated by bistable phenomena such as binocular rivalry—where distinct stimuli presented to each eye produce alternating percepts rather than a stable mixed image. Contemplative

traditions suggest that meditation can attenuate the influence of such priors, potentially expanding the scope of conscious perception. In this study, we examined whether volitional control over rivalry is possible via meditation-induced attention modulation. Advanced meditators ($n=26$) viewed flickering rivalry stimuli under three conditions: focused attention (FA), open monitoring (OM) and a no-meditation control. Each condition included self-report, no-report, and localizer blocks. We hypothesized that FA, by stabilizing attention on the current percept, would prolong individual percept durations, whereas OM, by fostering non-selective attention to all aspects of experience, would increase mixed percept durations. This, in turn, should reduce self-reported perceptual switches in both meditation conditions, albeit for different reasons. Perceptual switches were quantified using behavioral self-reports and changes in perceptual content were tracked over time by combining two EEG analysis methods: pattern-classification applied to rhythmic entrainment source separated (RESS) time series. As hypothesized, behavioral reports revealed fewer switches in both meditation conditions, with some meditators reporting exceptionally long percept durations (up to 180s). First EEG analyses show high decoding accuracy and strong correspondence between neural and self-reported switches in the no-meditation condition, supporting the feasibility of objective rivalry-tracking. Strikingly, initial EEG-RESS decoding findings confirm exceptionally long percept durations in some meditators. Full sample results will illuminate the extent to which control over conscious perception is possible through meditation.

23.431 DIFFERENTIAL MODULATION BY SPATIAL ATTENTION ACROSS THE VISUAL HIERARCHY IN AUTISM SPECTRUM DISORDER

Brenda (Siyue) Qiu¹, Kristin M. Woodard¹, Bridget Leonard¹, Hannah M. Rea², Sara J. Webb^{1,3}, Scott O. Murray¹; ¹University of Washington, ²West Virginia University, ³Seattle Children's Research Institute

Sensory hypersensitivity is frequently reported in Autism Spectrum Disorder (ASD), yet neural evidence in early visual cortex is mixed. Studies have reported increased (Brieber et al., 2010), typical (Dinstein et al., 2012), or reduced (Kolodny et al., 2020) V1 responses in autistic individuals. It has been proposed that reduced V1 activity may be attributed to enhanced suppressive feedback from downstream regions such as MT+ (Kolodny et al., 2020), suggesting that atypical responsivity in ASD may reflect differences in how higher-order areas modulate early signals rather than alterations in V1 encoding itself. To evaluate this, we used fMRI and a spatial attention paradigm to examine responses in V1 and higher-order, attention-modulated regions in 50 non-autistic and 24 autistic participants performing a motion discrimination task. Following an attention cue, participants covertly attended to one of two bilaterally presented drifting Gabor gratings at 0% (stimulus absent), 6%, or 50% contrast. Participants judged the motion direction of the attended grating. Accuracy did not differ between groups. Consistent with prior work, non-autistic participants showed an additive effect of attention in V1

and MT+ (Murray, 2008). Here, we found that autistic participants demonstrated the same pattern. Additive attentional modulation was also observed in IPS and FEF for both groups. However, group differences emerged at different levels of the visual hierarchy. In V1, non-autistic participants showed stronger attentional modulation than autistic participants, despite comparable overall response magnitude. In contrast, MT+, IPS, and FEF showed similar attentional modulation across groups, but autistic participants showed higher responses at 50% contrast, reflecting increased responsivity to strong sensory drive. The emergence of this effect in higher-order regions, but not in contrast-sensitive V1, supports the notion that hypersensitivity in autism may arise from later integrative processes where sensory input and top-down signals interact, rather than from altered early sensory encoding.

This work was supported by funding from the National Institutes of Health (R01MH131595 to S.J.W. and S.O.M.)

23.432 ATTENTION IMPROVES PROBABILISTIC REPRESENTATION IN THE HUMAN VISUAL CORTEX

Yuqi You¹, Shurui Li², Yuanning Li², Hsin-Hung Li³, Ru-Yuan Zhang¹; ¹Shanghai Jiao Tong University, ²ShanghaiTech University, ³The Ohio State University

How does the brain utilize attention to enhance the efficiency of information processing? The Bayesian brain hypothesis posits that the brain combines noisy sensory information with prior knowledge to construct a posterior distribution over stimulus value and uses the posterior to guide actions. Although attentional enhancements on visual behaviour are well-established, it remains unclear how attention improves probabilistic neural representation of visual stimuli. Here, we analyzed an fMRI dataset (3T; N = 8) where face stimuli appeared randomly at one of sixteen locations tiling the visual space (4 x 4 grid, 2° spacing) while human participants performed two tasks in different runs: attending to centrally presented digits (digit task) or attending to the face stimuli (face task). We investigated how stimulus position is represented probabilistically in trial-by-trial multivariate cortical responses and how attention modulates these representations. By building a probabilistic voxel-encoding model linking stimulus position and multivariate cortical responses, we decoded the posterior distribution over position from fMRI responses in single trials. We found that, compared to the digit task, the face task markedly reduced both spatial bias and uncertainty according to the decoded posterior in several visual regions, with the strongest effects in hV4. These findings provide direct evidence for attentional enhancement on probabilistic population codes. By parametrically manipulating generative and inference (decoding) models, we further showed that changes in both voxel tuning and covariance are sufficient and necessary to account for the observed reductions in bias and uncertainty. Finally, trial-by-trial fluctuations of decoded uncertainty and its attentional modulation in hV4 correlated with activity in distributed attention-orienting and self-monitoring networks (e.g., parietal cortex and anterior cingulate cortex). Together, our findings bridge attentional control and

probabilistic inference, two canonical cortical computations in the brain, and provide new evidence for how top-down modulation optimizes stimulus coding.

Supported by the National Natural Science Foundation of China (32441102 to R.-Y. Z., 32371154 to Y.L.), Shanghai Municipal Education Commission (2024AIZD014) to R.-Y.Z. and Shanghai Rising-Star Program (24QA2705500) to Y.L..

23.433 DOES THE LINK BETWEEN SPONTANEOUS MICROSACCADES AND EEG ALPHA ACTIVITY DEPEND ON EXTERNAL VISUAL INPUTS THAT MOVE ACROSS THE RETINA WITH EACH MICROSACCADE?

Freek van Ede¹ (freek.van.ede@vu.nl), Baiwei Liu; ¹Vrije Universiteit Amsterdam

We have previously reported how spontaneous microsaccades – made in the absence of any incentive to look anywhere but a central fixation marker – are associated with the transient lateralisation of posterior 8-12 Hz EEG alpha-band activity according to microsaccade direction (Liu et al., *Progress in Neurobiology*, 2023). We now ask to what extent such a link between spontaneous microsaccades and EEG-alpha activity depends on visual inputs. If the reported link reflects microsaccades directly, it should be observed irrespective of the concurrent presence or absence of visual inputs. In contrast, if the link reflects a secondary ‘visual consequence’ of microsaccades because microsaccades displace visual inputs across the retina, then the reported link should be contingent on the presence of visual inputs. To address this, we recorded spontaneous microsaccades and EEG activity across three fixation conditions: (1) in the absence of any fixation marker or other visual inputs (i.e. after an initial fixation marker disappeared), (2) in the presence of only a central fixation marker, or (3) in the presence of a central fixation marker together with peripheral visual stimuli. Critically, transient alpha lateralisation following microsaccades was observed even in the condition without the fixation marker and the peripheral stimuli. These data argue for an intrinsic link between the direction of spontaneous microsaccades and the spatially corresponding lateralisation of posterior alpha-band activity.

This research was supported by an ERC Starting Grant from the European Research Council (MENTICIPATION, 850636) and an NWO Vidi Grant by the Dutch Research Council (grant number 14721) to F.v.E.

Functional Organization of Visual Pathways: Cortical visual processing 2

**SATURDAY, MAY 16, 8:30 AM – 12:30 PM,
PAVILION**

23.434 COMPUTATIONAL ADVANTAGES OF PATHWAY SPECIALIZATION FOR OBJECT RECOGNITION AND VISUALLY-GUIDED GRASPING

Aida Mirebrahimi¹, David C. Plaut¹, Marlene Behrmann^{1,2}; ¹Carnegie Mellon University, ²University of Pittsburgh

Decades of neuroscience and psychology evidence support a division of the visual cortex into dorsal occipito-parietal and ventral occipito-temporal pathways. Ventral regions are commonly linked to object recognition, whereas dorsal regions are closely associated with spatial processing and visually-guided action, suggesting a functional dissociation between these pathways. Yet, the computational motivations for such dual-stream architecture remain unclear. We introduce a computational framework to test whether an early-shared, late-modular dual-pathway architecture provides a measurable advantage over a single unified pathway for jointly supporting object recognition and visually guided grasping—behaviors traditionally linked to ventral and dorsal streams, respectively. To do so, we trained capacity-matched single- and dual-pathway convolutional neural networks on a new dataset of ~450 naturalistic 3D objects spanning 15 subordinate categories across three superordinate categories. Objects were rendered from 250 viewpoints with transformations in position, scale, and rotation, yielding ~150,000 images. Images include a category label and geometry-based grasp endpoints for joint training on recognition and grasping. We found that late-stage pathway separation accelerated grasp learning without compromising categorization ability. Across three increasingly challenging out-of-distribution tests, the dual-path multitask network (DP_MT) reached grasp thresholds in significantly fewer epochs than a capacity-matched single-path multitask model (SP_MT), while matching a categorization-only baseline in learning rate and final accuracy. At convergence, post-threshold gain analyses showed that both multitask models achieved larger grasp improvements than a grasp-only baseline under milder OOD shifts (novel viewpoints and novel instances), but under the strongest OOD shift (novel categories) only DP_MT maintained this advantage. Task-gradient analyses provide a mechanistic account: DP_MT's semi-modular organization mitigates late-layer cross-task interference while preserving early shared representations benefiting both tasks. These results quantify the computational benefits of late-stage segregation, offering a rationale for two-stream organization in primate vision and concrete design principles for artificial systems that must jointly support recognition and action.

23.435 FEATURE-BASED PRINCIPLES OF CATEGORY SELECTIVITY IN HUMAN VISUAL CORTEX AND DEEP NEURAL NETWORKS

Leonard E. van Dyck^{1,2}, Katharina Dobs^{1,2}; ¹Justus Liebig University Giessen, ²Center for Mind, Brain and Behavior, Universities of Marburg, Giessen, and Darmstadt

Category-selective responses are a hallmark of human visual cortex, most prominently for faces, bodies, and scenes. However, it remains unclear how widespread category-selective processing is across object space and which object properties best account for it. To address these questions, we analyzed fMRI responses in inferior temporal cortex (ITC) of three participants viewing images from hundreds of object categories in the THINGS database. Using a shared response model, we derived a common feature space across participants and introduced a sparseness-based selectivity metric quantifying how strongly each category engages selective features. This approach confirmed robust selectivity for faces, bodies, and scenes under broad category sampling, but also revealed previously unknown selectivities. Applying the same framework to high-level representations of deep neural networks (DNNs), we found only moderate correspondence between category selectivity rankings in ITC and in DNNs. To identify the object properties underlying these rankings, we fit an encoding model using behavior-derived object dimensions to predict each category's selectivity. In ITC, category selectivity was explained by a small set of primarily semantic dimensions, whereas in DNNs it was captured by a larger set of mostly visual dimensions. Together, our findings show that category-selective processing extends across object space beyond classical domains and is organized by interpretable, feature-based principles. These principles point to a fundamental dissociation between biological and artificial visual systems: visual cortex prioritizes ecologically important categories, whereas current computational models emphasize visually distinctive ones.

23.436 IN SILICO MAPPING OF VISUAL CATEGORICAL SELECTIVITY ACROSS THE WHOLE BRAIN

Hossein Adeli¹, Ethan Hwang¹, Wenxuan Guo¹, Andrew Luo², Nikolaus Kriegeskorte¹; ¹Columbia University, ²University of Hong Kong

A fine-grained account of categorical selectivity in the cortex is essential for understanding visual processing in the brain. Classic experimental studies have identified multiple category-selective regions; however, these approaches rely on preconceived hypotheses about categories. Subsequent data-driven methods have sought to address this limitation by building encoding models that predict neural activity from different image features, but they are often limited by simple, typically linear, mapping models. We propose an in silico approach for data-driven discovery of novel category-selectivity hypotheses. The brain is first divided into 1000 functional regions. We then train an encoding model that incorporates a brain-region to image-feature cross-attention mechanism, enabling efficient mappings between deep network features and semantic patterns in the brain activity. The model significantly outperforms the linear mapping across the cortex in predicting the fMRI responses in the Natural Scene Dataset. Using this encoder, we use diffusion-based image generative models and large-scale datasets (e.g. ImageNet) to synthesize and select images that maximally activate each region. These images are

then used to form a label (by averaging their CLIP representations) to characterize the selectivity. We validate that our method assigns correct labels for each region by testing that label-image-similarity predicts how well the images activate the region in held-out sets and that those images match selectivity for known areas (e.g. face, place, and word areas). We then apply the method to all the other regions in the brain that were visually responsive but had no known categorical selectivity. Our approach reveals regions with complex, compositional selectivity involving diverse semantic concepts (e.g. tool use), which we validate both within and across subjects. Our work shows that using a brain encoder as a “digital twin” offers a powerful, data-driven framework for generating hypotheses about selectivity in the human brain, which can be tested in future experiments.

Research reported in this publication was supported in part by the National Institute of Neurological Disorders and Stroke of the National Institutes of Health under award numbers 1RF1NS128897 and 4R01NS128897. The content is solely the responsibility of the authors.

23.437 FREQUENCY-TAGGED FMRI WITH SIMULTANEOUS VIDEO STIMULATION REVEALS SUPPRESSIVE FACE-PLACE INTERACTIONS IN VISUAL CORTEX

Chelsea Kim¹, Reebal Rafeh², Taylor Schmitz², Marieke Mur^{1,3};
¹Department of Psychology, Faculty of Social Science, Western University, ²Department of Physiology and Pharmacology, Schulich School of Medicine & Dentistry, Western University, ³Department of Computer Science, Faculty of Science, Western University

Everyday vision involves multiple visual objects competing for attention, yet most neuroscience studies still rely on static, isolated stimuli. Here, we emulated naturalistic vision using simultaneously presented dynamic videos of faces and scenes. We used a recently developed frequency-tagged fMRI method (ft-fMRI; Ngo, Rafeh et al. 2024) to tag visual cortical BOLD responses synchronized to either face or place stimulation frequencies. We examined if object-object competition produces mutual suppression of face- and place-responsive regions in the visual cortex. Three adults freely viewed face videos (multiple identities, neutral/positive expressions) and place videos (e.g., kitchen, stairs, bedroom) in a 3T MRI scanner. Luminance contrast was modulated at 0.1 Hz for face videos and 0.125 Hz for place videos, allowing us to temporally disentangle responses in the frequency domain. Videos appeared centrally (face-center, place-center) or in a face-center-place-surround configuration; place videos were also shown in isolation in the surround. Each participant viewed each condition for eight minutes. Our ft-fMRI paradigm evoked oscillatory BOLD responses at the tagged frequencies. Centrally presented stimuli activated the expected face- and place-selective regions in lateral and ventral temporal cortex. Face videos also drove a lateral temporal social visual pathway (Pitcher & Ungerleider, 2021). When faces and places were presented simultaneously, face-evoked and place-evoked responses were reduced compared to isolated presentation, consistent with mutually suppressive competitive

interactions between category-responsive pathways. We demonstrate that ft-fMRI provides access to fine-grained cortical maps of competing object representations under naturalistic viewing. This opens the door to investigating cortical computations subserving object-object competition and attention across the visual hierarchy and in lateral temporal pathways relevant to social cognition. Because the method leverages free viewing and frequency tagging, it holds strong potential for studying multi-object processing in infants and other challenging populations, providing a powerful new tool for naturalistic visual neuroscience.

NSERC

23.438 HOW DOES THE MACAQUE BRAIN DETECT OPTOGENETIC PERTURBATION OF INFEROTEMPORAL NEURONS?

Behnam Karami^{1,2}, Alvin Dinh¹, Reza Azadi¹, Arash Afraz¹;
¹Laboratory of Neuropsychology, National Institute of Mental Health, NIH, Bethesda, MD, USA, ²Donders Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behavior, Radboud University, Nijmegen, The Netherlands

The inferotemporal (IT) cortex is the final stage of visual processing in the primate ventral stream. While IT responses to visual objects have been extensively characterized, far less is known about how direct stimulation of IT neurons modulates perception. In previous work, we showed that behavioral detection of focal IT stimulation depends on the image shown to the monkey, with detection performance lowest when no image is present. We next investigated the neural mechanism underlying this image dependence. We considered two hypotheses: (1) recurrent dynamics within IT suppress perturbations when global activity is near baseline, effectively nullifying optogenetic drive in the no-image condition; or (2) neurons are activated by optogenetic pulses, but in the absence of concurrent visual input, the evoked activity fails to reach the detection threshold for downstream decoders. To enable optogenetic control, we injected pAAV-CaMKIIa-C1V1(t/t)-TS-EYFP into IT cortex, expressing an excitatory opsin in pyramidal neurons. A macaque monkey was trained to maintain fixation while viewing object images or a blank screen. Each trial lasted 1 s and was separated by a 200 ms interval. On 75% of trials, 200 ms optogenetic pulses of varying intensity were delivered 400-600 ms after onset via a 120 μ m Doric optical fiber integrated with a 24-channel Plexon V-probe, enabling simultaneous stimulation and recording. Optogenetic stimulation produced similar increases in IT spiking for image and no-image trials, contradicting hypothesis (1). However, ROC analysis showed a variable true positive rate across images, with the lowest value in the no-image condition (74% vs. 92%), supporting hypothesis (2): detection depends on a downstream thresholding mechanism, and without visual drive, IT stimulation alone is less likely to cross it. At least under passive viewing, these results suggest that a global readout, rather than local recurrent control, gates the impact of selective IT perturbations.

This research was supported by the Intramural Research Program of the NIMH ZIAMH002958.

23.439 AN EXPANDED TRIPLE-N DATASET REVEALS HIERARCHICAL DYNAMICS AND CROSS-SPECIES CORRESPONDENCE FROM V1 TO IT

Pinglei Bao¹ (pbao@pku.edu.cn), Yipeng Li¹, Xieyi Liu¹, Jia Yang¹; ¹Peking University

The initial release of the Triple-N (Non-human Primate Neural Responses to Natural Scenes) dataset targets high-level vision, focusing on inferotemporal (IT) cortex (Li et al., 2025), leaving the upstream stages that feed IT much less well characterized under matched conditions. Here, we introduce a new component of the Triple-N dataset that centers on early to intermediate visual areas and uses it to quantitatively link macaque V1, V2, and V4, human visual cortex, and deep convolutional neural networks. Using Neuropixels recordings in two macaques, we targeted V1, V2, and V4 based on each animal's anatomy. Across 19 sessions, we obtained 8,794 visually responsive units with split-half reliability larger than 0.4 while the animals passively viewed 1,000 shared NSD images. With this dataset, we first confirmed that temporal dynamics and representational structure evolve systematically along the ventral stream. Response onset and peak latencies shifted earlier from IT to V4, V2, and V1, confirming a robust temporal hierarchy under naturalistic stimulation. Encoding models based on a deep convolutional network revealed a clear layer-by-layer mapping: V1 and V2 were best predicted by early convolutional layers, V4 by intermediate layers, and IT by deep fully connected layers. More interestingly, receptive fields estimated from spike-triggered averages of local contrast showed orderly, stable spatial tuning within each session. Correlating macaque recordings with human fMRI responses to the same stimuli further revealed an organized cross-species correlation map, with high correspondence in both visual hierarchy and retinotopic organization. Taken together, the expanded Triple-N dataset delivers a continuous, experimentally matched set of neural measurements from V1 through IT, allowing quantitative alignment with human visual cortex and deep networks. Recordings from V1, V2, and V4 thus provide the essential bridge from early feature processing to the high-level representations observed in IT.

23.440 EVIDENCE THAT VENTROLATERAL PREFRONTAL CORTEX IS NOT A READOUT OF THE VENTRAL STREAM BUT REFINES IT REPRESENTATIONS FOR HARDER IMAGES

Maren Wehrheim^{1,2}, Kohitij Kar¹; ¹York University, Department of Biology and Centre for Vision Research, Centre for Integrative and Applied Neuroscience, Toronto, Canada, ²Mila - Quebec AI institute, Montreal, Canada

Causal evidence shows that the ventrolateral prefrontal cortex (vIPFC) is necessary for core object recognition: vIPFC inactivation degrades behavioral performance and IT-based object decodes, with larger

impairments for late-solved, challenging images. Yet, we still lack a computational account of how vIPFC participates in the ventral-stream hierarchy. Does vIPFC act as a downstream readout of IT, or does it supply recurrent, image-specific signals that help IT refine object representations, particularly for challenging images? These hypotheses make distinct predictions. A pure readout model holds that vIPFC should (i) produce object decodes more aligned with behavior than IT, and (ii) contribute no additional task-relevant structure beyond what is already represented in IT. A recurrent-enhancer model instead predicts that (i) vIPFC will be less behaviorally aligned than IT, (ii) vIPFC activity will explain IT variance in an object-solution-time (OST)-specific manner, and (iii) removing vIPFC-derived components from IT should impair decoding, especially for late-solved images. To adjudicate between these models, we recorded from IT (437 neurons) and vIPFC (234 neurons) in two macaques performing object recognition on 1,320 images. As expected, early-solved images (OST 90–110 ms) yielded rapid, high-accuracy IT decodes, whereas late-solved images (>180 ms) yielded slower, weaker IT evidence. Contrary to the readout model, vIPFC decodes were not more behaviorally aligned than IT, and vIPFC contributed substantial image-dependent explained variance (~10% of IT responses), markedly larger for late-OST images. Residual analyses showed that removing vIPFC-predictable components from IT responses reduced decoding performance, with strongest impairments for late-solved images, demonstrating that vIPFC provides behaviorally relevant structure not present in feedforward IT. Noise-corrected dimensionality analyses further showed that vIPFC constrains IT representations toward task-relevant manifolds for late-solved images. Together, these results position vIPFC not as a downstream readout, but as a selective recurrent partner that shapes IT population structure during challenging object recognition.

MW is funded by Connected Minds Postdoctoral Fellowship (supported by CFREF). KK was supported by the Canada Research Chair Program (CRC-2021-00326), Canada First Research Excellence Funds (VISTA Program), & the National Sciences and Engineering Research Council of Canada (NSERC, RGPIN-2024-06223).

23.441 EVOLUTIONARY REFINEMENT OF PRIMATE THALAMOCORTICAL PROJECTIONS THROUGH CONSERVED MOLECULAR GRADIENT

Angela Fan^{1,2}, Maryam Majeed³, Huihui Qi⁴, Noah Kuehn^{1,5}, Jack Scott¹, Sabrina Cheng^{1,6}, Huiqing Zhan⁷, Justus Kebschull⁴, Xiaoyin Chen³, James Bourne¹; ¹National Institutes of Health, ²Monash University, Clayton, Australia, ³Allen Institute, ⁴Johns Hopkins University, ⁵Georgetown University, ⁶University of Washington at Seattle, ⁷Cold Spring Harbour Laboratory

The primate cortex has expanded significantly across evolution, giving rise to primate-specific areas through duplication and specialization. In comparison, all major thalamic nuclei are conserved from rodents to primates and maintain connectivity to the whole cortex. These distinct

evolutionary strategies create a paradox: how do the rapidly duplicated and specialized cortical networks remain functionally integrated with subcortical hubs that did not duplicate in parallel? Here we investigate the evolution of thalamocortical wiring rules by integrating single-cell and spatial transcriptomics with high-throughput single-cell connectivity mapping in both the marmoset and mouse thalamus. We sequenced 1.5 million neurons from the marmoset thalamus in-situ, and mapped the projections and transcriptomic identities of 480 thalamocortical neurons in the marmoset, and 1750 thalamocortical neurons in the mouse. In both species, we found continuous gene expression gradients that span across thalamic nuclei, suggesting a gradient-like organization that goes beyond nucleus borders. In contrast to the conserved transcriptomic architecture, projections in the primate thalamus project more focally and topographically compared to their mouse counterparts. Surprisingly, the more precise projections in the marmoset are associated with conserved gene expression axes that are also predictive of projections in the mouse thalamus, suggesting that the increased specificity in primate projections arises through sharpening of ancestral gene-projection relationships. We propose that this refinement of existing connectivity architecture represents an efficient strategy to accommodate projections to evolutionarily expanded cortical areas without requiring reorganization of relatively conserved subcortical hubs.

23.442 RETROGRADE VIRAL TRACING OF AFFERENT CONNECTIVITY TO V1 AND V4 IN NONHUMAN PRIMATES

Francesca Sciacotta¹, Carlos Mejias-Aponte¹, Martin Bohlen², Adam Snyder³, Ian Fiebelkorn³, Farran Briggs^{1,3}; ¹National Eye Institute, National Institutes of Health, ²Duke University School of Medicine, ³Center for Visual Science, University of Rochester

Primary visual cortex (V1) and extrastriate area V4 occupy central positions in the primate visual hierarchy, supporting early feature encoding and intermediate-level shape and color processing, respectively. Here, we evaluate retrograde viral tracing approaches to map afferent projections to V1 and V4 in macaque monkeys, with the dual goal of characterizing their anatomical connectivity and comparing the performance of different viral vectors. We performed targeted injections of retrograde viruses into V1 and V4 and analyzed labeled neurons across cortical and subcortical regions. Preliminary results reveal widespread projection patterns across multiple animals, with dense and distributed corticocortical inputs spanning early visual, extrastriate, and higher-order cortical areas. Across both target regions, cortical projections substantially outnumber subcortical inputs, including those from pulvinar. In parallel, we are directly comparing lentiviral and retrograde AAV-based approaches in macaque cortex. Early observations suggest broadly similar large-scale projection patterns across virus types, while quantitative differences in labeling efficiency and distribution are under active investigation. Ongoing analyses will provide systematic quantification across animals, target areas, and viral vectors. Together, this work provides new comparative data on afferent connectivity to V1 and V4

and informs methodological choices for retrograde circuit mapping in the primate visual system.

Face and Body Perception: Models

SATURDAY, MAY 16, 8:30 AM – 12:30 PM,

PAVILION

23.443 RECURRENT PROCESSING RESHAPES TUNING OVER TIME IN DEEP NEURAL NETWORKS

Amirhossein Farzmaidi¹, Hossein Adeli¹, Wang Boran¹, Chase King¹, Nikolaus Kriegeskorte¹; ¹Zuckerman Mind Brain Behavior Institute, Columbia University

The primate ventral visual stream is rich in recurrent connections. Unlike feedforward models, recurrent networks change their internal representations over time, supporting dynamic processing of visual information. Here we ask: how does recurrence transform neural network representations across hierarchy and time during naturalistic object recognition? We trained a 6-layer recurrent convolutional neural network with bottom-up and lateral connections to recognize different face identities. Using a controlled stimulus set of faces and objects with systematically varying pose (yaw, pitch, roll) and identity, we probed the model's internal representations across layers and recurrent steps at both population and single-unit levels. Early convolutional layers showed largely stable representations over time, indicating that recurrence did little to change low-level feature tuning. In contrast, deeper layers exhibited pronounced tuning shifts: as recurrence unfolded, representations transitioned from supporting coarse face detection to fine identity discrimination, accompanied by increased invariance of face identities to in-depth and in-plane pose changes. At the same time, object representations became less distinctive and less invariant under the same transformations. Importantly, recurrent processing in deeper layers produced tuning shifts at the single-unit level: for example, narrowing from many faces to a smaller set of face identities. RISE-based importance maps suggested a mechanism for these shifts: the diagnostic image features and face parts driving each unit's response also changed over time (e.g., moving from eyes to mouth). Together, these unit-level effects reveal genuine tuning shifts, rather than simple sharpening of a fixed preference, underlying the coarse-to-fine transition at the population geometry. These results show that recurrent processing can flexibly reshape representational tuning over time, especially in deeper layers of a network. The global-to-specific transition and increasing face invariance replicates neurophysiological findings in the primate IT cortex and suggest that recurrent dynamics are a key mechanism for achieving brain-like visual recognition.

This work was supported in part by the National Institute of Neurological Disorders and Stroke, National Institutes of Health, under award RF1NS128897.

23.444 COMPUTATIONAL MODELING OF THE PERCEIVED FACIAL TYPICALITY AND ATTRACTIVENESS OF DISTINCT FACE POPULATIONS

Logan Trujillo¹ (logant@txstate.edu); ¹Texas State University

Previous computational modeling research investigating the mental representation of faces suggests that the positive relationship consistently observed between the perceived typicality and attractiveness of young adult Caucasian female faces may be best explained by a representational format where faces are encoded in memory with respect to a hypothetically perfect representative of a face population (ideal face). This study extends this prior research by modeling the mental representation of young adult Caucasian, African American, and East Asian female faces. Human judgments of typicality, attractiveness, and pairwise similarity were gathered for 17 faces from each face group. Perceived typicality and attractiveness judgments were then predicted using four computational models expressed within a perceptual embedding space (“face space”) derived from the similarity ratings via multidimensional scaling: the Ideal Dimension Model (IDM), in which typicality and attractiveness reflect the similarity of a face to the ideal face of a face population; the Generalized Context Model (GCM), in which typicality and attractiveness reflect the similarity of a face to other faces encoded in memory; the Central Prototype Model (CPM), in which typicality and attractiveness reflect the similarity of a face to the central tendency of a population of faces; and the Varying Abstraction Model (VAM), in which typicality and attractiveness reflect the similarity of a face to clusters of related faces (“subprototypes”) within a face population. Face typicality for all three face groups was best explained by the IDM as quantified by Bayesian metrics of model generalizability. The attractiveness of the Caucasian and African American faces was also best explained by the IDM, whereas the VAM best explained East Asian facial attractiveness. These findings suggest that facial typicality is mentally represented by humans via an ideal format; however, facial attractiveness may be mentally represented using different formats depending on the face population.

This work was partially supported by NIH grant #HD021332.

23.445 HOW VISUAL FAMILIARITY AND PERSON KNOWLEDGE SHAPE THE COMPUTATIONAL DYNAMICS OF FACE RECOGNITION

Raphaël Fournier^{1,2}, Paul E. Downing³, Richard Ramsey^{1,4}; ¹Neural Control of Movement Laboratory, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland, ²Neuroscience Center Zurich, University and ETH Zurich, Zurich, Switzerland, ³Department of Psychology, Bangor University, Bangor, United Kingdom, ⁴Social Brain Sciences Laboratory, Department of Humanities, Social and Political Sciences, ETH Zurich, Zurich, Switzerland

The ability to quickly and accurately recognise a familiar face is an important aspect of everyday life, as it guides social behaviour. The way that one might interact with a family member or friend is different

to a stranger. Moreover, for familiar individuals, we not only know what they look like, but we also have access to person knowledge, such as their traits, attitudes, name and professional status. To date, most prior face recognition research has used manifest (observed) measures, such as response times and accuracy, to assess one's ability to recognise faces. As such, there is currently a rather limited understanding of the computational processes that support face recognition processes. Using evidence accumulation models (EAM), we investigated how different types of face familiarity (visual familiarity and stored person knowledge) impacted the signal-to-noise ratio (drift rate parameter) and response caution (threshold parameter) during a face memory task. We demonstrated that different computational processes support the recognition of different types of faces. In Experiment 1, visually familiar faces produced a lower signal-to-noise ratio (i.e., drift rate) than novel faces. In contrast, in Experiment 2, the combination of visual familiarity and semantic knowledge about a face induced a higher signal-to-noise ratio compared to novel faces. Given that visual familiarity is associated with the “core” face perception network in the ventral visual stream, whereas semantic knowledge is associated with the “extended” face perception network in more anterior brain regions, these results give some tentative insight into the computational processes that take place in different parts of the face processing network.

23.446 MAPPING THE REPRESENTATIONAL GEOMETRY OF SUBJECTIVE FACE IDENTITY DISCRIMINATION WITH A GENERATIVE FOUNDATION MODEL

Veronica bossio botero¹, Nikolaus Kriegeskorte; ¹Columbia University, Zuckerman Institute, ²Columbia University, Zuckerman Institute

Humans are experts at face recognition, yet the underlying perceptual space that drives subjective identity judgements remains to be characterized quantitatively. Deep image-generative models can synthesize photorealistic face images from latent embeddings in representational spaces. Such models offer a powerful new framework for studying face perception, operating as flexible stimulus generators and computationally explicit candidate models for the perceptual space itself. Here, we leverage Arc2Face—a diffusion-based generative model—to characterize the geometry of subjective face space and individual variability in face discrimination. We used Arc2Face to generate morph-lines radiating from 240 reference identities, each representing a trajectory of identity change in latent space. In a behavioral task, 40 participants judged whether pairs of faces along these trajectories depicted the same identity. We modeled these judgments using a hierarchy of probabilistic models to test hypotheses about the properties of human perceptual face space: its uniformity (is sensitivity invariant to location in face space?), isotropicity (is sensitivity invariant to direction of change?), and interindividual variability (do subjects exhibit idiosyncratic sensitivity profiles?). We established that a model based solely on the angular distance between embeddings is highly predictive, capturing 76% of the explainable variance in human judgments. Model comparison

revealed that perceptual space is non-uniform and anisotropic, with subjective identity thresholds varying significantly depending on reference face and direction of change. Importantly, interindividual variability extended beyond differences in global sensitivity (larger or smaller thresholds for perceiving a match). An interaction model allowing unique observer-stimulus interactions significantly outperformed one that assumes observers differ only in their response criterion, demonstrating observers have stable, idiosyncratic sensitivity patterns to specific facial features unexplainable by global bias. These findings quantitatively characterize the geometry of perceptual face space. We introduce a validated tool for predicting human identity judgments and provide evidence that face discrimination relies on idiosyncratic, observer-specific facial cues.

23.447 MATHEMATICALLY TRACTABLE NEURAL FIELD MODEL FOR ACTION-PERCEPTION COUPLING

Martin A. Giese¹, Xinrui Jiang¹; ¹Hertie Institute / CIN, University Clinic Tuebingen, Germany

Visual perception is modulated by concurrent action execution, and action execution can be modulated by concurrent action vision. This mutual dependence can be captured by dynamically coupled neural representations for perceived actions in the visual pathway, and for planned actions in the motor pathway. The computational mechanisms that link such dynamic representations are poorly understood, and popular data-driven large-scale neural network models, due to their complexity, do not allow a deeper mathematical analysis of the relevant interactions. METHODS: We present a dynamic neural model for the coupling between action perception and action execution of hand movements. It is based on coupled nonlinear dynamic neural fields / neural mass models. Input movies are mapped onto inputs of vision fields by a feed-forward deep neural network that accomplishes hand pose recognition on sequences of real images. This level is coupled to the second dynamic level of the model, which is based on motor fields that represent hand actions in terms of self-generated propagating activation patterns. We study how such representations can be learned by time-dependent Hebbian learning, and we mathematically analyze the dynamics of the individual and the coupled representations. RESULTS: The model reproduces several experimental results on perception-action coupling within a unifying framework (influence of action execution of visual action detection, influence of concurrent action vision on accuracy of action execution, influence of baseline delay on the perception of delay changes between action perception and execution). In addition, it provides a mathematical understanding of why these interactions occur, and how they might emerge through learning. CONCLUSIONS: Neural mass models are meaningful tools to understand dynamic interactions in multi-level dynamic cortical representations in the perception-action loop. Opposed to data-driven standard neural network models they allow an understanding of the underlying dynamical mechanisms.

The work was funded by ERC 2019-SyG-RELEVANCE-856495. The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Xinrui Jiang.

Decision Making: Perception 1

**SATURDAY, MAY 16, 8:30 AM – 12:30 PM,
PAVILION**

23.448 EVIDENCE FOR WINNER-TAKE-ALL DYNAMICS IN LONG-LASTING VISUAL INTEGRATION

Marie Holdsworth¹ (marie.holdsworth@epfl.ch), Maëlan Q. Menétrey^{1,2,3}, Can Oluk¹, Michael Herzog¹; ¹Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, ²Psychophysics and Neural Dynamics Lab, Department of Radiology, Lausanne University Hospital (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland, ³The Sense Innovation and Research Center, Lausanne, Switzerland

Conscious perception is preceded by long-lasting windows of unconscious processing. This is evident in the Sequential Metacontrast Paradigm (SQM), in which vernier offsets are integrated over several hundred milliseconds along a motion-stream of lines. When one line contains a vernier, all other lines are perceived as having the same offset. When the stream contains two verniers with opposite offsets, they integrate mandatorily, i.e., observers cannot distinguish the individual offsets. Previous work suggests that offsets are unconsciously summed and then read out as the conscious percept; under this model, two opposite offsets are predicted to cancel each other. Here, we aim to understand the phenomenology (or subjective experience) of the integrated percept. Using a continuous-response method (mouse adjustment), participants reported the size and direction of the perceived offset. Consistent with prior results using a two-alternative forced-choice method, participants reliably reported an offset in the single-vernier condition. In conditions with two opposite offsets, we expected an increase in the straight-line reports. However, participants reported offsets in either direction with almost equal probability, and the perceived size of these offsets was similar to that in the single-offset condition. When several offsets were shown in the same direction, participants reported larger offsets compared to single-offset conditions, in line with the model's prediction of offset-size summation. To test the limits of these integration mechanisms, we designed another experiment in which smaller and spatiotemporally closer offsets in opposite directions were interleaved. Even then, straight-line responses did not increase, nor did the perceived offset size differ from control conditions. Overall, these findings suggest that if two offsets are in opposite directions, a winner-take-all mechanism operates and only one offset is consciously perceived. However, if offsets are in the same direction, their sizes

sum. Thus, conscious perception results from a dynamic interplay between competitive and summative processes.

23.449 THE INFLUENCE OF REWARD MAGNITUDE AND PROBABILITY ON THE SELF-PRIORITIZATION EFFECT: A STUDY BASED ON THE PERCEPTUAL MATCHING PARADIGM

CHAOYING DAI¹, Yang Sun¹; ¹Shenyang Normal University, Liao Ning, China

Abstract Self-associated stimuli typically enjoy prioritized access within visual processing, suggesting that self-relevance functions as a high-value signal within the perceptual priority map. Yet it remains unresolved whether external reward value alters this intrinsic prioritization—or whether the visual system treats self-related information as a value source independent of incentive cues. To address this, we conducted five perceptual matching experiments manipulating reward magnitude (high/medium/low) and reward probability (high/low), and examined how these value parameters modulate evidence accumulation and attentional gain for self- vs. other-associated stimuli. Across studies, we replicated a robust Self-Prioritization Effect (SPE), reflected by faster and more accurate responses to self-associated stimuli. Crucially, modulation of SPE emerged only when reward magnitude and probability were jointly considered. Neither dimension alone was sufficient to alter perceptual priority, indicating that a single value input does not override baseline self-weighting in visual selection. However, when expected value was equivalent, high-magnitude/low-probability reward selectively increased the attentional gain of self-related stimuli relative to other stimuli. Moreover, probability effects surfaced only under high-magnitude conditions, suggesting that reward value influences perceptual evidence weighting only when motivational strength surpasses a threshold—consistent with a non-linear or gate-like integration mechanism. These findings support the idea that self-relevance establishes a strong baseline in the visual priority map, while reward magnitude provides an additive value signal that can amplify perceptual weighting. In contrast, reward probability contributes only under sufficiently high motivational drive, indicating that visual value integration is not purely multiplicative but conditionally modulated. This work refines our understanding of value-driven visual selection and reveals how intrinsic (self) and extrinsic (reward) value jointly shape early perceptual prioritization. **Keywords:** self-prioritization effect, value-driven attention, reward magnitude, reward probability, perceptual matching

Shenyang Normal University

23.450 DESIGNERS CHOOSE TO SHOW DATA VISUALIZATIONS THAT OBSERVERS CAN RELIABLY INTERPRET

Kushin Mukherjee¹ (kushinm@stanford.edu), Holly Huey², Lauren A Oey³, Judith E Fan^{1,4,5}; ¹Department of Psychology, Stanford University, ²Department of Psychology, University of California, San

Diego, ³Department of Psychology, Princeton University, ⁴Department of Computer Science, Stanford University, ⁵Graduate School of Education, Stanford University

Data visualizations are used to efficiently communicate patterns in quantitative data. While substantial work has investigated which visualization properties affect observers' ability to make fast and accurate judgments about data (e.g., using scatterplots to expose correlations rather than grouped bar plots), less work has investigated which factors visualization designers consider when trying to anticipate what would be helpful to an observer. Are designers more likely to follow simple heuristics (e.g., "Use scatterplots because they are more informative than barplots.") or do they choose different strategies depending on the question at hand? English-speaking adult participants (N=119; mean age: 38.87 years) performed a visualization design task where they decided which graph to show an observer to help them answer a question about a dataset (e.g., "What is the average departure delay of LaGuardia flights leaving at 2pm-6pm?"). Building on preliminary work (Huey et al., 2023), we devised scenarios where participants chose between different graph types (i.e., bar, line, scatter) and how many variables to show (i.e., between 3-5). We found that participants' choices were far from uniform ($\chi^2(8)=1760.40$, $p<.001$), and that simple heuristics, such as a persistent preference for fewer variables or for a specific graph type, failed to explain these patterns. Instead, participants chose different visualizations depending on which question they were presented with, and they were more likely to choose visualizations that an independent sample of observers (N=1,229; mean age: 21.14 years) found useful for accurately answering that question. These results are consistent with the possibility that participants make their choices by simulating how another person would answer the question, given a visualization. More broadly, such studies suggest that people make flexible use of their own visualization understanding abilities to communicate information effectively to others.

This work was supported by NSF CAREER Award #2047191, NSF DRL award #2400471, and a Stanford Hoffman-Yee HAI Grant to JEF

23.451 THE INFLUENCE OF UNCERTAINTY ON SERIAL DEPENDENCE

Natalya Phillips¹, Luis D. Ramirez¹, John T. Serences¹; ¹University of California San Diego

Neural processing capacity is finite, necessitating efficient and predictive mechanisms for rapidly processing sensory information. One such mechanism is serial dependence, a phenomenon where perception is biased toward recent stimuli, especially under uncertainty. Here we investigated whether uncertainty, induced by altering contrast or orientation precision, influences the magnitude and tuning of serial dependence. Because contrast selectively alters neural response strength and orientation precision selectivity alters the dispersion of neural responses, we hypothesized that contrast primarily modulates the strength of serial dependence, and orientation

precision primarily modulates tuning (width). We manipulated contrast or orientation precision within blocks of a 2AFC orientation discrimination task (orientation filter widths: 2°, 40°, 80°; contrasts: 90%, 50%, 25%). On every trial, an oriented bandpass-filtered noise stimulus (uniformly-sampled filter width or contrast level) was followed by a line that had an independently and uniformly-sampled orientation offset (7 offsets, linearly-spaced from 0–15°). Participants (n=11, each subject tested for 3–5, 1.5 hour sessions) reported whether the line was clockwise or counterclockwise from the filtered stimulus (2160–3510 trials per subject). Response bias was estimated with a Gaussian cumulative density function within bins of stimulus history (difference between the previous and current trial's stimulus orientation; 32° sliding window) and for each condition (9 combinations of contrast and orientation precision levels). To quantify serial dependence, we estimated the amplitude and width of bias with a derivative of Gaussian function for every combination of uncertainty. We found that the magnitude of attraction decreased with certainty (though less so for contrast). Moreover, the width of attraction decreased with orientation precision, but increased with contrast. Our findings suggest that serial dependence is primarily influenced by current sensory evidence and less so by the strength of evidence on previous trials, with mechanistic differences under different types of uncertainty.

This research was funded by National Institutes of Health Grant R01EY025872 to J.T. Serences and supported by K00EY036804 to L.D. Ramirez and R25NS119707 to N.K. Phillips

23.452 FUNCTIONAL ADVANTAGE OF SERIAL DEPENDENCE ACROSS MULTIPLE LEVELS OF VISUAL PROCESSING

Fiammetta Marini¹ (fiammetta.marini1@abdn.ac.uk), Meike Ramon^{2,3}, Julia Föcker^{4,5}, Mauro Manassi¹; ¹School of Psychology, University of Aberdeen, King's College, Aberdeen, UK, ²Applied Face Cognition Lab, Business School, Bern University of Applied Sciences, Bern, Switzerland, ³AIR – Association for Independent Research, Zürich, Switzerland, ⁴Danish School of Education, Aarhus University, Aarhus, Denmark, ⁵College of Health and Science, School of Psychology Sport Science and Wellbeing, University of Lincoln, Lincoln, UK

In serial dependence, our actions, perception, decisions, and memories are pulled towards recently seen stimuli. This bias has been proposed as a purposeful mechanism, enhancing efficiency by merging past and present visual input. In fact, initial evidence in orientation perception showed that serial dependence between consecutive similar orientations leads to higher precision (i.e., reduced variability) in current perceptual judgements. Recently, this functional advantage of serial dependence has been questioned, as the same increase in precision has been observed even when successive stimuli were largely different and serial dependence was absent. To unravel these seemingly contrasting findings, we analysed multiple datasets from adjustment tasks related to orientation (N=45), position (N=72), shape (N=42), facial expression (N=41), and face identity (N=45) stimuli to test the consistency of these results. First, we found evidence

for serial dependence across all the tested stimuli. Second, we observed evidence for improved efficiency in perceptual judgements at small stimulus differences across all tested datasets. Crucially, only datasets with orientation and position stimuli revealed a precision increase at larger stimulus differences, whereas no efficiency improvement was found for shapes and faces. Based on these findings, we propose that the efficiency advantage at larger stimulus differences is strictly related to the observer's knowledge of the stimulus space and not to serial dependence itself. When observers have a clear representation of the stimulus space (orientation or position stimuli), they exhibit enhanced efficiency in recognising orthogonal stimuli. Conversely, when this clear representation of the stimulus space is absent (shape or face stimuli), this specific efficiency advantage does not occur. Overall, these results confirm that serial dependence improves efficiency across multiple levels of visual processing independently of concomitant stimulus-related biases, further reinforcing the idea of serial dependence as a purposeful, beneficial mechanism that enhances the efficiency of vision.

This research was supported by the UKRI ESRC Research Grant ES/Z503381/1.

23.453 DIFFERENT EFFECTS OF INTERNAL REPRESENTATIONS ON DECISIONS AND EYE MOVEMENTS

Jozsef Fiser^{1,2}, Saleha Siddiqui^{3,4}, Barnabas Molnar^{1,2}, Rana Banai-Tizkar^{3,4}, Frederic Chavane^{3,4}, Anna Montagnini^{3,4}; ¹Central European University, ²Center for Cognitive Computation, ³Institut de Neurosciences de la Timone, ⁴Aix-Marseille Université

Every conscious and involuntary action we make is strongly influenced by our internal model of the monetary environment and by the actual context. In our previous investigation of perceptual decision-making, we used response biases to identify which alternative internal representation observers used during their decisions. We found that, depending on the representation they adopted, observers could make opposite choices and, in particular, could follow decisions that appear irrational from the standpoint of a “standard” internal representation. In the present study, we explored whether the internal representation that sets the bias of conscious decisions also controls anticipatory eye movements, or conscious and involuntary actions are governed by different internal models. We used a modified version of the standard 2-AFC decision-making paradigm in which observers (N = 20) judged the direction of random-dot kinetograms moving either right or left. First, we identified each observer's direction-sensitivity threshold with an adaptive staircase method, and then, in the main phase of the experiment, we used randomly chosen coherence levels on each trial. Importantly, when transitioning from threshold measurement to randomized trials, we also changed the appearance probability of the right/left motion directions from 50/50% to 80/20%, thereby imposing a general bias in the probability of motion across trials. Confirming our earlier results from a shape-classification task, observers' decisions under these conditions showed a strong bias toward selecting the less

frequent direction during low-coherence trials. This decision bias is long-lasting (over hundreds of trials) and contradicts the common expectation that observers should choose the more frequent direction. In contrast, measurements of anticipatory eye movements showed a significantly stronger bias toward the more frequent direction. Thus, the internal representations governing conscious decisions and the mechanisms of eye-movement control are at least partially dissociable.

This work was supported by Grant ANR-FWF I 6793-B

23.454 REPEATED ENCODING OPPORTUNITIES MITIGATE BELIEF-BASED BIAS IN CORRELATION ESTIMATION

Adam Malitek¹, Minsuk Chang², Cindy Xiong Bearfield², Keisuke Fukuda^{1,3}; ¹University of Toronto, ²Georgia Institute of Technology, ³University of Toronto Mississauga

The accurate interpretation of statistical data is critical in the modern information landscape, yet evidence suggests that objective information evaluation is compromised by pre-existing beliefs. For instance, pre-existing belief biases individuals' correlation estimates of a scatterplot to be more consistent with that belief (Xiong et al., 2022). Recently, we showed that gaze pattern predicted the magnitude of this belief-based bias only early on (i.e., the first 1.5 seconds) during scatterplot viewing. This suggests that there exists a limited time window (i.e., critical period) during which the visual information can interact with the pre-existing belief to form a mental representation of the scatterplot (Malitek et al., 2025). Therefore, we tested the hypothesis that belief-based bias can be mitigated by increasing the number of critical periods that individuals are exposed to. In Experiment 1 (N = 64), participants viewed eight belief-laden scatterplots (e.g., X: commute duration, Y: stress level of commuters) for five seconds each. Importantly, four scatterplots were presented three times across the experiment (repeated viewing condition) while the remaining four were shown only once (single-viewing condition). Here, we found that the belief-based bias was significantly reduced as the number of encoding opportunities increased, especially for those whose belief was comparatively weak. Critically, by keeping the total viewing time constant between the repeated and single-viewing conditions, Experiment 2 (N = 72) confirmed that this bias reduction was due to the increase in the number of critical periods rather than the total viewing time. Taken together, our results highlight the importance of the number of encoding opportunities in mitigating belief-based bias and provide insight into how we can promote bias-free interpretation of visualized data.

This research was supported by the Natural Sciences and Engineering Research Council (5009170).

23.455 VISUAL AND MOTOR BEHAVIOUR FOR OBJECT MANIPULATION IN A DYNAMIC ENVIRONMENT

Markus Solbach¹, Mohit Goyal², Sakar Khattar², Jayant Varma¹, Bjorn Vlaskamp³, Konstantine Tsotsos⁴, John Tsotsos¹; ¹York University, Toronto, Ontario, Canada, ²Google, Mountain View, California, USA, ³Google, Seattle, Washington, USA, ⁴Google, Toronto, Ontario, Canada

Functional vision relies on active gaze to acquire task-relevant information, yet how the visual system adapts to environmental instability remains poorly understood. This study investigates adaptive strategies during a complex 3D assembly task under dynamic constraints. We recruited 90 participants to replicate 3D magnetic block models on a central turntable rotating at 0, 5, 10, and 15 RPM. Participants were monitored under "Free" or "Restricted" mobility conditions using synchronized binocular eye-tracking and high-fidelity motion capture to record objective performance, oculomotor metrics, and 6D head-hand kinematics. Results reveal a critical interaction between mobility and workload. At lower speeds (5 RPM), the "Free" condition provided a clear performance advantage with faster response times. However, at the highest workload (15 RPM), this trend reversed: The quality of performance observed solution methods degrades as speeds increases. Conversely, the "Restricted" condition demonstrated greater stability, maintaining consistent efficiency even as difficulty increased. Longitudinal analysis indicates that mobility significantly influences adaptation ($p < .001$). Performance was best at 0 RPM, and participants were allowed to move "Free." Generally, the "Free" group showed better learning compared to the "Restricted" group. However, head translation revealed that while movement remained consistent at lower speeds, for the 15 RPM condition, a drastic increase appeared in cumulative head translation ($> 35m$) for "Free" moving participants. Additionally, oculomotor analysis revealed a processing shift: fixation counts increased linearly with speed while saccade amplitude contracted ($p < .001$), implying a transition from global scanning to localized tracking. We demonstrate that the benefit of physical mobility is not universal. While allowing participants to move generally aids adaptation, it becomes counterproductive during extreme dynamic tasks. By quantifying visual behaviour in a complex 3D environment, we move beyond simply observing that performance declines to documenting exactly what fails.

This research was supported by grants to the senior author (John K. Tsotsos) from the Air Force Office of Scientific Research USA and the Natural Sciences and Engineering Research Council of Canada and Google Canada.

23.456 WHEN THE WAIT MATTERS: DIFFERENCES IN COGNITIVE ABILITY CONTRIBUTIONS TO SIMULTANEOUS VS. SEQUENTIAL VISUAL DISCRIMINATION

Yoonsang Lee¹ (yoonsang.lee@gatech.edu), Jason Tsukahara², Randall Engle¹; ¹Georgia Institute of Technology, ²University of Miami

Although visual discrimination is often treated as a low-level perceptual task, many forms of discrimination—especially those requiring

comparisons across time—likely depend on higher-order cognitive mechanisms. In particular, sequential discrimination may draw on the same attentional-control and maintenance processes that support individual differences in working memory capacity (WMC) and fluid intelligence (gF). To examine whether different forms of visual discrimination rely on distinct cognitive abilities, 150 participants completed a battery measuring gF, WMC, and attention control (AC). A subset of 112 participants additionally completed both a sequential and a simultaneous line-length discrimination task. Structural equation modeling revealed a clear dissociation: sequential discrimination was predicted by both WMC and AC, with little contribution from gF, whereas simultaneous discrimination showed its strongest association with WMC and only a weak relationship with gF once AC was included. These patterns suggest that discrimination tasks differ substantially in their cognitive demands depending on how stimuli must be compared. More broadly, the results demonstrate that the link between perceptual discrimination and higher cognition depends on task format, and that separating maintenance-based WMC from executive and sustained-attention components of AC is essential for understanding individual differences in perceptual fidelity.

ONR Multidisciplinary University Research Initiative (MURI)

23.457 SENSORY PROCESSING SENSITIVITY IN BISTABLE AND CUED MOTION PERCEPTION

Rocio Asperilla¹, Liubov Ardasheva², Anna Montagnini³, Pilar Aivar⁴;
¹Universidad Autónoma de Madrid, ²Aix-Marseille University

Sensory Processing Sensitivity (SPS) may influence how individuals integrate sensory information, yet its role in controlled perceptual decision-making remains largely unexplored. Here, we examined how SPS—as measured by the Highly Sensitive Person Scale—interacts with stimulus ambiguity and cognitive priors during motion discrimination of an ambiguous, bistable rotating sphere. Across three experiments with independent participant samples, we manipulated stimulus ambiguity (variants of the Kinetic Depth Effect stimulus) and expectations about motion direction (probabilistic cues), while recording behavioral responses and high-precision eye movements (not analyzed here). In Experiment 1 (continuous 3-min bistable viewing), SPS correlated positively with perceptual switch frequency (Spearman $\rho = .47$, $p = .002$), indicating that individuals higher in SPS experienced more frequent alternations. Pearson correlations were nonsignificant, suggesting a nonlinear association. In Experiment 2 (trial-based unbiased direction judgments), mixed-effects models showed robust effects of stimulus ambiguity on both perceptual choices and reaction times (Ambiguous > Mid-Ambiguous > Non-Ambiguous; all $p < .001$). SPS influenced response promptness: higher-SPS individuals responded faster and more consistently when stimuli were less ambiguous. Within-trial switches were strongly reduced for non-ambiguous stimuli ($p < .001$) and did not interact with SPS. In Experiment 3 (priming with 50/70/90% cue validity), prior cues biased perceived direction but did not reliably affect reaction times or switching. Ambiguity again dominated perceptual stability, and SPS

did not modulate cue-induced biases. These findings provide initial evidence linking SPS to individual variability in bistable motion perception, with greater reactivity to ambiguous sensory input but not to externally imposed probabilistic priors. Given the modest sample sizes and the subjective nature of SPS assessment, interpretations should remain cautious. Nonetheless, the results identify a novel connection between trait sensitivity and perceptual decision-making.

23.458 PULLING THE LEVR: LEARNING ABOUT RISK IN A VISUAL FORAGING TASK

Hana Taha¹, Thomas W. James¹; ¹Indiana University Bloomington

A relatively new development in decision science involves studying decisions from experience. In this context, risk is particularly interesting to study, because learning can lead to preferences among risk conditions, but risk can also shape learning. To investigate the interplay between risk and learning, we systemically manipulated risk, expected value, and skew of resource distributions in a foraging task called Learning about Expected Value & Risk (LEVR). The task used a four-factor design: EV (-4, +4), risk level (low, medium, high), skew type (negative, no skew, positive), and grouping (One-Skew Multi-Risk, One-Risk Multi-Skew). Across 240 trials, subjects chose between six options that combined these factors. For One-Risk, options for each subject group varied skew type, but were all the same risk level. For One-Skew, options for each group varied risk level, but were all the same skew type. EV4 choices were picked more than EV-4 choices across most subject groups, indicating significant learning. Learning was most consistent for One-Skew, suggesting that choosing among options with different risk levels aided learning. In contrast, for One-Risk, learning was much greater for the low risk level group than the medium or high-risk level groups. This was supported by a significant 3-way interaction between risk, grouping, and learning. Grouping also influenced preferences among risk levels, with high-risk level choices picked most of all, specifically in the One-Skew condition. Finally, there were no significant effects of skew in the results. Overall, the grouping together of multiple risk levels emerged as the most influential factor in shaping choice behavior. We suspect that grouping of risk provided environmental scaffolding for learning. In the One-Skew Multi-Risk group, learning about low-risk level options may have aided learning of high-risk level options, which wasn't evident in the One-Risk group.

23.459 DON'T USE D' FOR VISUAL DETECTION: RT DATA WORK FOR UNEQUAL-VARIANCE SIGNAL DETECTION ANALYSIS

Kiyofumi Miyoshi¹ (miyoshi80@gmail.com), Dobromir Rahnev², Hakwan Lau^{3,4}; ¹Kyoto University, ²Georgia Institute of Technology, ³Institute for Basic Science, South Korea, ⁴Sungkyunkwan University

In visual detection tasks, participants often exhibit distinct behaviors between trials with and without target presence, commonly manifested as asymmetric type-1 receiver operating characteristic (ROC) curves. Such asymmetry suggests greater signal variability in target-present

trials, violating the equal-variance assumption fundamental to the classic d' metric. Here, we examined the utility of response time (RT) data for implementing unequal-variance signal detection theory (SDT) analysis as a cost-effective alternative to the traditional confidence-based approach. Across eleven datasets, RT- and confidence-based analyses yielded highly consistent estimates of the sensitivity metric d_a —an unequal-variance extension of classic d' . Furthermore, the d' metric systematically overestimated detection performance relative to d_a , underscoring the need to incorporate unequal-variance characteristics into the analysis. We also found that the degree of this overestimation varied with observers' response criteria: because unequal-variance ROCs expand more on the low-false-alarm side, adopting a conservative criterion can spuriously inflate d' . Taken together, these findings indicate that RT-based unequal-variance SDT analysis provides a reliable means of assessing visual detection performance, especially in situations where collecting confidence ratings is impractical.

This work is supported by JSPS KAKENHI Grant Number 22K13870 and 25K00896, awarded to KM. HL is supported by the Institute for Basic Science, South Korea (Grant Number IBS-R015-D2).

23.460 THE CONJUNCTION FALLACY IN PERCEPTION: THE EFFECT OF CONTEXT

Barnabás Molnár^{1,2} (molnar_barnabas@phd.ceu.edu), József Fiser^{1,2}, Ádám Koblínger^{1,2}; ¹Central European University, ²Center for Cognitive Computation

The conjunction fallacy is a well-known breakdown of Bayesian reasoning, but its cognitive roots remain unclear. Most studies rely on ambiguous, story-based vignettes that leave room for interpretation. Less is known about its emergence in simpler, perception-based tasks with tightly controlled visual stimuli. We aim to explore this gap in two controlled experiments. In Experiment 1, we tested how people learn probabilistic rules involving abstract visual features in a continuous stream. Participants ($N = 41$) judged the probability of individual and combined features, as in traditional conjunction-error tasks. Individuals could identify typical features for trial types with specific traits, suggesting that representativeness was learned. Conjunction errors were rare; moreover, varying the representativeness of a feature did not significantly affect the fallacy rate, even though representativeness has been argued to be the leading cause of the fallacy. In Experiment 2, we used a more complex stimulus space defined by color, shape, and pattern. Participants ($N=56$) estimated single and combined feature probabilities using ranking or slider methods. Error rates similar to those observed in the original tasks emerged only under high comparison load, when multiple statements were ranked simultaneously or when participants had to maintain earlier estimates to stay consistent. However, when participants assessed a single conjunction and its constituent, fallacy rates dropped. Task order also had a significant effect: starting with the probability assessment lowered fallacy rates in the sorting task. Together, these experiments offer new insights into the conjunction fallacy in perception-based

learning tasks. We found that the conjunction errors in perceptual tasks arise less from representativeness per se and more from constraints imposed by task structure. These findings support the theory of bounded rationality and show that fallacy rates can be reduced by presenting information in a visual sequence, thus reducing cognitive load.

Action: Navigation, locomotion

SATURDAY, MAY 16, 8:30 AM – 12:30 PM,
PAVILION

23.461 ADAPTING TO CHANGES IN ACTION CAPABILITIES DURING VISUALLY GUIDED STEERING

AJ Jansen¹, Brett Fajen¹; ¹Rensselaer Polytechnic Institute

Action capabilities pose critical constraints on how people move through the world. While running, biking, or driving, humans can only accelerate so quickly and turn so sharply. Previous work shows that people can accurately perceive whether actions are within their capabilities and adjust their behavior during continuous control tasks to accommodate their limits. These findings raise questions about how people initially learn and continually relearn their ever-changing action capabilities. When action capabilities are enhanced such that new, potentially more efficient solutions become possible, do people spontaneously switch to those solutions or stick with well-practiced solutions? To address this question, we conducted two experiments using a multiple-waypoint steering task performed in a virtual environment. Subjects guided a drone through two gates at a fixed distance on the centerline and a third gate at a variable distance and angle. Each block comprised 18 trials. The drone was sluggish in Blocks 1-5, gradually transitioned to agile in Block 6, and remained agile in Blocks 7-11. With a sluggish drone, the task could be performed most successfully by following a curved trajectory to the second gate that anticipated the third gate. With an agile drone, a more direct path was both possible and more time efficient. Nevertheless, no subjects immediately switched to the direct-path solution when it first became possible (Block 7) and only some subjects converged toward a direct-path solution by Block 11. In Experiment 2, we instead used an obstacle-avoidance task for Block 6 to test whether performing a different task during adaptation would promote switching solutions. As in Experiment 1, subjects tended to stick with the indirect-path solution that they learned when the drone was sluggish. These findings suggest that when action capabilities change, people may (at least initially) stick with well-practiced solutions even if more time-efficient solutions are available.

NSF 2218220

23.462 LEARNING THE AFFORDANCE LANDSCAPE FOR VISUALLY GUIDED STEERING

Brett Fajen¹ (fajenb@rpi.edu); ¹Rensselaer Polytechnic Institute

Most visually guided locomotor tasks allow for a range of solutions. When steering through a series of waypoints, for example, an actor might move directly toward the nearest waypoint or take a more indirect path that better anticipates upcoming turns. Individual movement trajectories differ not only in how effectively they satisfy task demands, but also in their stability and controllability given the actor's movement capabilities (e.g., how quickly they can move, turn). Together, these factors define the affordance landscape; that is, the space of possible actions. By learning the affordance landscape, actors can adapt behavior to task demands while maintaining stable control within their limits. The aim of this study was to extend existing information-based strategies of visual control to model how an agent might learn the affordance landscape for a basic locomotor task: steering to a goal. The model agent's translational and rotational accelerations are regulated by a simple error-nulling control strategy based on visual information about goal-heading angle, rotation rate, and speed. The control strategy has parameters that the agent can tune to alter trajectory shape and enact different solutions. The agent's movements are also influenced by motor noise and subject to action limits that shape the stability and controllability of individual solutions. Importantly, the controllability of the current solution is specified by the time evolution of the control strategy output (acceleration), which converges to zero only when the agent is following a stable, controllable trajectory. This information allows the agent to gradually learn a function from currently available visual information and control-strategy parameters to the controllability of individual solutions; in effect, the affordance landscape. Model simulations capture key aspects of human steering behavior that simpler models cannot, including the ability to adapt to changes in action capabilities and motor noise and to coordinate steering and speed control.

NSF 2218220

23.463 GUIDING LOCOMOTION WITH A VIBROTACTILE BELT: SENSORY SUBSTITUTION FOR VISUAL-SPATIAL INFORMATION

William Warren¹ (bill_warren@brown.edu), Sofia Ruiz¹, Michael Vail¹, Benjamin Kimia¹, Sina Feldmann¹; ¹Brown University

Can sensory substitution be used to enhance the guidance of locomotion for people with impaired vision? In previous research with a vibrotactile belt (VSS 2025), we demonstrated high accuracy and precision in judging the egocentric direction of single-tactor vibrations. Here we test whether such vibrations can guide active walking to targets in virtual reality (VR). Participants (N=16) wore a Quest 3 VR headset that tracked head position and orientation, and a vibrotactile belt that contained 16 pager motors spaced 22.5° apart. On each trial, the participant walked to a sequence of two targets using either visual or tactile information. Visual targets appeared as green poles, while tactile targets were specified by single-tactor vibrations in 0.5s bursts at 225 Hz. The burst rate was either (a) constant, with a 0.5s pause between bursts, or (b) increased in 3 steps (0.5s, 0.1s, 0.01s pause) as the target was approached (3m and 1.5m). Target position was

manipulated, with eight egocentric directions (45° intervals) and two distances (4m, 8m). Walking performance was only moderately worse with tactile compared to visual information. Mean stopping distance from the final target was slightly larger for tactile (0.34m) than visual (0.15m) targets, $p < 0.001$, but there was no effect of tactile burst rate on stopping distance (constant 0.35m, increasing 0.33m). Mean tortuosity (normalized path length) was slightly greater with tactile (1.14) than visual (1.03) targets, $p < 0.001$. However, mean walking time to 2 targets was much longer with tactile (31.8s) than visual (19.3s) targets, $p < 0.001$. We are currently analyzing walking speed, path curvature, maximum path deviation, and energetic cost. The results indicate that tactile guidance increases path length by only 10% but walking time by 65% compared to sighted guidance; the latter may improve with training. These findings demonstrate the potential of the vibrotactile belt for guiding locomotion.

Supported by NIH R01EY029745

23.464 DIFFERENTIAL EFFECTS OF AGE IN A COGNITIVE-MOTOR DUAL TASK: A PARAMETRIC MANIPULATION OF COGNITIVE AND MOTOR LOAD

Catherine Reed¹ (clreed@cmc.edu), Chandlyr Denaro¹, Joseph Asta¹, Alan Hartley¹, Jenna Monroy²; ¹Claremont McKenna College, ²Natural Sciences, Claremont Colleges

Aging affects attention and executive control, making it more difficult to split focus between walking (motor) and thinking (cognitive), thereby increasing fall risk. Cognitive-motor dual tasks require simultaneous performance across tasks. Prioritization theory assumes shared resources between cognitive and motor tasks and predicts biomechanical stability is prioritized as motor difficulty and cognitive difficulty increase. However, few studies systematically vary both cognitive and motor difficulty to determine the conditions under which processes are prioritized. We investigated how cognitive and motor difficulty affect biomechanical and cognitive performance variables in younger and older adults. Participants walked at preferred speeds on a treadmill while performing auditory n-back working memory tasks. Cognitive load varied across 0-, 1-, and 2-back tasks. Motor load varied across 0-, 2.3-, and 4.6-degree inclines. Results showed cognitive performance was affected by biomechanical difficulty and cognitive load. For both groups, d-primes declined with increased load, but the decline was greater for older adults. Kinematics were affected by age, incline and cognitive load. Relative stride length and percent time in double support increased with incline, replicating effects of motor load on biomechanics. In older adults, kinematics changed with increased cognitive load particularly with the onset of incline, suggesting a prioritization of biomechanical stability. Our results suggest a non-linear relationship between cognitive load, biomechanical difficulty and performance, with younger and older adults reaching peak performance at different levels of combined cognitive and motor difficulty. Future studies on aging populations will investigate the allocation of resources while parametrically increasing both tasks.

23.465 THE DEVELOPMENT OF EYE-BODY COORDINATION DURING WALKING IN CHILDREN AGES 7-13 YEARS

Lauren Hoare¹, Krista Kelly¹; ¹*School of Optometry and Vision Science, University of Waterloo*

Introduction: Eye-body coordination is important for locomotion, with children shifting from relying largely on vision to integrating vestibular and sensorimotor cues around 7-9 years of age. Compared to adults, children 2-6 years slow down, look down more often, and move their eyes more while walking on uneven ground (Schroer & Hayhoe, 2025 VSS meeting). Here, we examine eye-body coordination during walking in conditions of increasing complexity in children 7-13 years. **Methods:** 28 typically-developing children were enrolled (13 younger, 7-9 years; 15 older, 10-13 years). Children completed 3 walking conditions: 1) Straight Walk (SW): walk on mat, 2) Isolated Target Walk (IT): walk and step on two-dimensional (2D) targets, and 3) Distractor Target Walk (DT): walk and step on 2D targets while avoiding 2D distractors. Gait outcomes recorded with the GAITRite pressure-sensitive walkway were normalized velocity, stepping accuracy, and coefficient of variability (COV). Eye movement outcomes recorded with the Tobii Glasses 2 were looking time and average number of fixations in areas of interest on the mat (nothing, target, distractor). **Results:** Younger children walked faster than older children in all 3 conditions ($p < 0.001$), and had higher COV (i.e., more variability) for normalized velocity ($p = 0.012$) and step width ($p = 0.016$) in the SW condition only. Younger children were also less accurate at stepping on the far target ($p = 0.012$), but not the near target or distractors ($p > 0.057$). During walking, younger children looked more ($p = 0.042$) and made more fixations ($p = 0.019$) in areas with targets or distractors compared to older children. **Conclusion:** Compared to older children, younger children walked faster, were more variable and less accurate, and made more fixations to targets and distractors in their path. This pattern of eye-body coordination may reflect more reliance on visual feedback during the development of walking as younger children build strategies for more efficient planning and execution.

NSERC Discovery Grant RGPIN-2024-03992 (KK), Canadian Optometric Education Trust Fund (LH)

23.466 DO CHILDREN ADAPT THEIR GAZE STRATEGIES WHILE WALKING ON REAL-WORLD CHALLENGING TERRAINS?

Sara E Schroer¹ (saraschroer@utexas.edu), Mary Hayhoe¹; ¹*University of Texas at Austin*

During natural walking, adults adapt their gaze and gait in response to the environment. On rough terrains, adults look at the ground close to their bodies, slow down, and take shorter steps. On all terrains, however, adults often use recent visual information (acquired $< 1.5s$ before a step) to guide walking. This information decay rate has been

demonstrated across many motor tasks. In our study, we examined whether young children use a similar strategy. 3- through 5-years-old children wore a head-mounted eye tracker while walking on three terrains: a Sidewalk, Stones (an unstable surface), and a narrow Wall. On the Sidewalk, only 31% of looks were on the path and these were an average of 3.9 steps ahead. Children's cadence was 2.4 steps/s, which suggests they may be acquiring information $\sim 1.6s$ in advance. On the challenging terrains, the proportion of fixations directed towards the path dramatically increased (82% on the Stones, 79% on the Wall; $p < 0.001$). Children looked closer to their bodies on the Stones (2.8 steps ahead) and the Wall (3.2 steps ahead; $p < 0.001$). On the Stones, children took shorter steps ($p = 0.03$) and decreased cadence by 10% ($p < 0.001$). Gait changes on the Wall varied – half of the children increased step length by 10.5% and cadence by 12.6% ($p < 0.001$); one third decreased step length by 14% ($p < 0.001$), but did not change cadence; and the remaining children did not change their gait. When children deviated from normal cadence – slowing down on the Stones or speeding up on the Wall – they looked $\sim 1.3s$ in advance. Children who did not change cadence on the Wall looked $\sim 1.4s$ in advance. These results suggest that children may need more recent information to guide walking on challenging terrains. As with adults, they flexibly adapt their behavior, but doing so requires different information for planning.

NIH R01 EY05729 and T32 EY021462

23.467 SPATIOTEMPORAL INTEGRATION OF MOTION INFORMATION IN HUMAN CROWDS

Jiayi Pang¹ (jiayi_pang@brown.edu), William H. Warren¹; ¹*Brown University*

The ability to perceive and coordinate locomotion in flocks, schools, and crowds is fundamental for social animals, including humans. A common example is collective motion, walking together with a crowd of pedestrians. To do so, individuals adjust their walking direction (heading) and speed to coordinate with multiple neighbors in the field of view. Most models of this 'flocking' behavior assume that visual information about the motion of all neighbors is averaged simultaneously, in a form of ensemble perception. However, there is evidence that some species only process one or two neighbors at a time. In this study, we investigate how visual information about multiple moving neighbors is spatially and temporally integrated by a participant walking with a virtual crowd. We evaluate three hypothesis: (1) Simultaneous sampling - the movements of all neighbors are integrated at once and averaged to produce a precise estimation of group motion; (2) Sequential sampling - neighbors are sampled one at a time, and successive samples are averaged over time; and (3) Subsampling - a subset of neighbors is sampled simultaneously, and successive samples are averaged over time. Participants followed a crowd of 9 neighbors viewed in a VR headset, while we recorded eye movements and changes in heading direction in response to brief perturbations of neighbors' heading. In Experiment 1, we identified a 0.5s perturbation as sufficient to elicit reliable heading responses by

manipulating both perturbation duration and the number of perturbed neighbors. In Experiment 2, we applied an equivalent noise analysis by manipulating the variability of neighbors' heading directions within this brief 0.5 s perturbation. The analysis indicated that all 9 neighbors were integrated simultaneously within a single sample, consistent with hypothesis 1. The results support ensemble perception in a crowd, at least up to 9 neighbors, consistent with simultaneous averaging models of human flocking.

23.468 THE ROAD AHEAD: KNOWLEDGE DOES NOT REPLACE VISION IN NAVIGATIONAL DECISION-MAKING

Cassandra Engstrom¹, William H. Warren¹; ¹Brown University

Previously, we found that humans walking through mazes reduce their angular deviation from the beeline to their goal, although they also reduce turns away from their current walking direction (heading) or body orientation. Simulations demonstrated that compared to other uninformed strategies, reducing beeline deviations generates the shortest paths. But does this strategy persist when choosing between routes of equal length? Participants (N=28) viewed displays of rectangular courtyards in a VR headset. Starting at one corner, they walked around the perimeter to a goal post at the opposite corner. If participants reduce deviations from the beeline, they should walk the long side first. To test the influence of heading, half the participants walked to the start corner on an 'on-ramp' placed at different angles. To test the role of body orientation, the other half began trials standing at the corner facing different directions. We found that neither group preferred the long side first (ns). However, both groups chose routes that minimized turns from their start orientation ($p < .001$). We considered two possible explanations for this unexpected result. (1) Because participants can see the paths ahead, they depend on the spatial layout and discount other visual variables that normally correlate with the shorter path, like beeline deviations. (2) Because the courtyard is rectangular, they learn that route lengths are equal and similarly discount visual variables. To disassociate these alternatives, we repeated the experiment ($n=30$) except that participants began each trial behind a wall with two doors. Only the goal post, but not the courtyard, was visible when they selected routes. Both groups now preferred the long side ($p < .001$), even if it meant turning away from their start orientation. These findings suggest that visual information and knowledge of the environment play distinct roles in wayfinding behavior, and that the former dominates route selection.

NSF 1405001060; T32MH115895 (NIMH)

23.469 RETINAL FLOW DYNAMICS AND GAZE BEHAVIOUR DURING HIGH-SPEED CYCLING

Kontessa Ioanna Zorpala¹, Joan Lopez-Moliner¹; ¹University of Barcelona (UB)

Although there are many studies addressing optic flow in the past, less is known about how gaze behaviour and optic-flow dynamics operate in natural and high-velocity scenarios, like cycling, where rapid

changes in heading, curvature and environmental conditions create different visuomotor demands. Our study aims to characterise the retinal information available during cycling and to determine how flow-based measures relate to the future path allowing for steering control. Participants ($n=4$) rode a paved, winding road (3 km, 6.5% average negative slope), while wearing a mobile eye-tracker (Neon pupil labs, 200Hz) with a world camera (30 Hz). Maximum speeds were between 50 and 70 km/h. Gaze, world-camera and GPS data were synchronised, enabling the reconstruction of the visual input associated with each location point. Optic-flow was computed in two ways: (1) directly from the world-camera and (2) from a reconstructed retinal-flow field, created by cropping each frame around the gaze location (Matthis et al 2022) resulting in head and retinal frame of reference respectively. Curl and Focus of Expansion (FOE) were extracted for both references, while road curvature was obtained from GPS data. Gaze in world coordinates was computed using eye-tracker and IMU data and was aligned with road coordinates. We used cross-correlation functions (CCFs) and obtained peak correlations and their temporal lags, between gaze direction, curl, FOE with respect to road curvature. Gaze showed a consistent anticipatory relationship to upcoming curves, indicating that cyclists orientated their gaze toward the turn-tangent before entering it. Both optic-flow measures correlated with road curvature, but retinal-flow curl displayed stronger and more temporally structured coupling, particularly for faster subjects. Retinal FOE-route correlations showed a similar pattern. Overall, the retinal frame of reference supplies robust predictive information for future-path estimation and steering, expressed through FOE, curl, or both, highlighting its contribution to high-speed cycling.

The research was supported by Grant PID2023-150081NB-I00 funded by MICIU/AEI/10.13039/501100011. KIZ was supported by fellowship PREP2023-001890.

23.470 INTRODUCING EYE-LEAN (LOCOMOTION, EXPLORATION, ACTION AND NAVIGATION WITH EYE TRACKING): A BEHAVIORAL RESEARCH TOOLKIT FOR DATA RICH VIRTUAL REALITY EXPERIMENTS

Jakub Suchojad¹, Kavindya Dalawella¹, Serena DeStefani², Karin Stromswold¹, Jacob Feldman¹; ¹Rutgers University, ²Ohio State University

The advent of affordable Virtual Reality (VR) Head-Mounted Displays (HMDs) has opened brand-new avenues for designing ecologically valid behavioral studies and collecting rich datasets from freely locomoting human subjects. Here, we present a comprehensive toolkit for scientists looking to utilize VR in their experiments, which includes both ready-to-use assets as well as a list of resources and best practices for designing VR experiments. We address the entirety of such a project, from creating the experimental environment inside the Unity game engine, through obtaining avatar models, all the way to animating them by using off-the-shelf motion capture data or recording your own. As a part of our setup, we provide a robust, C#-based eye tracking data collection and visualization system. Our approach

utilizes the open- source OpenXR standard and can be deployed on HMDs equipped with an eye-tracker that supports the standard. The eye tracking system allows for dynamic, vergence-based estimation and visualization of the current gaze point (based on Duchowski et al.), allowing proper calibration of the eye tracker. The OpenXR API allows for collection of detailed raw gaze characteristics data, which in turn permits robust secondary data analyses, including pupillometry. We provide code to replay the recorded gaze data inside Unity, allowing for informative visualization of gaze patterns. To illustrate the toolkit, we present selected results from several recent studies from our own lab and others demonstrating practical applications of our system, illustrating how one base paradigm can be adjusted to support behavioral studies across disciplines. These examples, which span navigation, decision making, and linguistics, highlight the advantages of our paradigm over existing solutions by allowing simultaneous locomotion and display of complex stimuli. We see our toolkit as useful both for young researchers newly starting behavioral work, and seasoned groups looking for novel research solutions.

NSF training grant NRT-FW-HTF 202162811 NSF BCS Grant 2324598 NSF BCS Grant 2122119 Rutgers Research Council Grant

23.471 HEAD-SCANNING DYNAMICS DURING STREET-CROSSING IN ULTRA-LOW VISION

Batuhan Erkat¹, Dinesh Venugopal¹, Roksana Sadeghi², Tran Chau³, Will Gee³, Brittnee Livingston⁴, Gislin Dagnelie², Arathy Kartha¹; ¹SUNY College of Optometry, Department of Biological and Vision Sciences, New York, NY, United States, ²The Johns Hopkins Wilmer Eye Institute, Baltimore, MD, United States, ³BaltiVirtual, Baltimore, MD, United States, ⁴Central Association for the Blind and Visually Impaired, Utica, NY, United States

People rely on vision to guide them in crossing streets safely, yet it remains unclear how severe visual impairment alters visual behaviors, including head-scanning strategies, that support this visual guidance. To address this gap, we examined horizontal head-scanning behavior in a virtual street-crossing task with systematically varied environmental demands conducted with individuals with ultra-low vision (ULV). Ten normally sighted adults (NV) completed the task under both normal and simulated ULV (sULV). Fourteen adults with clinical ULV completed the same task across four complexity levels: A (empty road), B (parked cars with pedestrians), C (moving cars that yield to pedestrians), and D (yielding moving cars and pedestrians). Measures included latency to initiate walking, total absolute angular displacement along the yaw axis (total horizontal scan), scanning speed, and a pre- versus in-motion scanning index, indicating the proportion of scanning occurring before versus during walking. Before crossing, sULVs showed greater total scan and longer latencies than NVs (respectively, $p = 0.004$; $p < 0.001$), with similar scanning speed ($p = 0.33$). During crossing, total scan and scanning speed did not differ significantly between the NV and sULV conditions ($p = 0.78$; $p = 0.35$), nor between the NV and ULV groups ($p = 0.50$; $p = 0.98$). However, we observed that higher environmental complexity

modulates total scanning. NV participants showed similar scanning index values across simple and complex conditions (-78% , $p = 0.24$), whereas both sULV and ULV showed decreased horizontal scanning during street-crossing in complex scenarios (-1.46 -fold, $p = 0.03$; -0.61 -fold, $p = 0.002$). Our results suggest that ULV participants exhibit more head-scanning before stepping off the curb and less during walking in visually demanding environments. By characterizing these behavioral differences, we identify specific behaviors that can be targeted through mobility training for individuals with ULV to support safer street-crossing.

R00EY033031 (AK), and R01EY028452 (GD), Research to Prevent Blindness (GD)

23.472 INCORPORATING ATTENTION-BASED WEIGHTS INTO THE SCRUM MODEL OF COLLECTIVE MOTION

Kei Yoshida¹ (kei_yoshida@brown.edu), Sina Feldmann¹, William H. Warren¹; ¹Brown University

When following a crowd, pedestrians combine information from multiple neighbors in their visual neighborhood to guide walking, consistent with distance-weighted averaging (the SCrUM model of Self-organized Collective Motion, Rio et al., 2018). Last year, we reported a real-crowd experiment in which four confederates attempted to steer or split a group of walking participants ($N=16-22$). When participants were unaware of the confederates (covert), they could steer but not split the crowd, consistent with weighted averaging. But when participants were instructed to follow confederates carrying flags (explicit), they generated both steering and splitting. To account for explicit leaders, here we extend the weighted-averaging account by incorporating attention-based weights into the SCrUM model. Modified models were tested using multiagent simulations, which took confederate trajectories as input and computed simulated participant trajectories. We then compared the simulated final headings with the human data. For covert confederates, the original SCrUM model reproduced the empirical patterns of steering (KS test, $p < .001$ for left and right) and not splitting (overlapping GMM components with means -0.86° and 0.24°). However, the model could not replicate the strong influence of explicit confederates. Forcing the confederate weights to 1 reproduced the large left and right turns observed with explicit confederates ($p < .001$ for left and right). But splitting only emerged when agents allocated their full attention to the nearest confederate and the weights of the other three confederates were set to zero (two GMM components with means -8.0° and 8.4°). Neighbors retained their distance-based weights, yielding coherent sub-group motion. These findings support distance-weighted averaging as a generally robust mechanism for collective motion, but the addition of attention-based weights is needed to explain collective behavior in scenarios with explicit leaders. The results imply that followers attend to a single nearest leader in a crowd.

Supported by NSF BCS-1849446, NIH 1S10OD025181

Action: Miscellaneous

SATURDAY, MAY 16, 8:30 AM – 12:30 PM,

PAVILION

23.473 A NEW STIMULI SET AND DESIGN FOR STUDYING WORKING MEMORY FOR VISUALLY PERCEIVED MOVEMENT TRAJECTORIES

Rajalakshmi Usha¹, Brad Wyble¹; ¹The Pennsylvania State University

Recreating observed movements is a critical part of learning dance, handicraft and sign languages. We report a new stimulus set and experimental design to study the nature of complex motor representations in working memory. Prior works typically used simple reaching movements or simple actions to demonstrate a motor-specific subcomponent within working memory (Smyth & Pendleton, 1990; Hillman et al., 2024; Wood, 2007), distinct from the visuospatial module. However because these studies used either highly discrete, repetitive motions or focused only on the spatial endpoint, the memories did not reflect complex patterns and may not have been motor plans per se. We created 30 unique, complex 2-D movement patterns containing curves and loops. Videos of a hand moving a computer mouse served as stimuli. Each video was also rotated to create a non-ego version, presenting the movement from an “other-person” viewpoint. Participants viewed each video and immediately recreated the movement using their own mouse. Regardless of viewing conditions, all reproductions were made from the participant’s own perspective. We predicted better performance in the ego condition as non-ego stimuli require mental rotation into an ego-centric motor plan. Twenty-five students ($N = 25$) from Pennsylvania State University participated in the experiment. We used dynamic time warping (DTW) in MATLAB to compare stimulus trajectories with participant recreations. After normalization, a paired-samples t -test ($\alpha = .05$) showed significantly better matches in the ego condition ($M = 209.96$) than the non-ego condition ($M = 236.2$), $t(24) = -2.544$, $p = .0178$, indicating better performance for ego stimuli. Future work using concurrent motor and visual tasks will test whether this difference reflects the involvement of two distinct subsystems—one encoding egocentrically viewed movement directly as motor plans and another requiring mental rotation in the visual working memory prior to motor planning for non-egocentric movement.

23.474 I’LL BELIEVE IT UNLESS IT’S TOO ABSURD: SPONTANEOUS VISUAL PERSPECTIVE-TAKING AS PRIOR-BASED HEURISTIC INFERENCE

Xucong Hu¹ (xuconghu@zju.edu.cn), Yitong Zheng¹, Qinyi Hu¹, Haokui Xu², Hui Chen¹, Mowei Shen¹, Jifan Zhou¹; ¹Zhejiang University, ²Zhejiang University of Technology

The mechanism underlying visual perspective-taking (VPT) remains contested: do people rely on perceptual simulation, reconstructing

what others see, or on heuristic inference grounded in naïve optics? Although recent evidence favors heuristic-account, most support comes from explicit VPT tasks in which participants report the agent’s view. Whether implicit (spontaneous) VPT—where participants perform an unrelated task yet are influenced by the agent’s viewpoint—relies on the same mechanism remains unclear. This matters because explicit tasks permit strategic top-down correction based on high-level knowledge, potentially masking underlying visual processes. Moreover, simulation-account may seem unsupported simply because observers lack access to the agent’s visual experience—yet providing such experience in explicit tasks would reveal the answer. We therefore tested whether prior visual experience could dissociate the two accounts in spontaneous VPT. In Experiment 1, participants received plausible, no, or implausible agent-view priors but completed a viewpoint-irrelevant line-length reproduction task. The spontaneous VPT effect was defined as a systematic bias in reproduction consistent with the agent’s view. The simulation-account predicts such a bias only under plausible priors, requiring viable sensory input; the heuristic-account predicts following any prior, even implausible, because it relies on abstract, knowledge-based rules. Results showed a bias for plausible priors, no bias for no prior, and a weak bias for implausible priors. A control experiment ruled out priming. The instability of the implausible-prior effect (ratio analyses yielded only marginal effects) suggests plausibility-based integration rather than blind adherence. Experiment 2 used a Ponzo-style stimulus with five levels of prior-plausibility. A pattern emerged: participants followed the prior when it was plausible or only moderately implausible, but abandoned it when highly implausible. This pattern reflects the bounded rationality of heuristic inference: observers integrate the provided prior with visual input while it remains within a plausible range, but revert to self-visual evidence once the “plausibility” threshold is exceeded.

This research was supported by the National Natural Science Foundation of China (32371088, 62337001, and 32471093), the Science and Technology Innovation 2030 —“Brain Science and Brain-like Research” Major Project (2022ZD0210800), and the Fundamental Research Funds for the Central Universities (226-2024-00118).

23.475 DISSOCIATING THE EFFECTS OF THE SEEN HAND FROM THE ACTING HAND IN MOTOR CONTROL USING VIRTUAL REALITY

Olivia Ipwansek^{1,2} (oiipwanish@uwo.ca), Karsten Babin², Paul L Gribble^{1,2,3,4}, Jody C Culham^{1,2,3}; ¹Department of Psychology, Western University, Canada, ²Centre for Brain and Mind, Western University, Canada, ³Neuroscience Program, Western University, Canada, ⁴Department of Physiology & Pharmacology, Western University, Canada

Virtual reality (VR) offers a unique opportunity to investigate how the brain integrates vision and action, as it allows the visual representation of the body to be manipulated in ways otherwise impossible in the real world. A central theoretical question is how the appearance of one’s

own hand shapes the planning and control of movement. We used a 3D VR reaching task to investigate how seeing a right or left virtual hand affects motor performance when controlling a KINARM robotic device with either the dominant (right) or non-dominant (left) arm. Right-handed participants (n=36) performed centre-out reaches while wearing a VR headset calibrated to match the physical space of the robot, allowing for the independent manipulation of the physical controlling arm (right or left) and the virtual hand seen (right or left). This design created match and mismatch conditions that allowed us to disentangle the effects of the seen hand from the acting hand in motor control. Trajectory analyses revealed that when participants controlled the robot with their dominant right arm, seeing it visually represented as a left hand produced significantly greater curvature error than seeing a right hand. Notably, this effect did not occur for the non-dominant left arm, which showed comparable performance regardless of virtual hand identity. These findings suggest that accurate visual representation of the right/dominant arm may be particularly important for precise motor control, and that visuomotor mismatch selectively disrupts performance when it involves the right/dominant limb. More broadly, this work demonstrates how VR can be leveraged to investigate the contributions of vision to motor control and may have applications for motor rehabilitation, as in mirror-therapy treatments for stroke and phantom-limb pain.

Natural Sciences and Engineering Research Council of Canada Discovery Grant (04271-2022-RGPIN)

23.476 LARGE LANGUAGE MODEL PERFORMANCE IN OPHTHALMIC PATIENT EDUCATION: AMBLYOPIA AND AGE-RELATED MACULAR DEGENERATION

Sunwoo Kwon^{1,2,3}, Artashes Yeritsyan², Dennis Levi^{2,3,4}; ¹Exponent, ²Herbert Wertheim School of Optometry and Vision Science, ³Center for Innovation in Vision and Optics, ⁴Helen Wills Neuroscience Institute

Artificial intelligence (AI) chatbots are increasingly used for patient education, with 22% of Americans reporting seeking health advice from these tools. However, the reliability of chatbot-generated ophthalmic information remains unclear. We conducted a multi-condition evaluation of chatbot responses to patient-focused questions across two prevalent vision and visual-system disorders: amblyopia and age-related macular degeneration (AMD). We assessed the accuracy, comprehensiveness, and readability of these chatbot responses in comparison to the American Academy of Ophthalmology (AAO) and the American Optometric Association (AOA) patient materials. Two condition-specific question sets were constructed from AAO/AOA patient brochures: 21 amblyopia questions and 12 AMD questions. Each question was entered twice into six publicly available AI chatbots (ChatGPT-3.5, ChatGPT-4, Gemini, Meta AI, Snap AI, and Copilot), generating 252 amblyopia and 144 AMD chatbot responses. Using a 5-point Likert scale, amblyopia responses were rated by three optometrists with expertise in amblyopia while AMD responses were rated by five optometrists with expertise in retinal disease. Accuracy and comprehensiveness were analyzed using the Friedman and post

hoc Wilcoxon signed-rank tests, while readability was analyzed using ANOVA and post hoc TukeyHSD tests. Due to differences in question sets and raters, datasets were analyzed independently and synthesized qualitatively. Across both ocular conditions, GPT-3.5, GPT-4, Copilot, and Gemini consistently produced more accurate and comprehensive responses than Meta AI, Snap AI, and AAO/AOA brochures. With the exception of Copilot, all AI chatbots produced significantly harder-to-read texts than AAO/AOA patient brochures for both ocular conditions. Our results demonstrate that across two ophthalmic conditions, AI chatbots, particularly GPT-3.5, GPT-4, Copilot, and Gemini, outperformed AAO/AOA materials in accuracy and comprehensiveness but displayed persistent readability challenges. Collectively, these findings delineate both the emerging capabilities and current limitations of AI systems in ophthalmic patient education.

23.477 THE ROLE OF ANCILLARY BODY MOVEMENTS IN INTERPERSONAL SYNCHRONIZATION DURING JOINT MUSIC MAKING

Sara Fernandez Abalde^{1,2}, Felix Bigand¹, Trinh Nguyen¹, Roberta Bianco¹, Peter E. Keller^{3,4}, Giacomo Novembre¹; ¹Neuroscience of Perception and Action Lab, Italian Institute of Technology, Rome, Italy, ²The Open University Affiliated Research Center, United Kingdom, ³Center for Music in the Brain, Aarhus University, Denmark, ⁴The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia

Joint music making is a statistical human universal, occurring in nearly all documented human societies, suggesting an underlying predisposition for collective musical behavior. However, research in this domain has predominantly focused on trained musicians. To study interpersonal musical coordination independently of musical expertise, we used e-music boxes: digital instruments that allow participants to control the tempo of pre-registered songs through cyclic hand movements. We examined whether ancillary body movements, i.e., movements not required to produce music, facilitate interpersonal synchronization, and whether this effect depends on musical expertise. Forty dyads (20 musician pairs, 20 non-musician pairs) synchronized their musical output, either with or without visual access to each other, but never to each other's instrumental movements. Musical output was encoded as continuous oscillatory signals, enabling precise computation of interpersonal synchrony. In parallel, we recorded full-body kinematics, and electroencephalography from both participants simultaneously. Results showed that seeing the partner significantly improved interpersonal musical synchronization, irrespective of musical training. Using principal component analysis, we separated instrumental from ancillary movements in the kinematic data. While all movements exhibited periodicities aligned with the musical bar, ancillary movements additionally showed a faster periodicity aligned with the musical beat (i.e., the basic temporal pulse of the music). This beat-related periodicity was enhanced when partners were visible, particularly in upper-body anteroposterior sway

and head bobbing. Moreover, visual access increased interpersonal synchronization across several ancillary movements, including upper body sway, head movements and lateral sway, specifically at beat-aligned frequencies, but not at bar-aligned frequencies. A mediation analysis showed that interpersonal synchronization of ancillary movements mediated the effect of vision on musical synchronization. This indicates that visual information extracted from a partner's body movements provides coordination-relevant cues for joint action, independently of musical training. Ongoing electroencephalography analyses aim to identify the neural mechanisms underlying this visual-sensorimotor coupling.

23.478 USING FMRI TO UNDERSTAND THE FUNCTIONS OF THE "QUIET EYE" IN HIGH PRECISION MOTOR TASKS.

Michelle Marneweck¹ (mmar@uoregon.edu), Jackson Zenti¹, Brady Decouto³, Deborah Barany⁴, Nicholas Kreter^{1,2}; ¹University of Oregon, ²University of North Carolina Greensboro, ³Florida State University, ⁴University of Georgia

When performing aiming tasks, experts exhibit a long-duration gaze fixation on a task-relevant landmark just before the final movement, termed the "quiet eye". Behavioral studies have provided evidence for two main hypotheses explaining why the quiet eye aids motor performance. One possibility is that the quiet eye supports the time needed to develop a precise motor plan (programming hypothesis). Alternatively, the quiet eye may support the necessary time to inhibit task-irrelevant distractors (inhibition hypothesis). However, behavioral studies alone have not yet disentangled the extent to which the quiet eye contributes to each of these purported functions. Here, we used Bayesian pattern component modeling of task-based fMRI data during a virtual archery task (N = 20) to test how neural representations of task-relevant demand (i.e., aiming at a near vs. far target) and task-irrelevant distractors (i.e., with or without a peripheral visual distraction) are modulated by the extent of gaze stability (fixed/'quiet' vs. moving/'noisy'). Results showed strong evidence for distinct representations of 1) gaze stability in visual cortex (V1), superior parietal lobe (SPL), dorsal premotor area (PMd), and the primary motor cortex (M1); 2) task-irrelevant visual distractors in V1 only; and 3) task-relevant demand features in V1 and SPL. Consistent with the two hypotheses of quiet eye, the extent of gaze stability modulated representations for both task-relevant demand (V1 and SPL) and task-irrelevant distractors (V1, SPL, PMd, and M1). These results support both quiet-eye hypotheses. However, effects related to the inhibition hypothesis were widespread across visual, parietal, and motor regions, whereas effects related to the motor programming hypothesis were localized to the visual cortex and superior parietal lobe. This divergence of representations across regions indicates that the quiet eye engages multiple, distinct neural processes, rather than a single unified mechanism.

Wu Tsai Human Performance Alliance

23.479 ONLINE VISUAL AND PROPRIOCEPTIVE FEEDBACK IN ACTION SIMULATIONS

Stephen Wood¹ (woods1@myumanitoba.ca), Jonathan J Marotta¹; ¹Perception and Action Lab, University of Manitoba

Motor imagery (MI) refers to the internal simulation or rehearsal of movements without physical execution. According to motor simulation theory, MI is functionally equivalent to motor performance, as it evokes the same cognitive and neuronal processes involved in executing movements. One significant difference between MI and execution is the availability of online sensory feedback for their performance. Movements rely on online visual and proprioceptive feedback to monitor and correct ongoing actions. Alternatively, MI does not produce the actions required to receive online sensory feedback. Due to this difference, sensory and sensorimotor networks represent one area in which MI may not accurately replicate the processes involved in motor performance. The present study evaluated whether online sensory feedback is involved in MI. Fifty participants completed the Hand Laterality Judgement Task (HLJT) across four sensory feedback conditions, including a control, no-vision, disrupted-proprioception, and no-vision + disrupted-proprioception condition. The HLJT is a common measure of MI where participants identify the laterality of a hand stimulus as quickly and accurately as possible. To make their judgements, participants tend to imagine moving their own hand into the orientation of the hand stimulus presented to them, especially when identifying palm-view stimuli. Visual and proprioceptive feedback were manipulated by placing a box over participants' hands and applying transcutaneous electrical pulses to the median and ulnar nerves, respectively. Each participant completed 64 trials of the HLJT in all four sensory feedback conditions for a total of 256 trials. This experiment found that participants took longer to identify the laterality of hands when visual and proprioceptive feedback were disrupted. Moreover, reaction time increases were exclusively observed for palm-view stimuli. Overall, the results of this experiment indicate: 1) Online sensory feedback is used to simulate actions effectively; 2) both visual and proprioceptive feedback contribute to MI performance in the HLJT.

This research is supported by the University of Manitoba Psychology Graduate Fellowship, The University of Manitoba Graduate Fellowship, and an NSERC Discovery Grant.

23.480 REVEALING THE MECHANISMS OF REAL-WORLD ACTION PERCEPTION: BEHAVIORAL AND EEG EVIDENCE FROM LIVE VS. VIDEO ACTIONS

Elif Ahsen Çakmakci¹, Sezan Oral², Burcu A. Ürgen^{1,3}; ¹Bilkent University, Ankara, Turkey, ²University of Luxembourg, Esch-sur-Alzette, Luxembourg, ³Aysel Sabuncu Brain Research Center, Ankara, Turkey

Perceiving others' actions is essential for survival, interaction, and communication. Yet, despite decades of research, we remain far from developing neurally inspired computational models that approach human action perception. A major limitation is that most neuroscience

studies rely on 2D images or videos, which lack the real-time presence and social affordances of actual actions. As a result, the fundamental mechanisms of real-world action perception remain poorly understood. Here, we investigate the behavioral and neural mechanisms underlying the perception of real (live) versus video-based actions. Using a novel experimental setup (Pekçetin et. al., 2023), we conducted a two-session EEG study (N = 26) in which participants observed peripheral actions presented either live or via video while performing a central task under low vs. high attentional load. We analyzed the data using mass univariate ERPs, time-frequency analyses, and time-resolved representational similarity analysis (RSA). Behaviorally, central-task performance was consistently lower for Real than for Video stimuli, with the largest Real–Video difference emerging under high attentional load. ERP analyses revealed reliable differences between live and video conditions during the 150–450 ms window following action onset. Time-frequency analyses over occipital and parietal regions showed that video actions evoked substantially weaker alpha (8–12 Hz) and beta (15–25 Hz) suppression than live actions, indicating reduced perceptual engagement. Complementing these findings, time-resolved RSA robustly distinguished live from video actions within the 250–750 ms window. Taken together, our results demonstrate that live actions engage perceptual systems differently and more powerfully than their video-based counterparts. These findings underscore the limitations of screen-mediated paradigms and highlight the need for more ecologically grounded approaches in social and action perception research.

This study was funded by The Scientific and Technological Research Council of Turkey (TÜBİTAK), Project no: 122K915

23.481 MODEL-BASED SIMULATION VS MODEL-FREE PATTERN MATCHING: WHAT COMPUTATIONS UNDERLIE HUMAN ACTION PLANNING?

Aryan Zoroufi¹, Nishad Gothoskar², Leslie Kaelbling², Joshua Tenenbaum¹, Nancy Kanwisher¹; ¹Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, ²Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology

Vision enables us to understand our surroundings, predict future events, and plan actions. Advances in AI have produced computational models that serve as testable hypotheses for human visual cognition. In action planning, two main strategies have emerged. "Model-based" strategies estimate the world state (3D object shapes, spatial relationships, forces) from 2D inputs, forming beliefs about the environment and planning through mental simulation. While computationally expensive, model-based approaches generalize well to new scenarios. "Model-free" approaches learn direct mappings from sensory inputs to actions through pattern matching, typically as end-to-end neural networks. Model-free strategies are faster at test time but less adaptable to novel scenarios. Which framework better explains human behavior? We asked participants to search for occluded objects in virtual environments, examining which occluder

they moved first. We implemented two algorithm classes: (1) model-based strategies that construct 3D scene representations from RGB images, simulate potential target locations, calculate reveal probabilities, and select the highest-probability occluder; and (2) model-free algorithms trained end-to-end to map directly from RGB images to occluder selection. Model-based algorithms better predict human behavior than model-free approaches. Specifically, a modular perception architecture—decomposing the task into separate components (depth estimation, segmentation, 2D amodal completion, and 3D object reconstruction) wired together—fit human behavior ($r^2 = 0.33$ where human reliability is 0.45) better than end-to-end networks trained directly from RGB to 3D reconstruction (e.g. $r^2 = 0.07$ for GPT5). These findings suggest human action planning relies on explicit 3D mental models and planning through mental simulation, rather than learning direct 2D-to-action mappings. The advantage of modularity in our models mirrors modular human visual cognition, where distinct perceptual processes combine to enable robust scene understanding and action planning.

23.482 PEOPLE'S COST FUNCTIONS IN SENSORIMOTOR DECISION-MAKING REVEAL COMPLEX EFFORT COSTS

Tobias Fabian Niehues¹ (tobias.niehues@tu-darmstadt.de), Dominik Straub^{1,2}, Constantin Rothkopf¹; ¹Technical University of Darmstadt, ²University of Cambridge

Human visually guided decision-making is influenced both by the desirability and the uncertainties of possible outcomes. Bayesian models of perception and action very often represent these quantities as quadratic costs and Gaussian prior beliefs about stimulus distributions. By contrast, several previous studies have suggested different non-quadratic functional forms for, e.g., sensorimotor learning and visual working memory tasks, as well as asymmetrical cost functions for incorporating resource-rational behavior, thereby aiming to explain often observed systematic biases. Incorporating more intricate cost functions makes it challenging to solve the decision-making problem and, consequently, renders it infeasible to solve the inverse problem of determining the parameters that govern subjects' behavior, which include perceptual uncertainty, motor variability, cost functions, and prior beliefs about the stimulus. Here, we employ a recently introduced method that reverse-engineers various families of cost functions directly from observed behavioral data. This reconciles normative and empirical models of sensorimotor behavior. We analyze a variety of previously published experimental datasets and find that none of the observed behaviors is best described by the commonly assumed quadratic costs, but always by asymmetrical cost functions, most of which contain explicit costs representing effort costs for performing the movement. In most tasks, we found that prior beliefs are well-calibrated to the actual stimulus distributions and that subjects usually relied more on sensory information than on prior beliefs, but the degree to which they do and the variability across subjects within a task vary across tasks. Furthermore, we find that identifiability problems between priors and cost functions incorporating effort costs

can be resolved by using behavioral data containing various levels of perceptual uncertainty, and show how our modeling framework can guide the design of experiments to best disambiguate alternative hypotheses about people's implicit cost functions in visually guided behavior.

This research was supported by the European Research Council (ERC; Consolidator Award "ACTOR"-project number ERC-CoG-101045783). We gratefully acknowledge the computing time provided to us on the high-performance computer Lichtenberg at the NHR Centers NHR4CES at TU Darmstadt.

SATURDAY AFTERNOON POSTERS IN BANYAN BREEZEWAY

Spatial Vision: Clinical

**SATURDAY, MAY 16, 2:45 – 6:45 PM, BANYAN
BREEZEWAY**

26.301 ILLUSORY SIZE EFFECTS IN SCHIZOPHRENIA AND BIPOLAR DISORDER

Hannah Moser¹, Helen Yu¹, Killebrew Kyle¹, Scott Sponheim^{2,1}, Michael-Paul Schallmo¹; ¹University of Minnesota, ²Minneapolis VA Health Care System

People with psychotic disorders (e.g., schizophrenia, bipolar disorder) commonly experience abnormal visual perception, such as hallucinations. They also exhibit performance differences in paradigms designed to test visual context processing (e.g., contrast surround suppression). Here, we examined whether there are differences in contextual modulation of perceived size in people with schizophrenia (n=24) or bipolar disorder (n=16) diagnoses, as well as controls (n=39) without a history of psychosis, using an illusory size task (i.e., 'ball in the hallway'). The task consisted of two checkerboard circles presented on the upper and lower parts of a screen, in the absence or presence of a hallway image. With the hallway, the lower circle appears relatively smaller than the upper circle due to contextual modulation by the background, even when the circles are the same angular size. We used an adaptive staircase procedure to measure the difference in perceived size of the circles in the absence versus presence of the hallway. Each participant group showed a significant condition effect, with the presence of the hallway decreasing the perceived size of the lower circle. The strength of this size illusion did not significantly differ between people with schizophrenia, people with bipolar disorder, or controls. A subset of participants also completed a self-report measure of real-world perceptual anomalies, the Sensory Gating Inventory (SGI). The size illusion was negatively correlated with SGI total scores ($r(63)=-0.26$, $p=0.035$), indicating reports of everyday perceptual anomalies were associated with weaker contextual

modulation of perceived size. Our results suggest weaker perception of contextual size illusions may be related to anomalous perception regardless of diagnosis.

I01 CX001843, K01 MH120278, P41 EB027061, UL1 TR002494, T32 DA037183

26.302 VISUAL FIELD ASYMMETRIES IN REDUNDANCY MASKING IN AMBLYOPIA

Anne Danneels¹, Ângela Gomes Tomaz¹, Dennis M Levi¹, Bilge Sayim²; ¹Herbert Wertheim School of Optometry & Vision Science, University of California Berkeley, Berkeley, United States of America, ²Laboratoire de Sciences Cognitives et Psycholinguistique, Département d'Études Cognitives (DEC), École normale supérieure (ENS), CNRS, Paris, France

Amblyopia is a neurodevelopmental disorder characterized by reduced visual acuity. Observers with amblyopia have increased spatial uncertainty and poor spatial localization. In redundancy masking (RM), observers perceive fewer items than presented (e.g., perceiving two lines when three were presented) predominantly with regular arrangements of items. RM is strongest along the horizontal meridian revealing a reverse horizontal-vertical anisotropy compared to crowding. Here, we explored the strength and pattern of redundancy masking in amblyopia. We hypothesized that RM would be reduced in amblyopia, although still strongest along the horizontal meridian, as accurate (relative) localization is required to detect stimulus regularity. Sixteen observers (5 amblyopes, 11 controls) were presented with arrays of 3, 4, or 5 lines at 5° and 10° eccentricity for 150 ms. Eight locations were tested (polar angles: 0° to 315° in steps of 45°). To ensure that the extent of the line array was not used as a cue to the number of lines presented, inter-line spacings were varied: 0.40°, 0.58° and 0.78° at 10° eccentricity (and scaled for presentation at 5° eccentricity). Observers' minimum angle of resolution was smaller than the smallest spacing at each experimental eccentricity. Observers' task was to indicate the number of lines perceived. Preliminary results showed a surprising trend for high accuracy (correct reporting of the number of lines presented) in control observers and overestimation (reporting more lines than presented) in observers with amblyopia. Along the right horizontal meridian, performance was more accurate (less overestimation and no RM) compared to all other locations, except the left horizontal meridian. Similarly, accuracy was higher along the left horizontal meridian but only compared to the left intercardinal locations. These results, together with previous RM studies, highlight possible individual differences in RM strength, and suggest potential group-level differences between individuals with amblyopia and controls.

Supported by a grant from the National Eye Institute awarded to Dennis M. Levi (R21EY030609) and a France-Berkeley Fund grant awarded to Dennis M. Levi and Bilge Sayim, as well as a fellowship of the Belgian American Educational Foundation awarded to Anne Danneels.

26.303 REDUNDANCY MASKING ACROSS ECCENTRICITIES IN INDIVIDUALS WITH AMBLYOPIA AND CONTROLS

Ângela Gomes Tomaz¹, Bilge Sayim², Dennis M. Levi¹; ¹Herbert Wertheim School of Optometry & Vision Science, University of California Berkeley, Berkeley, USA, ²Laboratoire de Sciences Cognitives et Psycholinguistique, Département d'Études Cognitives (DEC), École normale supérieure (ENS), CNRS, Paris, France

Redundancy masking (RM) is the reduction of perceived number of items in repeating patterns (e.g., perceiving two lines when three are presented). RM occurs predominantly with regular arrangements of items, suggesting that regularity is a necessary condition for RM. In amblyopia, a neurodevelopmental disorder of spatial vision characterized by reduced visual acuity, observers have impaired spatial localization and noisy positional perception. As accurate (relative) localization is required to extract stimulus regularity, we hypothesized that RM would be reduced in amblyopia. Here, we explored the extent to which the amblyopic visual system compresses information in RM across the visual field. Seventeen observers (5 with amblyopia, 11 controls) were presented with arrays of 3, 4 or 5 lines for 100 ms (randomly to the right or left visual field). Lines were presented binocularly at 2.5°, 5°, 10°, or 20° eccentricity, as well as monocularly to the non-dominant/amblyopic eye at 2.5° and 5° (eye order counterbalanced). Eye-tracking ensured peripheral presentation of the stimulus. Five inter-line spacings were used from 1.50° to 4.30° in steps of 0.7° (at 20° eccentricity). Line size and inter-line spacings were linearly scaled at the other eccentricities and confirmed to be above observers' minimum angle of resolution at each eccentricity. Observers' task was to report the number of lines perceived. Preliminary results showed stronger RM at 20° eccentricity compared to all other eccentricities, and for the smallest spacing compared to all other spacings (and the second smallest compared to the largest). We found a strong trend for weaker RM across the visual field in the amblyopic compared to the control group. Our findings show that RM increases with eccentricity, and may be reduced in amblyopia hinting at a potential advantage in amblyopic visual processing.

Supported by a grant from the National Eye Institute awarded to Dennis M. Levi (R21EY030609) and a France-Berkeley Fund grant awarded to Dennis M. Levi and Bilge Sayim.

26.304 ATYPICAL VISUAL FIELD ASYMMETRIES IN REDUNDANCY MASKING EXPLAINED BY MAGNITUDE OF POSITIONAL NOISE: EVIDENCE FROM AMBLYOPIA

Bilge Sayim¹, Taylor Lui², Dennis M. Levi², Ângela Gomes Tomaz²; ¹Sciences Cognitives et Psycholinguistique, Département d'Études Cognitives (DEC), École normale supérieure (ENS), CNRS, Paris, France, ²Optometry & Vision Science, University of California Berkeley, Berkeley, USA

In amblyopia, a neurodevelopmental disorder of spatial vision, observers have impaired spatial localization. Redundancy masking

(RM) –the reduction of the perceived number of items in repeating patterns– has been found to strongly depend on stimulus regularity, with strongest RM in regular arrangements and weak or no RM in irregular arrangements, suggesting that regularity is a necessary condition for RM. Here, we explored whether RM exists in amblyopia and, as accurate (relative) localization is required to extract stimulus regularity, we hypothesized that RM would be reduced in amblyopia. Six observers with amblyopia (2 strabismic, 1 anisometropic and 3 mixed) and twelve controls were presented with arrays of 3, 4 or 5 lines for 150 ms randomly to the left or right of fixation at 8° eccentricity. Line arrays were presented binocularly, as well as monocularly to either eye (while the other was patched). Five inter-line spacings were used from 0.45° to 1.38° in steps of 0.23°, all above observers' minimum angle of resolution. The task was to report the number of lines perceived. Preliminary results showed weaker RM for observers with amblyopia compared to controls, and for the non-dominant/amblyopic eye compared to binocular presentation. RM at the largest spacing was weaker compared to all other spacings (except the second largest). The current findings support the notion that the atypical visual field asymmetries of RM—with the strongest RM along the horizontal meridian and no upper-lower asymmetry—may be due to spatial localization accuracy, which is greatest along the horizontal meridian. We suggest that, counterintuitively, higher magnitudes of positional noise (both internal and external) may improve performance, i.e., reduce RM.

Supported by a grant from the National Eye Institute awarded to Dennis M. Levi (R21EY030609) and a France-Berkeley Fund grant awarded to Dennis M. Levi and Bilge Sayim.

26.305 CONTINUOUS PSYCHOPHYSICS IN THE CLINIC: A PILOT STUDY IN PATIENTS WITH RETINAL DISEASE

Veronica Pisu¹, Aleksandrs Koselevs¹, Sarah Cutler², Maria Patsiamanidi², Sam Khandhadia², Isabel Escofet³, Jay Self^{1,2}, Andrew Lotery^{1,2}, Guido Maiello¹; ¹University of Southampton, UK, ²University Hospital Southampton, UK, ³Salisbury District Hospital, UK

Standard clinical vision tests (e.g., acuity charts, perimetry) trade off ease of administration against informativeness: some are quick but narrow, others richer but slow and effortful, especially in challenging clinical populations. Continuous psychophysics offers a promising alternative: a fast and intuitive target-tracking task that yields reliable estimates of visual and motor function (Bonnen et al., 2015; Straub & Rothkopf, 2022). In a pilot study, we evaluated the feasibility, acceptability, and reliability of continuous psychophysics in the clinic, and explored the association of tracking performance with visual acuity (VA). Twelve adults with retinal disease (age range: 54–86 years) attending an intravitreal injection clinic completed two same-day target-tracking sessions and rated the task using the System Usability Scale (SUS). In the tracking task, participants used a mouse to track a randomly moving concentric Gabor target (10 trials/session); testing was binocular and unaided. The task was feasible within the clinic workflow, with short per-session runtime (M 4.03 minutes, SD 0.45;

including self-paced inter-trial breaks; 11/12 completed both sessions). Tracking performance indices showed excellent test-retest reliability ($r_s \geq .84$, $p_s < .05$, $N = 6$). Performance improved markedly between first and second tracking sessions in post-injection assessments ($N = 5$), consistent with short-term recovery, and in one participant who wore corrective lenses in the second session. SUS scores (0–100 scale, higher = better) indicated excellent acceptability ($M = 87.1$, $SD = 11.0$). Surprisingly, associations with VA were non-significant. While this may partly be due to our limited sample, the high test-retest reliability we observed suggests that the method is capturing much more than simple visual acuity. These findings support the feasibility of continuous psychophysics in clinical practice and motivate larger studies, including the development of methods to disentangle visual and motor contributions to target-tracking performance.

This research was funded by Fight for Sight (Grant Reference: RESSGA2302, awarded to GM) and Wessex Medical Research (Grant Reference: AF04, awarded to GM).

26.306 MEG OUTPERFORMS EEG IN CAPTURING FREQUENCY-TAGGED VISUAL RESPONSES

Ayberk Ozkiri¹ (ayberk.ozkiri@gmail.com), Angela Ingrid Renton^{1,2}, Debajyoti Sengupta¹, Jesse Livezey³, David J. Klein³, Yaqing Su³, Adam Hanina³, Dimitri Van De Ville^{1,2}; ¹Neuro-X Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), Geneva, Switzerland, ²Department of Radiology and Medical Informatics, University of Geneva (UNIGE), Geneva, Switzerland, ³Dandelion Science Corp, USA, and Dandelion Science Sarl, Switzerland

Recent work demonstrated that spatial-frequency-tagged natural-scene videos elicit steady-state visual evoked potentials (SSVEPs) which act as robust neural markers of visual function in age-related macular degeneration (Renton et al., 2025). Building on this work, we sought to determine whether electroencephalography (EEG) or magnetoencephalography (MEG) provides a superior signal-to-noise ratio for detecting these frequency-tagged responses, and to capture how artificial scotomas modulate these visual cortical neural responses in healthy controls ($N=20$). To this end, neural activity was recorded using simultaneous MEG and EEG while participants performed a novel gaze-contingent two-alternative forced-choice (2AFC) letter-discrimination task (T vs. L). On each trial, the target appeared at one of 36 visual field locations relative to the participant's real-time gaze position. Participants completed the task with free-gaze under three central-scotoma masking conditions (None, small, large). Throughout the experiment, full-screen videos composed of either natural-scene or artificially-generated texture were displayed in the background, with low- and high-spatial-frequency information distinctly tagged at 5 and 6 Hz (counterbalanced across trials). We found that SSVEP signal-to-noise ratios were substantially higher when detected with MEG compared with the simultaneously acquired EEG, with a medium effect size when averaged across experimental conditions. These results demonstrate a clear advantage for MEG in capturing

frequency-tagged responses in early visual cortex. Moreover, increasing scotoma size was associated with stronger correlations between occipital and occipitoparietal sensors (medium-to-large effect sizes), consistent with prior evidence of cortical reorganization and compensatory recruitment under central vision loss. These findings highlight the efficacy of combining gaze-contingent stimulation, naturalistic video-embedded frequency tagging, and neuroimaging to sensitively probe the neural dynamics of visual function and plasticity. This approach offers a promising framework for individualized mapping of functional reorganization and for developing neurofeedback-based vision-rehabilitation interventions in clinical populations.

This work was funded by the the Swiss Innovation Agency (Innosuisse, Project Enlighten, No. 116.587 IP-LS).

26.307 PSYCHOPHYSICAL MEASUREMENT OF VISUAL ACUITY IMPROVES PRECISION OVER CLINICAL TESTING

Helen Treankler¹, Mina Gaffney³, Mohammed Herzallah², Adam Greenberg⁴; ¹Medical College of Wisconsin, ²Marquette University

During standard visual acuity (VA) testing, participants are asked to read aloud (row-by-row) a series of letters on a chart containing 14 rows (5 letters each) of logarithmically-scaled visual angle. The tumbling E ETDRS chart is comprised of a series of E's in 4 possible orientations via rotations of 0, 90, 180, and 270 degrees. Participants are asked to report the orientation of each E, and VA is typically reported via the line of the smallest E's, in which the majority is correctly identified. We hypothesized that a psychophysical approach to VA measurement would yield a more precise outcome metric and could potentially reduce bias, quantify uncertainty, and differentiate real changes above-and-beyond normal variability. To test this, we developed a digital version of the traditional tumbling E task that uses a staircase paradigm to display a range of E sizes (with orientation randomized), generating psychometric curves from which acuity thresholds can be estimated. Participants performed two experiments at a distance of 2 meters: one with the traditional analog chart and one with our digital task. The digital task contained 280 total trials in two blocks, each containing 3 or 4 interleaved staircases via the QUEST algorithm. Each experiment took approximately 10-15 minutes to complete. Results from both experiments were converted to LogMAR acuity. Our digital method allows for a more continuous measurement of VA rather than assessing at discrete levels. The digital test demonstrated superior precision over analog testing, with adaptive staircases producing tighter clustering and more stable threshold estimates for detecting subtle acuity changes. Our psychophysical approach to measuring acuity demonstrates the potential to improve clinical assessment of VA while preserving patient encounter time, and to potentially identify deficits beyond VA when incorporating reaction time.

Perceptual Training, Learning and Plasticity: Rehabilitation

SATURDAY, MAY 16, 2:45 – 6:45 PM, BANYAN BREEZEWAY

26.308 HOME BASED VISUAL TRAINING FOR AMBLYOPIA USING CONSUMER VR HARDWARE

Lucy Starling¹, Juju Fars¹, Po-Chun Tseng¹, Junyun Zhang², Dennis M. Levi³, Laura Breach⁴, Vijay K. Taylor-Hamblin^{5,6}, Pedro Quijada Leyton⁷, William Latham⁸, Stephen Taylor¹, I. Betina Ip¹, Holly Bridge¹; ¹University of Oxford, Oxford, UK, ²Peking University, Beijing, China, ³UC Berkeley, USA, ⁴Oxford Eye Hospital, John Radcliffe Hospital, Oxford, UK, ⁵NIHR Moorfields Biomedical Research Centre, London, UK, ⁶Moorfields Eye Hospital NHS Foundation Trust, London, UK, ⁷University College London, UK, ⁸Goldsmiths, University of London, UK

Virtual reality (VR) training offers a promising platform for vision science research conducted outside of the laboratory and for visual rehabilitation. Here we aimed to determine whether an established computer-based training for amblyopia could be implemented into VR. We tested whether a home-based VR training protocol, based on the dichoptic de-masking task described in Liu and Zhang (2018), could induce perceptual learning in participants with amblyopia (N=18, 12 female, aged 18-50, 15 training days). Normally sighted controls (N=16, 13 female, aged 21-30, 5 training days) provided normative performance measures. The training task was to indicate which of two sinusoidal gratings, presented to the weak eye, had a higher contrast. The stronger eye was presented with visual noise. The delta (contrast difference between the two gratings) was used to quantify changes during training. In addition, both groups performed tests of visual acuity, contrast sensitivity, interocular balance and stereopsis before and after training. VR training improved performance on the dichoptic de-masking task for participants with amblyopia (first day mean delta = 20.42 (SD=4.09), final day mean delta = 12.55 (SD=5.79), $t(17)=5.89$, $p<.001$) and controls (first day mean delta = 18.13 (SD=2.02); final day mean delta = 13.49 (SD=3.62); $t(14)=4.19$, $p<.001$). Participants with amblyopia significantly improved their interocular balance ($t(17)=2.386$, $p=.011$), and five gained stereopsis. There was no difference in visual acuity or contrast sensitivity in the amblyopic participants, and those with healthy vision showed no change in any outcome measures. While all except one participant completed the required training amount, feedback suggested training was repetitive, tedious, and lacked variety. These findings suggest that VR training outside of the lab is feasible and can be completed with high compliance. Training resulted in plasticity in controls and visual rehabilitation in amblyopes, opening avenues for larger scale studies in a variety of visual conditions.

This work was funded by the Medical Research Council (MR/V034723/1 to HB and IBI), and The Royal Society (University Research Fellowship to HB, Dorothy Hodgkin Research Fellowship to IBI). LS is funded by a Linacre EPA Cephalosporin scholarship.

26.309 BOOSTING BLIND-FIELD PLASTICITY WITH VIRTUAL REALITY (VR)-BASED PERCEPTUAL TRAINING AND OCCIPITAL TRNS

Michele Bevilacqua¹, Rebecca Frinco², Meghna Uzgare¹, Dean Thurston³, Peige Wang³, Aaron R. Seitz³, Lorella Battelli¹; ¹Berenson-Allen Center for Noninvasive Brain Stimulation, Department of Neurology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA, ²Department of Psychology, University of Turin, Torino, Italy, ³Department of Psychology, Northeastern University, Boston, MA

Stroke-related damage to primary visual cortex (V1) is one of the main causes of chronic homonymous hemianopia ("cortical blindness", CB). Standard rehabilitation mainly relies on compensatory strategies, yet gaze-contingent visual perceptual learning can partially restore blindfield motion processing and expand perimetric sensitivity when training targets the blindfield border. Although effective, this approach often requires prolonged training, limiting compliance. More recently, pairing high-frequency transcranial random-noise stimulation (tRNS) with visual training has been shown to accelerate recovery in CB, likely by increasing cortical excitability and enhancing detection through noise-based (stochastic resonance) mechanisms. To extend such interventions to patients without clinic access, we piloted VR-based perimetry and motion-discrimination training paired with tRNS. Eight patients completed immersive, gaze contingent VR training with integrated eye tracking, using the new open science XRPSys platform, to ensure motion stimuli remained anchored to each patient's blindfield border. Training used a coarse direction discrimination and integration task that adaptively varied motion coherence (normalized direction range, NDR) to estimate thresholds. Occipital tRNS (110-640 Hz) was delivered during training blocks. Thresholds were tracked across 10 days (Day1-Day10), with performance in each patient's intact visual field as an internal control. Visual-field extent was assessed pre- and post-training using VR-based perimetry, alongside Humphrey perimetry and contrast sensitivity (qCSF). Group-level blindfield performance improved by mid-training. Paired Wilcoxon tests showed significant baseline-to-day6 improvements in NDR ($p=0.031$), while the intact hemifield showed no significant change ($p\geq 0.148$). VR perimetry showed a mean seeingfield expansion of +5.73% (range -9.5% to +13.3%; 6/8 improved, 1 unchanged, 1 decreased). Precisely targeted VR-based perceptual learning paired with occipital tRNS produced concurrent gains in motion processing and measurable increases in visual-field extent, consistent with retained plastic potential at the blindfield border. This VR+tRNS approach is feasible in chronic hemianopia and yields promising functional and perimetric gains, supporting its potential for home-based rehabilitation.

This work was supported by the NC NM4R 2024-2025 Pilot Project Grant (NIH Grant UM1TR004406 for NC TraCS) and by NEI-R61EY037510-01.

26.310 ENHANCED VISUO-ATTENTIONAL CONNECTIVITY UNDERLIES TRAINING-INDUCED RESIDUAL VISUAL IMPROVEMENT IN OCCIPITAL STROKE SURVIVORS

Lilie Gailloud¹, Rebecca Millington-Truby¹, Hanna Willis^{1,2}, Lucy Starling¹, Matthew Cavanaugh³, Krystel Huxlin³, Marco Tamietto^{4,5}, Holly Bridge¹; ¹University of Oxford, ²École normale supérieure, ³University of Rochester, ⁴Tilburg University, ⁵University of Torino

Damage to the primary visual cortex (V1) following stroke typically causes contralesional visual field loss, yet many individuals retain residual vision (also known as blindsight), likely supported by extrastriate areas such as motion-sensitive hMT+. Here, we tested the hypothesis that extensive blind-field training restores vision in the blind field by modulating functional connectivity across V1, hMT+, and frontoparietal attentional regions. Seventeen participants with visual field deficits (5 female; 24-71 years; >6 months post-stroke) completed ≥6 months of blind-field motion discrimination training (2x20min/day; 102-271 sessions) and underwent pre- and post-training MRI. Sixteen controls (6 female; 28-71 years) were scanned once. Psychophysical measures included motion discrimination thresholds and detection of moving Gabors. Resting-state connectivity analyses identified regions showing BOLD correlations with ipsi- and contralesional V1, hMT+, intraparietal sulcus (IPS), and dorsolateral prefrontal cortex (DLPFC). Patient-control differences were assessed with two-sample t-tests; pre-post training effects with paired t-tests including performance as covariates ($Z > 2.3$; cluster-corrected $p < 0.05$). Relative to controls, stroke survivors showed widespread V1 hyperconnectivity, lower contralesional hMT+–V1 and IPS–V1 coupling, and greater ipsilesional hMT+–somatosensory and IPS–somatosensory connectivity, consistent with disrupted visual input and compensatory multisensory recruitment. Training-related motion discrimination improvements correlated with increased ipsi- and contralesional hMT+–paracingulate, ipsilesional DLPFC–V1 and hMT+, and contralesional DLPFC–V4 and hMT+ connectivity, suggesting top-down facilitation of residual motion processing. Gabor detection gains correlated with strengthened dorsal attention network connectivity, with increased ipsi- and contralesional IPS–IPL and precuneus coupling. Ipsi- and contralesional DLPFC–extrastriate (V2, V3, V4, and hMT+) connectivity also increased, underscoring the central role of frontoparietal regions. To conclude, visual improvements in cortically-blinded fields appear to depend critically on strengthening visuo-attentional connectivity to amplify residual signals. Rehabilitation should therefore not only re-engage spared visual circuits but actively boost attentional engagement, with multisensory and virtual reality-based approaches offering promising routes to more efficient recovery.

26.311 TRAINING TO USE PERIPHERAL VISION AFTER SIMULATED CENTRAL VISION LOSS: BEHAVIORAL OUTCOMES

Elam Cutts¹, Madeline Ragland¹, Pinar Demirayak¹, Marcello Maniglia², Jaap Munneke³, Aaron Seitz³, Nicholas Turk-Brown⁴, Kristina Visscher¹; ¹University of Alabama at Birmingham, ²University of California Riverside, ³Northeastern University, ⁴Yale University

Patients who lose vision in the center of the retina must adapt to this vision loss by learning to use peripheral vision for daily tasks. Many develop an area within the retina which they use preferentially for daily tasks, called the Preferred Retinal Locus (PRL). Learning to use this peripheral retina for daily vision involves improving low level factors like visual sensitivity, mid-level factors like spatial integration, and high-level factors like attention and eye movements. We trained healthy vision participants to use a peripheral location, assigning participants to be trained on low-level, mid-level, high-level factors, or a combination training, using a gaze contingent display with a simulated scotoma. Participants were tested before and after training on tasks that assess these factors of vision in their peripheral vision, as well as more ecological tasks, including reading. In general, we found that training on a given factor improved performance on that factor and less-so to other factors (near transfer). Further, improvements on low-level and mid-level tasks regardless of training type, predicted improvement in more ecological tasks. These results are consistent with the idea that training on low-level features in the artificial context of a training game may lead to improvements in real-world activities of daily living. I will discuss this work in the context of developing more efficient and effective training paradigms.

26.312 LOCALIZED MOTION DISCRIMINATION TRAINING IN PATIENTS WITH EARLY VISUAL CORTEX DAMAGE IMPROVES VISUAL FUNCTIONING GLOBALLY

Niki Lam¹ (niki_lam@urmc.rochester.edu), Berkeley K. Fahrenthold¹, Matthew R. Cavanaugh¹, Elizabeth Ginzburg¹, Neil Dogra¹, Maria Grothe¹, Bryan V. Redmond¹, Jingyi Yang¹, Duje Tadin¹, Marisa Carrasco², Krystel R. Huxlin¹; ¹University of Rochester, Rochester, NY, ²New York University, New York, NY

Occipital stroke causes postchiasmatic visual cortex damage and cortical blindness (CB), decreasing visual functioning and quality of life. Here, we asked whether motion discrimination training with feature-based attention (FBA) at a single, blind-field (BF) location for at least 6 months is sufficient to (1) increase performance at an untrained BF location, (2) improve functional field scores computed from Goldmann perimetry, and (3) improve quality of life as assessed with the National Eye Institute Visual Function Questionnaire (NEI-VFQ). Data were obtained from 39 subacute CB patients enrolled in clinical trial NCT04798924 (mean age/SD: 53.5/13.5 years; time post-stroke: 3.4/1.1 months; male/female: 25/14). Before and after training, patients completed eye-tracker-enforced psychophysical testing, perimetry, and the NEI-VFQ. Goldmann area for different isopters and

functional field score (which heavily weighs the lower hemifield and central 10°) were computed to estimate the extent of intact vision in each patient. The NEI-VFQs were manually scored and averaged to create a single composite score per person per test. At-home training involved performing 600 trials of a motion discrimination task with FBA and caused significant improvements at both trained and untrained BF locations. Additionally, training improved the functional field score, NEI-VFQ composite scores, and area of central isopter I2e, but not the total area of vision encompassed by the V4e isopter. Changes in motion-discrimination thresholds at trained BF locations did not predict changes in functional field score or NEI-VFQ composite score, but the number of sessions trained did, especially in the subacute period. In sum, localized BF motion discrimination training with FBA improved both local and global visual performance in CB patients, increasing the functional field of view and quality of life. Moreover, the amount of training during the subacute period predicted global visual function outcomes, suggesting that earlier and more frequent training is more effective.

26.313 TRAINING WITH FEATURE-BASED ATTENTION AFTER EARLY VISUAL CORTEX DAMAGE REDUCES LIMITS ON PERCEPTUAL RECOVERY DUE TO INTERNAL ADDITIVE NOISE

Berkeley K Fahrenthold¹, Matthew R Cavanaugh¹, Michael D Melnick¹, Duje Tadin¹, Marisa Carrasco², Krystal R Huxlin¹; ¹University of Rochester, ²New York University

Cortical blindness (CB) arising from damage to early visual cortex can be partially alleviated by visual retraining, although perceptual thresholds at trained blind-field (BF) locations typically remain impaired relative to corresponding intact-field (IF) locations. Previous results suggest that high internal additive noise underlies residual BF impairments. In normally-sighted individuals, feature-based attention (FBA) increases gain and sharpens neuronal population tuning, helping overcome internal noise and reducing the impact of external noise. Here, we asked whether training with FBA engages similar mechanisms in the BF and IF of CB patients. Twenty-nine patients (54±12 years old, 19 males) enrolled 1.2–5.1 months post-occipital stroke (clinical trial NCT04798924) trained on a direction discrimination task with FBA across two phases. In Phase1, they were randomized to train in either their BF or IF until 6 months post-stroke; in Phase2, all participants trained only in their BF for the next 6 months. Direction discrimination thresholds improved significantly in the BF after both training phases (Phase1 improvement by 36.14±34.19°; Phase2 improvement by 32.9±40.33°) but only after Phase1 in the IF (Phase1 improvement by 0.77±1.18°; Phase2 worsening by 0.13±0.81°). Threshold versus noise (TvN) functions were measured pre- and post-training and fit using the perceptual template model (PTM) framework. This revealed key differences between the BF and IF: in Phase1, TvN curves in the IF showed improvements consistent with a combination of enhanced external-noise exclusion and reduced internal additive noise, with no further changes in Phase2. In the BF,

TvN curves exhibited their largest improvements in Phase2, driven primarily by reductions in internal additive noise. Our results suggest that training with FBA elicits perceptual learning in the BF via different processing changes than in the IF. Whereas training in the IF changes both gain and tuning, training in the BF primarily reduces internal additive noise and increases population-response gain.

Visual Search: Search strategies, clinical

SATURDAY, MAY 16, 2:45 – 6:45 PM, BANYAN BREEZEWAY

26.314 VISUAL SEARCH DEMANDS MODULATE EFFORT INVESTMENT IN OPTIMAL STRATEGIES

Tianyu Zhang¹, Vladislav Khvostov¹, Andrew B. Leber¹; ¹The Ohio State University

In daily life, people frequently engage in visual search, such as finding a friend in a crowded mall or locating a car in a parking lot. Individuals vary in how strategically they guide their attention to find the target, and their strategies are often surprisingly suboptimal. Why would people prefer suboptimal strategies? One proposal is that they avoid the cognitive effort required to implement the optimal strategy, because they perceive that the required effort investment is not worth the performance gains realized. We used the Adaptive Choice Visual Search (ACVS) paradigm to measure the optimality of individuals' visual search behaviors (Irons & Leber, 2018). In this paradigm, the optimal strategy is to search through the less numerous of the two color subsets; search optimality is quantified as the percentage of trials on which participants report the target in the less numerous color subset. In order to test the aforementioned hypothesis, we manipulated the ACVS search difficulty across a series of experiments by varying target-nontarget similarity and set size. We designed the experiments in such a way that greater task difficulty corresponded with greater performance benefits for using the optimal strategy. Results from all experiments revealed a significant effect of search difficulty on strategy optimization, with optimality increasing with greater task difficulty. In addition, when we informed participants of the optimal search strategy, the effect of this instruction was greater in the hard search condition than in the easy condition. Taken together, these findings suggest that increased search demands encourage people to invest the cognitive effort needed to achieve more optimal search behaviors.

26.315 FIXATED BUT FAILED TO SEE: RECOGNITION ERRORS IN VISUAL SEARCH

Jonathan Nir¹ (jonathan.nir@mail.huji.ac.il), Carmel Ruth Auerbach-Asch¹, Leon Yona Deouell^{1,2}; ¹Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, ²Department of Psychology, The Hebrew University of Jerusalem

People routinely perform visual search – from everyday actions like locating our keys to high-stakes tasks such as airport security screening – yet they sometimes fail to notice and respond to clearly visible items. Authors miss typos, drivers fail to stop for pedestrians, and radiologists overlook tumors. These “looked but failed to see” (LBFTS) errors (Wolfe et al., 2022) offer a useful window into the mechanisms linking perception, awareness, and action. In a subset of LBFTS errors, corresponding with Kundel’s (1978) recognition- or decision-errors, observers fixate on a target but still fail to report it. The probability of this phenomenon varies remarkably across studies, from 2% to 34%, depending on target salience, prevalence and specificity. By incorporating eye-tracking within a visual search task that includes easily recognizable objects and allows observers to freely refixate on target-examples, we were able to minimize memory- and decision-based errors, thus focusing on errors where subjects fixate a target but fail to recognize it. A hierarchical Bayesian logistic regression modelled the probability that a fixated target was not recognized based on array type (color items, grayscale items, color items with noise background), target category (human, animal, inanimate), and rotation angle ($\pm 20^\circ$). Across observers, 26.2% of target visits were classified as recognition errors. Posterior predictive estimates showed similar error rates for color items with and without noise-background (23.9% and 25.0%), but substantially higher rates for grayscale arrays (31.8%). Target category also modulated recognition: animate targets increased error probability by 20.5 percentage points (95% HDI = [12.5%, 27.5%]). Rotation angle did not meaningfully affect error probability. Our systematic characterization of recognition-based LBFTS errors in visual search, reveals that both low-level visual features and high-level semantic categories shape whether observers fail to detect a target they have fixated.

26.316 VISUAL FORAGING PROMOTES PERCEPTUAL CATEGORY LEARNING OF TARGETS AND DISTRACTORS

Victoria L. Jacoby^{1,2} (vjacoby@bwh.harvard.edu), Jeremy M. Wolfe^{1,2}; ¹Brigham and Women’s Hospital, ²Harvard Medical School

Visual search often relies on category-based target detection, particularly in foraging tasks that require locating multiple targets among distractors, such as picking berries or detecting tumors. If categories are unfamiliar or perceptually similar, observers must learn how to distinguish targets from distractors. While incidentally learned information about targets and distractors can aid search performance (e.g., Addelman & Störmer, 2023), prior work has primarily examined statistical learning of single features and locations or memory for individual items. Here, we asked whether observers could move beyond individual item learning to discover and extract perceptual patterns among items when targets and distractors are drawn from similar, perceptually based categories. In Experiment 1, participants were asked to collect rocks from one geological rock category while avoiding those from another. Participants were shown three examples from each category. Next, a pretest evaluated their ability to categorize rocks based on this minimal training. This was followed by visual

foraging with feedback and post-foraging recognition and classification assessments. Assessments included previously seen and novel items from the two primary rock categories, as well as items from additional, unseen categories. Results revealed above-chance classification performance, with significant improvements after foraging. Exemplar recognition was strongest for target items, and classification accuracy was highest for previously seen targets, suggesting greater exemplar-level learning for targets. Categorization of novel exemplars was equivalent for targets and distractors, indicating similar category-level learning. In Experiment 2, we removed the pretest and all explicit cues indicating that distractors belonged to the same category prior to the assessments and found similar results: generalizable category learning was observed for both categories. These results suggest that visual foraging supports the learning of perceptual patterns in both targets and distractors, enabling generalization to novel items and influencing how observers search for these categories.

This work was supported by the Harvard Medical School Alber’s Fund

26.317 UNDERSTANDING THE COSTS OF TARGET FRAGMENTATION IN MEDICAL IMAGING DIAGNOSTICS

Cailey Collins¹ (cpcollins@mun.ca), Blaire Dube²; ¹Memorial University of Newfoundland and Labrador

Detection of abnormalities in medical imaging is essential for diagnosis, yet a substantial proportion of radiological errors arise because observers fail to inspect target-containing regions. Recently, segmented rapid serial visual presentation (RSVP; dividing a full image into centrally presented segments) has been proposed as a method to improve diagnostic search. By forcing observers to view every region in a controlled sequence, segmented RSVP aims to ensure full image coverage and reduce premature search termination. Although prior work demonstrates significant improvements in target sensitivity, the methods used ensured that targets always appeared intact within a single segment. In realistic image segmentation, diagnostically relevant features may instead be split across segments, potentially impairing target detection. The present study addresses this gap across two experiments. Experiment 1 tested whether fragmentation reduces target sensitivity. Participants performed an RSVP search for either a square among diamonds or a diamond among squares (intermixed). Each trial comprised a 24-frame search stream, with each frame containing four stimuli that were spatially jittered so that stimuli could appear partially outside the RSVP frame (edge-occluded) but never overlapped. 50% of trials contained a target; on full-target trials (25%), the target appeared within a single frame, on partial-target trials (25%), only half of the target was visible, with the other half occluded by a frame edge. Results showed a clear detection cost for partial targets. Experiment 2 tested whether repeated exposure to fragmented features—an inherent property of segmented RSVP—would offset this cost. Participants viewed streams in which fragmented targets appeared across two different frames. Although repetition improved overall sensitivity, a fragmentation cost remained, indicating that repeated exposure only

partially compensates for missing diagnostic information. Thus, target fragmentation imposes meaningful perceptual costs in segmented RSVP search, highlighting constraints on RSVP-based diagnostic aids that should be considered in their design and implementation.

Natural Sciences and Engineering Research Council of Canada Discovery Grant

26.318 PITFALLS AND TRIUMPHS OF VISUAL SEARCH FOR CANCEROUS MOLES

Zephyr Markley¹, Gabriel Conn¹, Katherine S. Moore¹; ¹Arcadia University

Decades of research on visual search in basic and applied settings have informed theoretical models of how we carry out such tasks in a variety of settings and which external and internal factors affect search accuracy and speed. Among these factors include target visibility, target-distractor similarity, and subsequent miss errors (e.g., early quitting upon finding one of several targets in a display). While these and other factors have been identified in laboratory and applied settings, such as TSA safety searches and radiology scans, little attention has been given to dermatological searches. In the present study, novice searchers (university students) searched for cancerous melanoma and carcinoma moles among benign moles, simulating a skin search. First, participants were trained to identify the three different mole types. Once they reached criterion, they performed a speeded skin search for the cancerous mole targets. We varied (1) target visibility with skin tone, (2) target-distractor similarity (melanoma is more visually similar to benign moles than is carcinoma), (3) the number of targets on a trial, to measure subsequent misses, and (4) time pressure, including the presence of a ticking clock on ½ of all trials, to simulate the time pressure that dermatologists often find themselves under during an office visit. Consistent with visual search literature, skin tone (target visibility) and mole type (target-distractor similarity) affected performance, but there was no significant subsequent miss error effect. Participants were faster and less accurate in the presence of a clock than when no clock was present, suggesting the time pressure included a speed-accuracy trade-off. Our findings may inform how novices can be trained to conduct self-screens. Reliable and valid screenings can not only save lives but also save on expensive and invasive procedures by avoiding false positives.

26.319 CONTEXTUAL CUEING IN MULTI-TARGET VISUAL SEARCH: EFFECTS ON SEARCH TERMINATION AND TARGET SELECTION ORDER

Jeunghwan Choi¹, Sang Chul Chong^{1,2}; ¹Graduate Program in Cognitive Science, Yonsei University, ²Department of Psychology, Yonsei University

People can find a target more quickly in a familiar context than in a novel one, a phenomenon known as contextual cueing (Chun & Jiang, 1998). In most contextual cueing experiments, participants are asked

to search for a 'T'-shaped target among 'L'-shaped distractors and report its orientation. The trial is terminated immediately after the response. However, real-world visual search often involves detecting more than one target. In such situations, not only is the rapid detection of the first target important, but also it is critical to determine whether another target is absent, or to locate it efficiently if it is present. To investigate these situations, we conducted a contextual cueing experiment in which participants searched for a 'T'-shaped target among 'L'-shaped distractors. Each display contained either one or two targets. Participants were asked to click on all targets present in the display and press the space bar to terminate the search when they believed no additional target remained. Both one-target and two-target conditions included repeated and novel displays, allowing us to measure contextual cueing. Our results showed that search termination was significantly faster in repeated displays than in novel ones for both the one-target and two-target conditions. This reduction in termination time was primarily driven by faster detection of the first target. In one-target displays, participants continued to search longer after detecting the target in repeated displays than in novel displays. In two-target displays, contextual cueing for the interval between the first and second target detection was weak; however, participants selected the two targets in a more consistent order in repeated displays than in novel ones. These findings demonstrate that contextual cueing facilitates visual search in repeated displays by allowing people to rely primarily on information from the first detected target, even when multiple targets may be present.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT)(RS-2022-NR070542).

26.320 VISIBILITY IMPACTS RELIANCE ON ANCHOR-OBJECTS DURING TARGET SEARCH

Makayla Souza-Wiggins¹, Joy J. Geng¹; ¹University of California, Davis, ²Center for Mind and Brain

Real-world scenes follow predictable spatial patterns, with some objects more likely to co-occur with other objects. Anchor objects (i.e., large, stable objects that predict smaller objects) are strong predictors of where smaller objects are likely to appear in a scene (e.g., a sponge in the kitchen sink). We hypothesized that under degraded visual conditions, observers will rely more heavily on predictable anchor objects to find targets. To test this, we developed 40 scenes in Unity and asked participants to search for a novel exemplar of an object within each scene under high- or low-visibility conditions. Within a scene, the target (sponge) could appear in a congruent location (by the sink), an incongruent location (on a chair), or be absent. Participants responded whether the target was present or absent (N=51). There were significant interactions between visibility and congruency in both RT and accuracy due to larger congruency effects in the low-visibility condition (RT: $\beta=55$, $p<0.001$; accuracy: $\beta=0.58$, $p<0.001$). This suggests that when low visibility hinders target detection, we rely more on anchor-based spatial predictions as proxies

to guide attention. This dependence produces a larger spatial congruency effect as expectations based on prior knowledge compensate for perceptual limitations. Dwell time on the anchors support this finding: under high-visibility conditions, only highly predictive congruent anchors (determined by a separate judgment task) were fixated longer. In contrast, all congruent anchors were fixated for longer under low-visibility conditions. This suggests that even less predictive anchors were relied upon under low-visibility to help disambiguate targets ($\beta=0.14$, $p=0.015$). Together, these results show that degraded visibility increases reliance on anchor objects for guiding attention, whereas under high-visibility conditions, anchors are used more selectively, specifically when they are highly predictive of the target.

NEI T32 Vision Training Grant

26.321 WHEN DOES WHAT MATTER TO WHERE? IDENTITY-LOCATION INTEGRATION IN SPATIAL CONTEXT LEARNING

Mikayla N. Krech¹ (krech036@umn.edu), Roger W. Remington¹, Vanessa G. Lee¹; ¹University of Minnesota

Visual search is faster on displays that occasionally repeat (contextual cueing, CC). Traditionally, CC is attributed to learning the spatial context of a search array independent of item identities. However, research using real-world objects showed that CC emerged only when both item identities and locations were repeated (Makovski, 2016). Building on these findings, we test whether identity repetition is necessary when (1) the stimuli do not perturb the perceived global configuration via changes in elongation or axis orientation, and (2) the task does not require identity-based search. Participants searched for a target among distractor items over 24 trial blocks. Across blocks, displays were either fully repeated (All-Repeat), repeated in locations but not identities (Location-Repeat), or entirely new (New-Display). Experiments 1 and 2 tested Chinese readers using Chinese characters—square-like stimuli with relatively homogeneous elongation and axis orientation. Participants performed either an identity search (finding an animal character; Experiment 1) or a non-identity search (finding a faint dot superimposed on a character; Experiment 2). Experiment 3 tested English speakers using visually variable common objects in the identity-irrelevant dot-search task. We examined whether CC depended on identity repetition when identities were neither task-relevant nor visually distinctive. In Experiment 1, CC emerged only in the All-Repeat condition; Location-Repeat yielded no CC, suggesting that in identity search tasks, identity and location were jointly learned even with visually homogeneous stimuli. In the dot-search task of Experiment 2, Location-Repeat alone was sufficient to produce CC, indicating task modulation. In Experiment 3, however, Location-Repeat failed to produce CC when stimuli were visually distinctive, even though identity was task-irrelevant. These findings demonstrate that identity information is incorporated into spatial contextual cueing except when it is both visually nondistinctive and irrelevant to the search task.

Supported in part by the McKnight Foundation and a University of Minnesota Graduate Student Research Award.

26.322 EFFORTFUL RESPONSES CHANGE QUITTING STRATEGIES IN VISUAL SEARCH

Austin Durgin¹ (adurgin@conncoll.edu), Maddie McDaniel¹, Jeff Moher¹; ¹Connecticut College

Even in simple visual search tasks, participants miss easily seen targets at a surprisingly high rate. In real-world situations, such as medical image reading or x-ray baggage screening, these errors can carry high costs. One factor that may influence miss rates is when and how participants decide to quit their search. In other words, if I have been searching for a target for a while and don't see one, when is it time to quit and move on? One factor that we believe may be leveraged to influence quitting strategies in search is effort. People avoid effort where possible, and anticipated effort has been shown to shift decision strategies in other cognitive domains. To examine this issue, we presented participants with a simple visual search task in which they determined whether a vertical target line was present among tilted lines. We varied the amount of effort required to indicate target absence across two conditions. In the "Absent Effort" condition, participants had to hit three different keys to indicate target absence and only one key to indicate target presence. In the "Present Effort" condition, these mappings were reversed such that three keypresses were required to indicate target presence. Participants completed both conditions in randomly counterbalanced order. In the Absent Effort condition, participants committed fewer miss errors, while also taking longer to respond on target absent trials. Together, these results suggest that participants were delaying quitting the search process when a target absent response required the additional effort of three keypresses. This result highlights important links between action and cognition, and also suggests a relatively simple but effective way in which search strategies might be manipulated in real-world search scenarios. Ongoing research in our lab is continuing to explore the role that anticipated effort can play in quitting during visual search.

This work was supported by National Science Foundation grant BCS-2218384 to JM

26.323 TESTING THE ADAPTATION TO CHANGING SPATIAL REGULARITIES: DISTRACTOR LOCATION SUPPRESSION PERSISTS, TARGET LOCATION PRIORITIZATION FAILS.

Mustafa Zeyd Söyük¹, Metin Üngör², Anna Schubö¹; ¹Cognitive Neuroscience of Perception and Action, Department of Psychology, Philipps-University Marburg, Marburg, Germany., ²Associative Learning, Department of Psychology, Philipps-University Marburg, Marburg, Germany.

Visual search is influenced not only by stimulus salience and goal-driven factors but also by prior experience with regularities in the environment. Salient distractors appearing more frequently at certain

locations can be suppressed over time, reflecting learned adjustments in the spatial priority map. We tested two key questions: (1) whether the degree of spatial suppression adapts to gradually changing spatial regularities of the distractor, and (2) how prior suppression of a high-probability distractor location affects its later prioritization when it becomes a high-probability target location. Participants performed an additional singleton task in which one location contained a color distractor more frequently than other locations (high- vs. low-probability distractor locations). During acquisition, distractor location ratios gradually increased (2:1 → 4:1 → 8:1; ascending group) or decreased (8:1 → 4:1 → 2:1; descending group). We hypothesized that distractor location learning (DLL) would emerge in both groups, and impact differently on subsequent target location learning (TLL) of the same location now being relevant for the target. Both groups exhibited DLL, but the descending group showed persistent suppression even under weaker ratios. When the previously high-probability distractor location became the high-probability target location, neither group showed TLL, suggesting that prior suppression constrained subsequent upweighting of the same location. These results indicate that the priority map exhibits asymmetric plasticity: down-weighting a location is rapid and persistent, whereas subsequent prioritization is limited.

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation – project number 290878970-GRK 2271, project 9).

26.324 EXAMINING AGE-RELATED DIFFERENCES IN VISUAL SEARCH BEHAVIOR USING CURSOR TRAJECTORIES

Shira Tkacz-Domb¹, Kelvin S. Oie², Dwight J. Kravitz^{1,3}, Stephen R. Mitroff¹; ¹The George Washington University, ²DEVCOM Army Research Laboratory, ³National Science Foundation

Visual search, looking for targets among distractors, is a critical cognitive skill that underlies many day-to-day tasks (e.g., finding your keys) as well as several vital professions (e.g., radiology, aviation security). Unfortunately, search is vulnerable to age-related performance decline, and it is important to understand why. To gain deeper insight into search behavior and how it changes with age, the current study continuously recorded participants' cursor trajectories (i.e., how they moved the mouse cursor around the display) during a search task. In an online experiment, participants searched for target Ts among pseudo-L distractors in visual arrays that contained 1, 12, 18, or 24 total items (50% target prevalence). Stimuli were rotated (0, 90, 180, 270°), and target-distractor similarity was manipulated by varying the symmetry of the L distractors through 7 levels (increasing the offset level moved the vertical segment of the L closer to the center of the horizontal segment, making it more similar to a T). The cursor trajectory data were used to calculate dwell time for the target and distractors, and they were then modeled to explore how behavior changed as a function of target-distractor similarity and age. Broadly, dwell times were longer for distractors that were most configurally

similar to the target T shape, indicating increased interference with higher similarity. Age-related changes in search behavior emerged across a number of analyses, revealing nuanced interactions between age and search behavior (e.g., increased dwell time under higher search difficulty for older participants). The results highlight the value of tracking cursor trajectories to reveal age-related changes in search processes.

This research was funded by US Army Research Laboratory Cooperative Agreements #W911NF-24-2-0188, #W911NF-23-2-0210, & W911NF-25-2-0104

26.325 THE EFFECT OF TARGET CONTRAST ON VIRTUAL REALITY SEARCH TASK IN OBSERVERS WITH CENTRAL FIELD LOSS

Andrew Freedman¹, Preeti Verghese¹; ¹Smith-Kettlewell Eye Research Institute

Individuals with macular degeneration (MD) must rely on peripheral vision, where contrast sensitivity is lower. Greater target contrast improves search performance in an on-screen search task, but it is unclear if this holds for naturalistic search while walking in a virtual reality (VR) environment. To investigate this, we measured search performance in a VR space for participants with MD (n=2) and controls with a 10° gaze-contingent artificial scotoma (n=9). Furthermore, we investigated whether the presence of a salient black-and-white striped border would improve performance regardless of target contrast. The task was to find a generic cylindrical target among similar distractors, distributed throughout a realistic scene. Participants saw a preview of the target object, then walked through the room and indicated when they detected the target object. Target contrast and the presence of the visibility enhancement were randomized across trials. Results show differential benefits from high target contrast and the visibility enhancement between the MD and control groups. Controls with the artificial scotoma performed better on average than MD participants, detecting targets faster and with higher accuracy, and did not benefit from the visibility pattern. The visibility pattern improved search efficiency for MD participants, reducing search time by 2.8 seconds, $F(1, 1) = 192.746$, $p = 0.045$. High-contrast targets significantly improved search performance for controls with the artificial scotoma, who detected those targets 5.6% more frequently, $F(1, 8) = 5.97$, $p = 0.04$, and 0.6 seconds faster, $F(1, 8) = 5.522$, $p = 0.047$, while MD participants saw no benefit. Thus, preliminary data suggest that high contrast, either from a target itself or an enhanced boundary, can improve search performance, but the benefit may depend on the nature and severity of the observer's central field loss.

This work was supported by the NIH grants T32EY025201 and R01EY027390.

26.326 ASSOCIATION BETWEEN REDUCED VISUAL SPAN AND WORKING MEMORY IN SCHIZOPHRENIA

Nwabunwanne Emele¹, (emelenwabu@gmail.com), Karin Reuwsaat², Gabrielly Mansour³, Maria Araujo⁴, Jeremy Wolfe⁵, Rogerio Panizzutti⁶; ¹Harvard Medical School, ²Federal University of Rio de Janeiro

Visual search efficiency relies on the ability to effectively allocate attention around the visual field to detect targets while suppressing distractors. Individuals with schizophrenia often show impairments in visual exploration, potentially reflecting a narrower visual span. Although visual span is believed to be influenced by higher-order cognitive processes, its relationship with cognitive functioning, symptomatology, and medication effects—particularly anticholinergic burden—remains unclear. **Methods:** In this cross-sectional study, we compared visual span during a conjunction search task between individuals with schizophrenia ($n = 21$) and age-matched healthy controls ($n = 21$). Visual span was operationalized as the distance of the fixation closest to the target prior to correct identification. Cognitive functioning was assessed using standardized measures of working memory and executive function. Medication effects, including anticholinergic load, were quantified to examine their contribution to visual span variability. **Results:** Individuals with schizophrenia exhibited a significantly narrower visual span than healthy controls, requiring closer fixations to identify the target ($F(1,40) = 12.27$, $p < .01$). Across participants with schizophrenia, smaller visual spans were significantly associated with poorer working memory performance. Higher anticholinergic burden was also linked to reduced visual span, suggesting that both cognitive capacity and medication-related effects influence visual search behavior. **Conclusion:** Findings confirm that schizophrenia is characterized by reduced visual span during conjunction search and demonstrate that this deficit is shaped by higher-order cognitive functioning and medication effects. Interventions targeting cognitive enhancement and minimizing anticholinergic load may help mitigate visual exploration impairments and improve functional visual processing in schizophrenia.

26.327 AUTISTIC AND NON-AUTISTIC ADULTS DIFFER IN SEARCH PATTERN, NOT SEARCH SPEED

Zoe (Jing) Xu¹, Bridget Leonard¹, Kristin Woodard¹, Hannah Rea¹, Sara J. Webb², Scott O. Murray¹; ¹University of Washington, ²Seattle Children's Research Institute

For the past 25 years, research has consistently shown that autistic individuals search faster than neurotypicals, but the underlying reasons remain unclear. Here, we conducted six experiments to clarify differences in search performance between autistic and neurotypical individuals. Participants (32 autistic and 40 non-autistic adults) completed four classic search tasks: a conjunction search (finding the letter R among Ps and Qs), a difficult feature search (ellipse among circles), a moderate feature search (letter N among Ps and Qs), and a color search with varying levels of difficulty (target color among distractor colors). We measured search time and eye movements. In two additional psychophysical experiments, we used a staircase procedure to estimate threshold display durations and color

discriminability during a color search task. Across the four classic search experiments, we did not observe significant group differences in search speed. However, there was a trend that autistic participants searched faster at a moderate-to-easy difficulty level (e.g., searching for a target color among a similar distractor color), but not in harder conjunction/feature searches or easier color-search conditions. This potential advantage may stem from their distinct search patterns observed in the eye-tracking data: at moderate difficulty, autistic participants used a larger functional viewing fields that covered more items, fixated for longer durations, and made saccades that covered a smaller convex hull area. Results from the psychophysical experiments also showed that autistic participants required shorter display durations at larger target eccentricities, suggesting faster peripheral visual processing. Together, these findings indicate that the critical distinction between groups lies in their use of peripheral vision: autistic individuals appear to rely more on parallel processing with peripheral vision and less on serial, item-by-item examination. Our results refine previous accounts of enhanced search performance in autism and suggest differences in underlying visual mechanisms, such as a peripheral vision preference.

NIH R01MH131595 to S.O.M. and S.J.W.

26.328 INVESTIGATING TRAITS OF ASD, ADHD, AND MIGRAINES ON VISUAL SEARCH PERFORMANCE IN IMMERSIVE VIRTUAL REALITY

Lex Stinson¹, Lindsay Houck¹; ¹Rhode Island College

Sensitivity to visual stimuli is often reported as a symptom of many conditions such as migraines, Autism Spectrum Disorder (ASD), and Attention-Deficit Hyperactivity Disorder (ADHD). Sensitivity to bright light and high color saturation are most common. Though these visual sensitivities are noticeable, there is little research on how they impact individuals during day-to-day visual tasks. This study explored how visual sensitivities associated with ASD, ADHD, and migraines affect performance on a visual search task in an immersive virtual reality (VR) environment. For each trial, the room brightness or color saturation was manipulated to high or low settings, and visual search performance was measured with reaction time (RT) and d' (d'). Other measures included questionnaires identifying traits of ASD, ADHD, general sensory processing, and migraine symptomatology. Preliminary analyses ($n=23$) found a correlation between sensory processing and d' ($r(136) = -0.21$, $p = 0.013$) with higher sensory sensitivity associated with better performance. ADHD traits also correlated with d' ($r(136) = -0.18$, $p = 0.031$) with fewer ADHD traits associated with better performance. For reaction time, there was a significant correlation with the number of days impacted by migraines ($r(136) = -0.17$, $p = 0.045$) meaning people with more days impacted by migraines responded more quickly. In the high saturation condition, more ASD traits were correlated with faster reaction times ($r(21) = -0.47$, $p = 0.025$). This suggests that certain environmental conditions can differentially impact individuals with more traits associated with ASD, ADHD, and migraines. Future analyses will investigate additional

sensory modalities, such as tactile and auditory processing, and their interaction with these and other individual difference measures. These findings have implications for those with sensory sensitivities associated with various conditions and the impact of visual perception and completion of day-to-day visual tasks.

The Rhode Island College Graduate Student Support Fund

Visual Search: Attention, memory, decision-making

SATURDAY, MAY 16, 2:45 – 6:45 PM, BANYAN BREEZEWAY

26.329 HOW CONTEXTUAL ASSOCIATIONS SHAPE VISUAL SEARCH ACROSS PROCESSING STAGES

Lu-Chun Yeh¹, Belma Seferovic¹, Marius V Peelen², Daniel Kaiser¹; ¹Justus Liebig University, ²Radboud University

Increasing evidence suggests that object semantics play a crucial role in guiding attention and shaping behavioral performance during visual search. Among such semantic influences on visual search, categorical effects have been widely examined. However, the influence of contextual associations — how objects commonly co-occur within specific scene contexts (e.g., a cooking pot and a rolling pin in a kitchen) — and their interaction with perceptual and categorical factors remain less understood. In this study, we independently manipulated contextual and categorical relationships between target and distractor objects in a visual search task. The stimulus set comprised 24 objects divided across two contexts (kitchen and garden) and two categories (tools and non-tools). Additionally, we conducted an independent same-different task ($N = 34$) to obtain continuous measures of perceptual target-distractor similarity for each stimulus pair. Behavioral responses and two event-related potentials (ERPs) components related to attentional allocation, the N1pc and N2pc, were then modeled using linear regression analyses incorporating contextual, categorical, and perceptual target-distractor similarity. The behavioral results ($N = 40$) showed that contextual associations influence search performance both independently and in combination with perceptual and categorical similarity. The ERP results ($N = 24$) indicated that contextual associations guide attention independently during early processing, as reflected in modulation of the N1pc component, and subsequently interact with categorical information, as reflected in modulation of the N2pc. Taken together, our findings demonstrate that contextual associations among real-world objects influence visual search across multiple processing stages, thereby supporting efficient attentional guidance in everyday visual tasks.

The authors thank Sirine Nouria for help with data collection. LCY is supported by the MSCA programme (101149060). DK is supported by the DFG (SFB/TRR135, KA4683/5-1, KA4683/6-1) and an ERC

Starting Grant (PEP, ERC-2022-STG 101076057). This work is further supported by the DFG under Germany's Excellence Strategy (EXC 3066/1 "The Adaptive Mind").

26.330 INDIVIDUAL CAPACITY LIMITS IN MEMORY-GUIDED VISUAL SEARCH

Anna Grubert¹, Ziyi Wang¹, Courtney Turner¹, Kaijie Zhang¹, Daisy Searle¹; ¹Durham University

Multiple attentional templates (mental target representations) can be activated in parallel during visual search when multiple targets are task relevant. Such target representations are held in visual working memory (vWM), which storage capacity is known to differ between individuals. We tested whether individual vWM capacity predicts an individual's ability to activate multiple attentional templates by correlating behavioural and electrophysiological measures of vWM capacity with indices of attentional selection efficiency. In a change detection task, participants memorised one, two, or three colours and then judged whether a subsequent test display was identical or contained a colour change. Individual vWM capacity was substantiated by means of Pashler's k and amplitudes of the contralateral delay activity (CDA). In a visual search task, participants reported the orientation (vertical/horizontal) of a target-colour bar presented among five differently coloured nontargets. One, two, or three possible target colours were cued before each block. Search efficiency, indicative of the ability to activate multiple templates simultaneously, was indexed by N2pc amplitudes. In addition, we assessed the time course of template activation by measuring N2pc components elicited in response to colour singleton probes presented prior to the search displays. Both tasks produced load effects: Increasing memory load yielded higher k values and larger CDA amplitudes in the change detection task, while increasing the number of target colours reduced N2pc amplitudes during search. Critically, substantial individual capacity differences emerged in the change detection task, and these differences predicted visual search performance reliably. Individuals with higher k values and larger CDA amplitudes showed smaller N2pc load costs, demonstrating their ability to search with multiple concurrent templates. These individuals also demonstrated earlier and simultaneous template activation in preparation for search. Together, our results suggest that individual vWM limits predict an individual's ability to activate multiple templates simultaneously in multiple-colour search.

This work was funded by a research grant of the Leverhulme Trust (RPG-2020-319).

26.331 AUTISTIC TRAITS ARE ASSOCIATED WITH REDUCED PRIMING EFFECTS WITHIN PERCEPTUAL DECISION MAKING

Emma Mueller¹, Miranda Scolari¹; ¹Texas Tech University

Priming occurs when a stimulus is repeated, leading to faster and/or more accurate responses to it. Individuals with autistic-like traits exhibit

behavior consistent with an underutilization of trial history (Pomè et al., 2020), but the underlying processes that give rise to this weakened priming effect remains unclear. We examined how visual priming influences distinct components of perceptual decision making, and whether they are associated with autistic traits. Twenty-five neurotypical college students completed an Autism Spectrum Quotient (AQ) survey and a visual search task. This task consisted of locating a bite mark (left or right) within a uniquely colored target apple presented in one quadrant among three distracting whole apples occupying the remaining quadrants. On half the trials, the color and position of the target apple repeated from the previous trial (repeat trials); for the remaining half, the color and position changed from the previous trial (switch trials). The location of the target bite mark was randomly selected on each trial. As expected, we observed a significant priming effect within reaction time (RT; $p < .001$). The RT and accuracy data were then submitted to a drift diffusion model (Ratcliff & Tuerlinckx, 2002) to investigate priming within decision processes. Among the estimated components, only non-decision time showed significant differences between repeat and switch trials ($p = .004$). This suggests that priming decreased the time spent searching for the relevant apple before accumulating evidence in favor of one of the two response options. Furthermore, non-decision time was moderately correlated with AQ scores for both repeat ($r = .405$) and switch ($r = .565$) trials; individuals with lower AQ scores exhibited a larger priming benefit in non-decision time than those with higher scores. These findings advance our understating of the priming effect on decision-making and its association with autistic qualities in neurotypical adults.

26.332 FRAMING EFFECTS IN EFFORT-REWARD TRADEOFFS DURING VISUAL SEARCH

Nusrat (Progga) Jahan¹, Brian Anderson², Darrell Worthy², Peter Kvam¹, Andrew Leber; ¹The Ohio State University, ²Texas A&M University

In our fast-paced world, visual search is often constrained by tradeoffs. For example, when rushing through a supermarket with a long shopping list, we may forgo our preferred brand when it is harder to find than a satisfactory alternative. The present study asks how people discount their reward preferences as a function of task demands during visual search. Participants ($N = 60$) first completed a Binary Choice Task where they established their preference among three color-coded options: a stable option (+60 points), a risky and variable option (+20/+60/+100 points; equiprobable), and a control option (0 points). Then they completed an Adaptive Choice Visual Search (ACVS) task where the same colors and rewards appeared, and participants freely chose which color target to search for, though one was sometimes easier to find than the other. We model behavior using a combination of expected utility and delay/effort discounting approaches to assess how reward history shapes attention and tradeoffs under uncertainty. Preference for consistent vs. variable reward schedules varied across participants, and participants were

explicitly aware of these preferences. These stable preferences generalized across the two tasks such that stronger initial preferences predicted a bias toward reporting similarly colored ACVS targets, even when searching for the preferred color would be less efficient. At the same time, participants appropriately adapted to differences in search difficulty, and this effort sensitivity was statistically independent of reward-schedule preference. These findings demonstrate that reward history and risk preferences jointly influence attentional strategies, such that individuals bias their search strategies in trial-by-trial decisions. We are currently running a second experiment with equivalent losses to understand how framing modulates these choices and visual search strategies.

26.333 PREDICTIVE CUES MEDIATE THE COST OF IMPRECISE ATTENTIONAL TEMPLATES IN VISUAL SEARCH

Jennifer Bartlett¹ (jdbartlett@mun.ca), Blaire Dube¹; ¹Memorial University of Newfoundland

During visual search, a representation of the search target (an attentional template) held in visual working memory (VWM) supports the automatic guidance of attention to feature-matching perceptual inputs. Previous work has demonstrated that the precision of these attentional templates - how closely the representation held in VWM matches the features of the search target - affects the efficiency of search performance: As templates become less precise, observers adopt broader attentional sets that guide attention less accurately. In the present study, we evaluated whether we could narrow the attentional set and mitigate the response time (RT) cost associated with using imprecise templates by providing predictive cues that would describe how the search target would differ from the initial template image. Participants searched for real-world objects/animal stimuli among an array of distractors under three cue conditions: precise template cue, where the target matched the initial template exactly ("search for this exactly"); imprecise template cue, where the target features were similar to the template, but with a change in state that was unknown to the observer (e.g., a template depicting a bird perched but a target stimulus that was a bird in flight; "search for something similar to this"); and predictive template cue, in which the target features were similar to the template, but the change in state from template to target was predictable based on a provided cue ("search for this in flight"). Predictive cues significantly reduced the RT cost seen when using imprecise templates, but RTs in the predictive template condition were still significantly higher than in the precise template condition. These findings suggest that predictive cues can narrow the attentional set somewhat, partially recovering the cost associated with using imprecise attentional templates and allowing for more efficient search; however, templates generated with predictive cues are less efficient than precise templates.

Natural Sciences and Engineering Research Council of Canada Discovery Grant

26.334 THE ROLE OF VISUAL WORKING MEMORY IN THE GUIDANCE OF ATTENTION FROM CONSISTENT TEMPLATE VALUES

Aditya Prakash¹, Andrew Hollingworth²; ¹University of Iowa

Attention during visual search can be guided by the content of visual working memory (VWM) and long-term memory (LTM). Previous research suggests that consistent target representations maintained in LTM can successfully control attention even when VWM is filled with an independent load (e.g., Woodman, Vogel, & Luck, 2001). In these experiments, although search RT intercepts increased with a VWM load, the slopes relating RT to set size did not, indicating that VWM load may have influenced initiation or confirmation processes but did not influence the guidance of attention during search per se. However, in these and related experiments, the template representation was primarily a “template for recognition” rather than a “template for guidance.” That is, the consistent target attribute (gap in a Landolt C) could not have provided strong guidance of attention, and thus the question of the role of VWM in attention guidance from consistent template values remains open. Here, we examined consistent color templates, which can provide efficient guidance to matching items. Participants searched for an object with a small target attribute within arrays composed of multiple objects drawn in two colors, a consistent target color and a trial-variable distractor color. We use eye-tracking to characterize the impact of a VWM load on selection dynamics. This allowed us to distinguish load effects on non-guidance processes (such as search initiation time and target confirmation time) from load effects on the guidance operation itself. In contrast with previous work, we found that a VWM load impaired template-driven selectivity during search, indicating that even with extensive repetition of a template value, the guidance function does not necessarily become fully independent of VWM.

26.335 THE LOW-PREVALENCE EFFECT VS. PREVALENCE INDUCED CONCEPT CHANGE: THE ROLE OF FEEDBACK IN VISUAL SEARCH FOR INFREQUENT TARGETS

Patrick Cox¹; ¹Lehigh University

When a target is rarely present in visual search, it is more likely to be missed—this is the Low-Prevalence Effect (LPE; Wolfe et al., 2005). Low prevalence typically reduces both hits and false alarms, suggesting that observers adopt a more conservative target-present decision criterion and shorten their search on target-absent trials, consistent with a lower quitting threshold (Wolfe & Van Wert, 2010). In contrast, other work shows that when examples of a category become rarer, the boundaries of that category expand—a phenomenon known as Prevalence-Induced Concept Change (PICC; Levani et al., 2018). These two effects appear to pull in opposite directions. How can they be reconciled? Prior work using single-item categorization tasks suggests that feedback plays a decisive role (Lyu et al., 2021). Here, I extend this idea to visual search, showing that feedback again

determines whether lowering target prevalence produces the LPE or PICC. In Experiment 1, participants (N = 106) completed a T-among-Ls search task with trial-level feedback in blocks of 80% and 20% target prevalence (block order counterbalanced). Results revealed a classic LPE: increased miss rates and reduced target-absent search times under low prevalence. Experiment 2 used the same design but without feedback. Participants (N = 98) showed no reliable effect of prevalence on miss rates or target-absent search times. Instead, they exhibited increased false alarms at low prevalence, consistent with PICC. These results suggest that with feedback, observers are confident in their knowledge of the underlying low-prevalence distribution and shift toward a more conservative target-present criterion (i.e., the LPE). However, without feedback, observers may believe they are missing targets and instead adopt a more liberal criterion (i.e., the PICC).

26.336 HERE'S THE T: REPEATED SEARCH TASKS RESULT IN REDUCED SEARCH EFFICIENCY.

Katherine Stuart¹ (katherine_stuart2@mymail.eku.edu), Steven Pesina¹, D. Alexander Varakin¹; ¹Eastern Kentucky University

Sometimes searching through the same configuration can increase efficiency (e.g. contextual cuing), while other times searching through the same configuration has no effect on efficiency (e.g. repeated search). Pesina & Varakin (2020; VSS) found that repeatedly searching through the same configuration where targets were placed in previous distractor locations resulted in reduced search efficiency. The current experiments seek to replicate and extend Pesina & Varakin (2020, 2025; VSS Poster). In the search task, participants (N=154) searched a visual display for a rotated T amongst rotated L's with varying set sizes (8, 12, or 16) over a series of 8 trials. In the target replaces distractor condition (TRD) stimuli were replaced in previous distractor locations, and in the target replaces empty condition (TRE) stimuli were replaced in previously empty locations. In our previous experiments, participants were informed that configurations would repeat for 8 consecutive trials and that the target location would not repeat. In the current experiment, participants did not receive these instructions - participants were only instructed to search for T's among L's. The results replicated those of previous experiments. In the TRD condition, search efficiency changed from ~40.6 ms/item to 52.5 ms/item from the first to the eighth repetition. In the TRE condition, the search efficiency changed from 47.0 ms/item to 38.5 ms/item from the first to second repetition, then changed to 50.0 ms/item by the eighth repetition. (Varakin, 2025) Our results are consistent with previous experiments, suggesting that the reduced search efficiency is not due to location-specific inhibition-carryover effects from trial to trial, and is not impacted by lack of instructions about the repetition.

SATURDAY AFTERNOON POSTERS IN PAVILION

Attention: Endogenous, exogenous

SATURDAY, MAY 16, 2:45 – 6:45 PM,
PAVILION

26.401 ENDOGENOUS AND EXOGENOUS ATTENTION REMAIN UNIFORM ACROSS CARDINAL MERIDIANS DESPITE DIFFERENTIAL ADAPTATION

Hsing-Hao Lee¹ (hsinghaolee@nyu.edu), Marisa Carrasco^{1,2};
¹Department of Psychology, New York University, ²Center for Neural Science, New York University

[BACKGROUND] Covert spatial attention and visual adaptation are two fundamental visual processes that can help manage the limited bioenergetic resources according to the task demand. Visual perception is heterogeneous around the visual field: it is better along the horizontal than the vertical meridian (horizontal-vertical anisotropy, HVA), and better along the lower than the upper vertical meridian (vertical meridian asymmetry, VMA), reflecting well-established performance fields. Neither endogenous (voluntary) nor exogenous (involuntary) attention alter performance fields, whereas adaptation is stronger at the horizontal than the vertical meridian, reducing the HVA. Here we ask, whether and how endogenous and exogenous attention affect the performance asymmetries before and after adaptation at the cardinal meridians. [METHODS] To equate discriminability across the four cardinal locations, we titrated contrast thresholds while observers performed an orientation discrimination task on a vertically-tilted Gabor ($\pm 2.5^\circ$, 5 cpd) in the neutral condition, with and without adaptation to 100%-contrast vertical Gabor adaptors. They then performed the same task with informative central (endogenous) pre-cues (Experiment 1) or uninformative peripheral (exogenous) pre-cues (Experiment 2) indicating the target location. The pre-cue either matched (valid-cue), mismatched (invalid-cue) or provided no location information (neutral-cue). Performance was quantified using d' . [RESULTS] Titration results were consistent with previous studies, showing (1) typical HVA and VMA in contrast thresholds in the non-adapted condition, and (2) stronger adaptation at the horizontal than the vertical meridian. Results from the attentional tasks showed that, in both adapted and non-adapted conditions, each endogenous and exogenous attention improved or impaired performance in the valid or invalid conditions, respectively, to a similar extent across cardinal meridians. [CONCLUSION] Despite the distinct flexibility and temporal dynamics of endogenous and exogenous attention, both types of attentional effects remained uniform across cardinal meridians before and after differential adaptation. Thus, they do not interact with adaptation around polar angles.

NIH NEI R01-EY027401 to MC. Taiwan's Government Fellowship for Studying Ph.D. Abroad, the Ministry of Education in Taiwan and 2025 National Science and Technology Council (NSTC) Taiwanese Overseas Pioneers Grants (TOP Grants) for PhD Candidates to HHL

26.402 ENDOGENOUS BUT NOT EXOGENOUS SPATIAL ATTENTION INFLUENCES GRADEDNESS OF VISUAL AWARENESS OF NATURAL SCENES

Suraj Kumar¹ (surajk22@iitk.ac.in), Narayanan Srinivasan¹; ¹Indian Institute of Technology Kanpur

Is visual awareness, especially perception of gist of natural scenes, graded? Also, does attention influence gradedness? Recently, we showed that perceptual load of a secondary task influences gradedness of visual awareness of scene gist. Awareness was more graded under high perceptual load condition compared to low perceptual load. Carrasco and colleagues have already shown that both exogenous and endogenous attention influence subjective appearance of content. However, it is still not clear whether exogenous/endogenous attention influence appearance in a graded or discrete fashion. Hence, we conducted two experiments in the current study in which we manipulated attention (exogenous cuing in experiment-1 and endogenous cuing in experiment-2). In both these experiments, we presented natural scene images with varying contrast to obtain psychometric curves while participants performed a scene discrimination (Type-1) task and perceptual awareness rating (Type-2) task. Exogenous cue was not predictive of spatial location, while endogenous cue predicted spatial location. A natural scene image (indoor or outdoor) was presented either to the left or to the right of fixation. A four-parameter model was fitted for valid, invalid, and neutral conditions in both the experiments. Previous studies have used changes in the slope of the psychometric function to infer for degrees/gradedness of awareness. As expected, we found differences in threshold as a function of cuing indicating that attention influences appearance. Moreover, endogenous cuing but not exogenous cuing significantly influenced the slope parameter of the psychometric curves for perceptual awareness ratings. Slope for invalid and neutral conditions were higher compared to the valid condition with awareness ratings; gist perception is less graded with endogenous attention. These findings show that while exogenous spatial attention does not influence gradedness of visual awareness of scenes, endogenous attention does. Taken together these results support graded theories of visual awareness and also show that attention influences gist appearance.

26.403 ISOLATING COVERT ATTENTION SHIFTS AWAY FROM SALIENT VISUAL INFORMATION USING A NOVEL "ANTI-SINGLETON" TASK

Jin Shi¹, Kirsten Adam¹; ¹Rice University

Long-standing accounts of attentional selection propose that visuospatial attention is captured by salient visual information, and attention must subsequently be disengaged and reoriented. Here, we have developed a novel task to measure the disengagement of covert attention from salient information. We call this task the "anti-singleton" task, because it is a covert attention analogue of the well-known "anti-saccade" task used to study overt shifts of attention (Hallett, 1978). On

each trial, participants viewed an array of 8 colored items arranged in a circle, and 1 of the items was a unique color singleton (e.g., 1 red item and 7 green items). Participants ($n=18$) reported the orientation of a line inside a target item (horizontal or vertical). Under the “pro-singleton” condition, the target was defined as the unique color singleton; under the “anti-singleton” condition, the target was defined as the item spatially opposite from the color singleton (i.e., 180 degrees around the circular array). In the pro-singleton condition, participants could shift attention directly to the singleton target. In the anti-singleton condition, participants had to first locate the singleton, and then rapidly shift attention away from the singleton to locate the target. Response times were substantially slower in the anti-singleton task (658 ms) compared to the pro-singleton task (1,008 ms; $p < .001$), consistent with an added shift away from the singleton. Pro-singleton trials elicited a robust N2pc (~180-280 ms), reflecting efficient target selection. Anti-singleton trials elicited an initial N2pc toward the color singleton (~180-280 ms), followed by a subsequent N2pc to the opposite hemifield containing the target (~280-400 ms). Together, our behavioral and electrophysiological results confirm that the anti-singleton task reliably measures serial shifts of covert attention toward and then away from salient information.

26.404 DO DEVIANT VISUAL CUES CAUSE DISTRACTION?

Nadja Jankovic¹, Rachel Yapp¹, Aaron Richardson¹, Vincent Di Lollo¹, Thomas M. Spalek¹; ¹Simon Fraser University

When a task-irrelevant cue (e.g., a tone) becomes familiar, ignoring it is easy. If it changes suddenly (e.g., increases pitch), it can cause distraction and slowed responses. This effect – called deviance distraction – is observed during visual tasks that are preceded by tones of either a standard pitch (e.g., 80% of trials) or a deviant pitch (20%). Yet similar conditions elicit the alerting effect: leading alerting cues quicken responses during visual tasks. Trials with alerting cues (on 50% of trials or more) are compared to no-cue trials. These two contradictory effects have been explained by invoking the same neural mechanisms: salient cues cause a brief increase in arousal by triggering locus coeruleus activity, which fires strongly to salient, novel, or task-relevant stimuli. Notably, distraction and alerting are rarely considered in tandem, and accounts reconciling both effects remain to be proposed. One possible source of the conflict is the different probabilities of cue occurrence: because deviant cues occur infrequently (20%) compared to alerting cues (50% or more), simply reducing the number of alerted trials may change the alerting effect to distraction. In Experiment 1, participants performed a pop-out search task that was either preceded by a visual cue (white flash, 20% of trials) or a blank interval (80%). We found no differences between cue conditions, but the cue reduced the response repetition effect, which has been reported in distraction studies. Experiment 2 had identical procedures except for the inclusion of a standard cue (black flash, 80% of trials). Surprisingly, there were no differences between cue conditions, and no response repetition interaction. Interestingly,

responses were significantly faster in E2 than in E1, indicating an overall alerting effect when a cue occurred every trial. The implications are discussed in the context of both deviant distraction and alerting.

Natural Sciences and Engineering Research Council of Canada

26.405 THE ROLE OF EXPLICIT AND IMPLICIT EXPECTATIONS FOR ATTENTION CAPTURE BY A NOVEL COLOR SINGLETON.

Gernot Horstmann¹, Stefanie Becker²; ¹Bielefeld University, ²The University of Queensland

An important question in attention research is to what extent attention is biased by bottom-up factors such as saliency versus top-down factors such as intentions and expectancies. A particularly relevant test is the surprise presentation of a critical stimulus. A surprise presentation means that the presentation is not expected in advance, and it is thus reasonable to assume that no intentions are directed to the critical stimulus. Surprise presentations may entail several components, such as not informing the observer of the critical stimulus, or rendering the critical stimulus novel in the sense that its features have not been presented before, and therefore have no top-down utility or history. Previous research found that announcing the critical stimulus, rendering it expected in some sense, has almost the same strong attentional prioritization effects as when the critical stimulus is not announced and thus unexpected. Here we measured eye movement parameters as a means to disentangle the predicted pre-attentive and post-attentive prioritization of the surprise stimulus, that is, the capture and binding of attention, respectively. We compared two groups with the same novel color singleton, where the change in the critical trial was announced in one group but not in the other. The novel singleton was strongly prioritized in the critical trial. The announcement of the change had little effect on pre-attentive prioritization, but shortened the post-attentive prioritization. Overall, the results suggest that implicit expectations generated during on-task experience and the incidental learning of statistical regularities are more important for the attentional prioritization of a surprising stimulus than explicit expectations generated by verbal instructions.

26.406 HEMISPHERIC ASYMMETRIES IN VISUAL CROWDING

Valorie Wiseman¹, Summer Sheremata², Sang Wook Hong¹; ¹Florida Atlantic University, ²Northeastern University

Visual Crowding occurs when a target that is easily identified in isolation becomes difficult to recognize when surrounded by distractors, reflecting a fundamental limit of attention. Although visual-field (VF) asymmetries in crowding have been attributed to reading, prior work with non-linguistic stimuli showed Left/Right asymmetries are not language-dependent and rightward crowding is modulated by attention. Additionally, hemispheric asymmetries have been reported in fMRI and EEG studies in response to high-task demand in visual attention tasks. Here, we hypothesize that visual attention is

asymmetrically organized and may underlie the mechanisms of visual crowding. However, neural data is required to confirm this effect. In the current work, we predict that if the right hemisphere preferentially processes crowded visual information under high task demand, then ERP components of visual attention will be greater at the right electrodes compared to the left, and that this effect will be further enhanced when the pre-cue is absent. We conducted a pilot EEG study to measure three indices of visual attention: the N2pc, SPCN, and P3b. Formal statistical analyses were not conducted due to the limited sample size ($N = 2$). For the N2pc, mean amplitudes for Left and Right Target with Cue Present were 0.41 μV and 0.87 μV , respectively, while amplitudes for Left and Right Target with Cue Absent were -0.44 μV and -0.18 μV . SPCN amplitudes ranged from -0.99 μV to 0.59 μV across conditions, and P3 amplitudes ranged from -2.44 μV to 3.13 μV . This project found that all three ERP components exhibited a trend towards larger amplitudes in the right electrodes compared to the left, with an additional increase in the absence of a pre-cue. These pilot results suggest rightward hemispheric biases in crowding, highlighting the role of attentional asymmetries in visual crowding and attentional processing.

26.407 ATTENTION IN VIRTUAL REALITY REAL-WORLD ENVIRONMENTS

Beatriz Gil-Gómez de Liaño¹, David Pascual-Ezama; ¹Universidad Autónoma de Madrid, ²Universidad Complutense de Madrid

Virtual Reality (VR) has changed science in the last decade, significantly transforming applied sciences, industrial settings, and society. In the present work, we adapted ANTI-V (Roca et al., 2011) and ANTI-VEA (Luna et al., 2018) experimental tasks based on Posner's theory to develop a behavioral assessment in the industry using VR: We created a VR real-world-like task for human resource selection and occupational risk prevention in which participants had to maintain a proper tank pressure by manipulating key-pressure tools arrow-like shaped, in a similar way as in ANTI-VEA Posner tasks. We run participants in different 2-D ($n=38$) and 3-D VR ($n=30$) tasks, getting measures of visual selective attention, visual attentional capture, auditory attentional capture and vigilance. The results validated classic Posner effects for measuring these variables within a VR industry selection task, showing the potential of VR in applied psychology using experimentally validated paradigms in cognitive attention science.

Attention: Features, objects

SATURDAY, MAY 16, 2:45 – 6:45 PM,

PAVILION

26.408 A SERIAL ATTENTIONAL BOTTLENECK IN TEXTURE PROCESSING: EVIDENCE FROM A DUAL-TASK CONTINUOUS-REPORT PARADIGM

Mary Catington¹, Michael Pratte¹; ¹Mississippi State University

Recent work suggests that simple visual features, such as colors, can be attended to in parallel across items. However, more complex information suffers from a serial processing limitation. For example, you can only read one word at a time, and have difficulty perceiving more than one object simultaneously. Here we explore the degree to which mid-level visual features like texture can be processed in parallel across objects. The results suggest that attempting to perceive two textures simultaneously incurs a significant performance cost compared to perceiving only one texture. The critical question is how exactly performance suffers when two visual stimuli must be attended: are you able to build less precise representations of both items, or is the trade-off discrete such that on some trials you can perceive both stimuli with high precision, but on other trials perceive only one? To address this issue we developed a task in which participants reported the orientation of a single texture, or of two textures presented simultaneously. Performance is worse when two textures must be perceived, and the accuracy of reports implies a discrete model: performance on some trials is high for both stimuli, but on other trials only one stimulus is perceived accurately while the other fails to be seen at all. These results support the proposal by Popovkina, Palmer, Moore and Boynton (2021) that the attentional bottleneck is serial, but that in some cases attention can shift rapidly across items such that both are perceived with high resolution. Our findings provide evidence that this serial bottleneck is present as early as mid-level visual texture segmentation, not just in higher-order object recognition.

This work was supported by National Institutes of Health Grant R15MH113075.

26.409 ATTENTIONAL TEMPLATES FOR VISUAL MOTION

Xiaoli Zhang¹, Taosheng Liu¹; ¹Michigan State University

Visual attention is guided by attentional templates that store task-relevant information. It has been shown that the maintained templates can capture attention during a secondary search task where the template feature is task-irrelevant, a phenomenon we refer to as template-driven capture. This capture effect has been used to assess the characteristics of attentional templates for static features such as color and orientation. Importantly, it remains unknown how attentional templates encode dynamic features such as visual motion, given that motion direction can also guide visual search effectively (e.g., Girelli & Luck, 1997). This study aims to explore attentional templates for motion direction using the template-driven capture paradigm. Participants were cued to attend to a particular motion direction and perform a sustained feature-based selection task. The selection task was intermixed with a visual search task containing an irrelevant distractor that either matched or mismatched the cued direction. We used the performance difference between these conditions to index the guiding strength of the attentional template. We found template-driven capture for motion direction, indicated by a higher inverse efficiency score in matched than mismatched search trials. A follow-

up experiment was conducted to explore whether motion direction is automatically recoded into orientation information during maintenance. Current results suggest that orientation features alone only produce salience-driven capture but not template-driven capture. Our results demonstrate that motion direction is actively maintained in attentional templates, providing a potential mechanism for effectively utilizing dynamic information to guide attention. More broadly, our study extends the template-driven capture paradigm to dynamic information and opens up new avenues for further exploration into the properties of attentional templates across diverse features.

This work is funded by NIH R01EY032071 to TL.

26.410 DEPLOYING COVERT ATTENTION IN THE VICINITY OF FIXATION: COLOR SEARCH AND SHAPE SEARCH PRODUCE DIFFERENT RESULTS

Cailey Tennyson¹, Jeremy Wolfe^{1,2}; ¹Brigham and Women's Hospital, ²Harvard Medical School

Observers can process about 4-6 items in the vicinity of a single fixation. In the present experiment, observers moved their eyes to a fixation point that appeared at different locations on a screen. After 300 msec, a ring of 8 stimuli was flashed around the point and participants had to indicate the orientation of a target object using a 4AFC decision. Exposure duration was staircased (3-up, 1-down) to produce ~25% errors. In prior work, more errors fell on the vertical meridian than horizontal. Of more interest, individual patterns of errors showed significant idiosyncratic deviations from this average pattern. Previous studies involved targets defined by shape – TvL, hammers among tools. In the present work, the target was defined solely by color; blue rectangles among distractors of yellow and red. Observers indicated the 4AFC orientation of the blue rectangle. Participants ran 400 trials for each task, alternating tasks in 200 trial blocks. Using a TvL task, we replicated past findings with most observers showing idiosyncratic deviations from the average result. These idiosyncratic patterns were consistent between sessions with an average $r=0.71$ correlation. Interestingly, results for the color task were quite different. The average pattern of errors was different from the TvL pattern with most errors in the lower left. For individuals, color search produced different error patterns from their TvL search, with only $r=.12$ correlation between the two tasks. For each observer, the two sessions of color search were not well correlated with each other. Finally, there was no evidence of idiosyncratic patterns in the color task. The differences between tasks might be related to task difficulty. The staircased threshold for the color task was much shorter than for the TvL task, indicating that the color task was easier. Alternatively, patterns of attentional deployment might differ between tasks involving different features.

NEI EY17001

26.411 DYNAMICS OF SPATIAL AND OBJECT-BASED ATTENTION

John Gonzalez-Amoretti^{1,2}, Adam Snyder^{1,2,3}; ¹Neuroscience Graduate Program, University of Rochester, ²Brain and cognitive Science, University of Rochester, ³Center for Visual Science, University of Rochester

Selective attention facilitates behavior by filtering out visual information and prioritizing goal-relevant stimuli. These processes depend on internal templates that inform the identity of selection targets, facilitating the guidance of attention to their possible locations. However, the neurocomputational mechanisms of how these templates are formed and then accessed to guide attention remain unclear. We propose that attentional templates are generated by transforming a sensory code to a task-ready latent code held in higher-level areas such as frontoparietal attention networks. A latent template representation provides a flexible code that can be integrated with other processes, such as spatial orienting and saccade planning, or retransformed into a sensory code for evaluating sensory input. To study this, we designed a task that distinctively engages spatial attention, object-based attention or a combined process based on cueing conditions. Unlike typical paradigms that rely on exact template matching, or low-level feature cues, such as colors or shapes, our design promotes higher-order search strategies by implementing categorical template matching with grayscale images. Latent factor analysis of human EEG data revealed varying spatiotemporal dynamics across conditions, suggestive of flexible coordination between visual and frontoparietal networks that changes with task demand. Image cues evoked activation of occipitotemporal regions that spread to frontoparietal regions during delay periods, suggesting that identity information is projected into higher-order regions. Spatial cues showed a reverse effect with centroparietal activation preceding occipital activation after cue onset, likely leveraging retinotopic maps for anticipation. The combined spatial-image cue recruited both pathways, revealing shared dynamics and suggesting that both processes can unfold in parallel. Moving forward, we will focus on multivariate pattern classification and representational similarity analysis to determine whether these dynamics play a role in transforming sensory codes into a latent template representation.

26.412 EFFICACY OF OBJECT-BASED ATTENTIONAL SELECTION AGAINST DISPROPORTIONATELY SALIENT DISTRACTORS ALONG THE VENTRAL VISUAL PATHWAY

Lasyapriya Pidaparthi¹ (lasyapriya.pidaparthi@vanderbilt.edu), Frank Tong¹; ¹Vanderbilt University

In our daily lives, we are often confronted with salient distractors that we must effectively ignore in order to focus on our task. Research has shown that object-based attentional modulation can selectively boost the neural responses in both low- and high-level visual areas to prioritize task-relevant information, even in the presence of a competing, overlapping distractor (Cohen & Tong, 2015; O'Craven et al., 1997). However, we do not fully understand how this attentional modulation is affected by distractor saliency, particularly when the distractor is overwhelmingly more salient than the target. Here, we

investigated this question by using an fMRI decoding-based approach (Kamitani & Tong, 2005) to examine top-down attentional biasing of neural responses along the ventral visual pathway, including early visual areas (V1-hV4) and category-selective regions (fusiform face area/FFA; parahippocampal place area/PPA). In our experiment, participants were presented with overlapping face-house images that varied in their relative contrast. On each trial, participants attended to either face or house and indicated whether the attended object across sequential pairs of images was the same or different. On some trials, only a single face or house was presented (“attend-single”). To measure object-based attentional selection, we trained an SVM classifier on the normalized multivariate responses during attend-single trials and tested how well it could decode the attended object from responses during attend-face and attend-house trials. Notably, modulatory effects of object-based attention were found throughout visual areas V1-hV4 and FFA/PPA, evidenced by greater than 50% decoding bias for the target, even when the distractor was twice as strong as the target in stimulus contrast. Further, we observed a trend of strengthening attentional modulation as one ascends the visual hierarchy. Overall, our novel findings showcase the efficacy of object-based attention in overcoming stimulus-driven effects to strengthen the neural representation of an attended object.

This research was supported by NIH grants R01EY035157 and R01CA240274 to FT, T32EY007135 to LP, and P30EY008126 to the Vanderbilt Vision Research Center.

26.413 INCIDENTALLY LEARNED AND EXPLICITLY CUED DISTRACTOR IGNORING DURING VISUAL SEARCH IN OLDER AND YOUNGER ADULTS

Pun Punyawish¹ (pun@docchula.com), Viola Stoermer², Chaipat Chunharas^{1,3}; ¹Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand, ²Department of Psychological and Brain Sciences, Dartmouth College, ³King Chulalongkorn Memorial Hospital, Thai Red Cross Society, Bangkok, Thailand

Previous research has provided mixed evidence as to how well older adults successfully ignore frequent distractors during visual search (Lega et al., 2023; Yao et al., 2025). In younger adults, studies indicate that the effectiveness of distractor ignoring may depend on the mechanisms engaged, in particular whether participants are explicitly cued to ignore specific features or are incidentally learning about likely distractors (e.g., Addelman & Störmer, 2023). We here tested two questions: (1) whether distractor ignoring is preserved in old age for cued and learned attention, and (2) whether its effectiveness is modulated by the usefulness of the feature regularities in mitigating distraction. Younger (n=46) and older adults (n=43, age>60) performed a block design visual search task looking for a target (left/right gap Landolt C) among other 7 distractors (top/bottom gap). Color probabilities of the items were manipulated such that the distractors occurred more often in one out of four colors used across the experiment. Participants either implicitly learned the color probabilities (learned ignoring; Exp. 1) or explicitly informed about

these probabilities (cued ignoring; Exp. 2). We also varied the distractor-to-target ratio on each block, such that there were either 4 or 6 high-probability-color distractors present in the search array. We found that both age groups were faster in finding the target on trials with the frequent distractor present, indicating that explicit cue and incidental learning of distractor features speeds response times (RT) during search. Interestingly, only older adults showed stronger RT benefits for trials in which more high-probability-color distractors were present in the search display. Overall, our results suggest preserved distractor suppression in aging, with an age-specific sensitivity to the distractor-target ratio. This may suggest that older adults up-weighted the color-probability information to support performance when particularly useful.

26.414 FROM RETINAL SIZE TO SEMANTICS: INTEGRATION OF SIZE INFORMATION INTO OBJECT-REPRESENTATIONS

Kelly McEvoy¹, Dick Dubbelde², Sarah Shomstein¹; ¹The George Washington University, ²Georgetown University

Semantic knowledge of an object's real-world size influences how attention interacts with objects, yet the mechanism underlying these effects remains unclear. Real-world size is a unique representational dimension because it can be encoded in two dissociable ways: retinal size, reflecting the visual angle projected on the retina, and inferred size, reflecting observers' knowledge of an object's true physical dimensions. Whereas retinal size is a low-level sensory measure, inferred size reflects higher-level contextual information. We hypothesize that real-world size shapes both perceptual processing as well as attentional selection. Experiment 1, using rectangles of different retinal size, replicated the classic finding that targets appearing within larger objects are processed less efficiently than those within smaller objects, demonstrating a robust effect of retinal size on perceptual processing. Experiment 2 tested whether this size-based cost also emerges when only inferred size differs. Using line drawings matched in retinal extent, responses were slower for semantically large objects (e.g., door) compared to semantically small ones (e.g., domino), indicating that a size effect originally arising from low-level inputs also manifests when size is defined purely by semantic knowledge. Together, these experiments support the hypothesis that attentional effects elicited by retinal size become integrated into higher-level object representations. Experiment 3 asked whether inferred size further constrains attentional allocation by probing object-based attention. In an adapted two-rectangle paradigm using small- and large-sized real-world object pairs, robust object-based effects were observed. Critically, the magnitude of the same-object advantage was significantly smaller for semantically small than semantically large objects. This pattern indicates greater attentional prioritization for smaller inferred-size objects. Across three experiments, our findings show that real-world size consistently modulates attentional performance, suggesting that size, both retinal and inferred, is

embedded within object representations and systematically governs how attention is deployed.

NSF BCS 2022572 to SS

26.415 INVESTIGATING ELECTROPHYSIOLOGICAL MARKERS OF LEARNED ATTENTIONAL SELECTION DURING REAL-WORLD OBJECT SEARCH

Yifan Fang¹ (yifan.fang.gr@dartmouth.edu), Kevin Ortego¹, Viola Störmer¹; ¹Dartmouth College

People learn from statistical regularities in their environments in support of more efficient behavior. For example, during visual search, people can rapidly learn which color or object frequently associates with a task-relevant target, improving search performance. Previous studies using simple features like color suggested that early and later stages of neural processing are involved in this learned prioritization (Ortego et al., 2024; Wang et al., 2023). Here, we investigate the neural underpinnings of learned attention during real-world object search. Participants ($n = 24$) performed a visual search task to indicate the direction of a tilted object (target) among three other upright objects (distractors) while EEG was recorded. On each trial the same four objects were present, but unbeknownst to the participants, one object appeared more frequently as the tilted target object during an initial “training phase”, which was followed by a “testing phase” where target frequencies were equal across all four objects. We found that participants rapidly learned these statistical regularities, reflected in faster response times during search for the frequent relative to infrequent objects during the training phase. These behavioral benefits were accompanied by changes in the N2pc component, an index of early attentional selection, and the later LPC component, demonstrating that selecting one specific object more frequently modulates neural processing and behavior. However, these learning effects did not extend reliably into the testing phase, suggesting that attentional tuning to specific objects may quickly adjust to the current regularities in the environment. Together, our results suggest that similar neural processes may give rise to how selection history influences search behavior for both simple features and complex, real-world objects.

26.416 SURFACE VS SEMANTIC FEATURE PRIORITY IN PHISHING EMAIL JUDGMENTS

Corey Bohil¹; ¹Lawrence Technological University

A debate in phishing detection research is whether e-mail observers allocate attention primarily to semantic or surface-level features. This study examined the relative salience of visible surface cues (abnormal e-mail structure, urgent-action hyperlinks) versus abstract semantic cues (premise plausibility, personalization, disproportionate benefit) and their relation to judgments of message authenticity and danger. Fifty-eight participants (model training $n = 46$, test = 12) evaluated 50 emails (2,900 observations), rating the salience of the five features and perceived levels of authenticity and danger. To understand causal

influences on these judgments, a set of Bayesian network models were fit to the discretized (low, medium, high) rating data, separately for authenticity and danger judgments. Three theoretically-motivated models were compared: a Surface Features model (with nodes for actionable links and abnormal structure), a Semantic Features model (with nodes for benefit, personalization, & plausibility), and an Interaction model, with one surface-feature node (links) and one semantic-feature node (benefit). For both authenticity and danger judgments, the Surface Features model provided the most parsimonious account. Despite its simplicity, this two-feature-node model achieved moderate predictive power (for authenticity and danger judgments respectively: $F1 = .65, .60$; average $AUC = .79, .76$; Accuracy = $.65, .60$). The Surface Feature model also demonstrated superior generalization compared to a more complex data-driven model that included nodes for all 5 predictive features. These results suggest the primacy of visible, concrete email features in observer judgments, supporting a “heuristic” account of risk assessment (i.e., judgments based on the presence of superficial cues). The model-based approach used here offers a rigorous basis for future research examining how cognitive states (e.g., fatigue, high workload) modulate feature salience.

26.417 TRACKING THE COMPETITION BETWEEN NOVELTY AND MEMORY FOR VISUAL ATTENTION

Dana Slabbekoorn¹ (slabbek2@uwm.edu), Eunji Back¹, Deborah E. Hannula¹; ¹University of Wisconsin - Milwaukee

Novel objects are strong attractors of attention and are viewed preferentially in the visual paired comparison task. At the same time, a growing body of work demonstrates that long-term memory can bias gaze toward the learned associate of a memory cue. The present study examined how novel items and associates compete for attention. Participants completed an incidental encoding task in which they viewed scene-face pairs and were asked to consider their fit. In a subsequent indirect test, each trial began with a studied scene, followed by the presentation of a 3-face display superimposed on the scene. One of these faces had been paired with the scene (i.e., the associate), another was familiar but had been paired with a different scene, and the remaining face was new. Participants were instructed to view the display and eye movements were recorded. No reference was made to memory. After incidental viewing, participants completed two explicit recognition tests. In one test, they attempted to identify the novel face and in the other they identified the associate; confidence judgments were made after each recognition response. During the incidental test block, participants spent more time looking at associates in the first gaze than at either of the remaining two faces. They also made more transitions to the associate and spent more time over the course of the entire trial viewing that face compared to the other two. When time-course analyses were performed, results indicated that disproportionate viewing of the associate was significant within the first second of the 3-face display onset and remained high for the rest of the trial. The only measure that was not sensitive to

learned associations was the landing position of the first saccade. Together, these results demonstrate that under conditions of direct competition, associates are stronger attractors of attention than novel items.

26.418 ATTENTIONAL LOAD SIMILARLY AFFECTS ENSEMBLE AND ITEM EMOTION PROCESSING: INSIGHTS FROM OCCIPITAL ALPHA BAND SUPPRESSION

Bradley E. Buchanan¹, Laramie L.J. Starling¹, Jacob Zepp², Rachel Gaynor¹, Gavin Heindorf¹, Harold Rocha¹, Chad Dubé¹; ¹University of South Florida, ²University of Zürich

When teaching in front of a class, it helps to get an idea of how the class is feeling so that we may adjust our teaching styles to engage those who may be less interested. However, our processing of group emotion, or Ensemble Emotion Processing (EEP), is not fully developed. More specifically, current models of EEP question the role of attention when making ensemble judgements of emotions. Previous work by Liu and Ji (2024) attempted to characterize the role of attention in EEP without neural markers. They reported the absence of an effect of attentional load on EEP in behavioral markers. In response, we carried out a replication of the author's task while measuring occipital alpha band (10-12 Hz) suppression, a direct neural marker of selective attention, to assess their findings. Bayesian linear-mixed effects models reveal that increasing attentional loads reduced alpha band suppression but, interestingly, suppression was not modulated by EEP conditions nor was there any interaction of attentional load and EEP on alpha band suppression. These findings suggest that selective attention during EEP may not differ than individual item processing and are similarly affected by increasing attentional loads. Altogether, these findings provide strong evidence that EEP occurs without the need of additional selective attention resources even when selective attention capabilities are limited by load.

26.419 INTERACTION MODULATES VISUAL INTERPRETATION AND TRUST IN UNCERTAINTY VISUALIZATIONS

Songwen Hu¹, Alex Endert¹, Cindy Xiong¹; ¹Georgia Institute of Technology

Forecast visualizations convey uncertain information intended to guide decision-making, yet the way this information is visually encoded can strongly influence how people interpret and trust the data. Some studies find that viewers place greater trust in simple, deterministic displays that emphasize one clear pattern (Elhamedi et al., 2023), whereas others show that visualizing uncertainty explicitly can foster more calibrated trust and more appropriate decisions under uncertainty (Padilla, Ruginski, & Creem-Regehr, 2017; Witt & Clegg, 2022). Such variability suggests that perceptual and cognitive responses to uncertainty visualization are not uniform across observers but may depend on individual differences in how visual information is processed and integrated into judgment (Grounds &

Joslyn, 2018). We drew on human-factors and visualization research to test whether interaction design could reconcile individual differences by allowing viewers to control how much uncertainty information they see. We designed interactive visualizations that either (1) communicated forecast uncertainty through confidence intervals (aggregated uncertainty) or (2) explicitly displayed multiple forecasts (individual model predictions). We compared these with non-interactive versions that fixed the display but contained the same information. Participants viewed humidity forecasts for two cities, made predictions about future trends, and rated their confidence and trust in the visualization. Results showed that participants trusted predictive visualizations and underlying data more when interactions were well designed than when they were static (MD=0.667, $p=0.012$). Interaction also altered participants' predictions, leading them to perceive larger differences between the cities' projected humidity values (MD=5.763, $p<0.001$), though confidence ratings were unaffected (MD=0.267, $p=0.229$). These findings demonstrate that interaction modulates how viewers perceive and mentally represent forecast uncertainty. Future work can further examine how interactive uncertainty visualizations shape attentional control and allocation during data foraging to better align interface design with the visual system's capacities for probabilistic reasoning.

26.420 THE IMPACTS OF ACUTE STRESS ON VISUAL ATTENTION AND MEMORY DURING AN EYEWITNESS EVENT

Keely Esch¹ (keelynesch@gmail.com), Abigail Kortenhoeven¹, Michael Serra², Miranda Scolari¹; ¹Texas Tech University, ²Texas Tech University Health Sciences Center

Stress can have negative consequences on cognitive processes, including memory and attention. For example, meta-analysis revealed that stress negatively affects the accuracy of eyewitness recollection (Deffenbacher et al., 2004). Individuals experiencing high levels of acute stress may also experience attentional bias toward threatening stimuli, leading to higher detection rates for such items compared to their non-threatening counterparts (Putwain et al., 2011; Rued et al., 2019). To determine how stress may impact both visual overt attention and memory in the same task, we manipulated acute stress before presenting participants with a brief video of a criminal event. We administered the Maastricht Acute Stress Task (MAST; Shilton et al., 2017) to induce acute physiological stress within the experimental group ($N = 28$), and compared their performance to a non-stressed control group ($N = 30$). Both groups then watched publicly accessible CCTV footage from the Denver Police Department of a convenience store armed robbery while we tracked their eye movements across key elements of the scene. Self-report measures of stress revealed that the experimental group experienced higher levels of stress compared to the control group, as expected. The experimental group spent significantly more time fixating both relevant, threatening elements in the scene compared to the control group (including a robber and a robber's gun), as well as relevant non-threatening items (the cash

register). They also made more regressions to the employee's face compared to the control group. Despite these fixation patterns to relevant elements, the experimental group performed marginally worse on a subsequent memory test of the event ($p = 0.07$). These results suggest that even just 10 minutes of physiological stress can impact the way individuals both attend to and remember important events.

Perceptual Organization: Features, parts, wholes, objects

SATURDAY, MAY 16, 2:45 – 6:45 PM,

PAVILION

26.421 EXPANSION OF PERCEIVED SPACE AND DISTORTION OF POSITION WITHIN A FIGURAL REGION

Abigail Way¹ (away@udel.edu), Ethan Thomas¹, Emma Sutcliffe¹, Timothy Vickery¹; ¹University of Delaware

“Object-based warping” occurs when two reference dots placed within an object are perceived as further apart than equally spaced dots placed on a ground region. Are these distortions due to misperceptions of the locations of the individual dots within the object, or is the illusion due to a misperception of their relative spacing alone? We previously found that expansion effects occur for figure regions compared with ground regions in classic figure-ground frame stimuli with alternating regions, where each region appears as figure or ground due to cues like convexity (Vickery, Sweetman, & Peterson, VSS 2025). Such stimuli offer novel potential to test whether position distortions occur, since all dots appear within the same rectangular frame. In one in-person ($N=25$) and one preregistered online experiment ($N=68$), participants viewed visual displays that were divided into alternating black and white regions, with convex or concave curves flanking each region and thus inducing alternating figure/ground percepts. To assess spacing, two reference dots were placed on either figure or ground, and two movable dots were placed in the opposite type of region. To assess position, one reference dot was placed on a figure or a ground region, with one movable dot on the opposite. Subjects matched spacing or position. Replicating earlier results, spacing judgments were larger when reference dots were on figure vs. ground ($p < .001$). Notably, position was also distorted when the reference dot was placed on a figure vs. ground ($p < .001$). Congruent with the spacing effects, figure dots were reported as nearer the outside boundary than ground dots. Additionally, spatial and position distortions were correlated (online study: $r = .26$, $p < .05$). These findings suggest that distortions in perceived spatial relationships due to figure assignment can at least partially be attributed to misperception of each dot's position.

26.422 INTERACTIONS OF PERCEIVED NUMBER, DENSITY, AND COMPRESSION OF VISUAL SPACE

EXPLAIN IMPROVED FEATURE DISCRIMINATION IN REDUNDANCY MASKING

Doğukan Nami Öztaş¹ (oztasdogukan@gmail.com), Li L-Miao², Nihan Alp¹, Bilge Sayim²; ¹Sabanci University, Istanbul, ²École Normale Supérieure, PSL University, CNRS

In redundancy masking (RM), the number of items in repeating patterns is underreported. For example, three identical peripheral items are often reported as two. RM is associated with a radial spatial compression (the perceived radial extent of stimuli is reduced) and an inflated item-to-item spacing (the perceived distance between items is increased). Here, we hypothesized that RM would improve feature discrimination of the perceived items. In two experiments, we presented arrays of 3-5 identical, radially arranged bars with uniform spacing, randomly to the left or right of fixation. Observers reported the number of bars they perceived. In Experiment 1, a foveal probe containing the reported number of bars was presented, and observers adjusted both perceived width and spacing to match the presented array. In Experiment 2, a single-bar probe was presented, and observers adjusted its perceived width. Participants reported fewer items than presented, showing strong RM in both experiments. In Experiment 1, when the number of bars was underreported (RM-trials), observers reported wider –and more accurate– bar width, and larger – but less accurate– spacing than in No-RM-trials (correct number reports). In RM-trials, the reported horizontal extent of the arrays was reduced, showing radial compression of visual space. Importantly, perceived density was similar in RM-trials and No-RM-trials, and close to the presented stimulus density. In Experiment 2, the reported width difference between RM- and No-RM trials was reduced compared to Experiment 1, but still closer to the correct width in RM-trials. Together, our findings indicate that RM may enhance feature perception, possibly directly linked to spatial compression and accurate density perception observed under RM. It seems that RM preserves global density perception by jointly compressing space and the number of items. We suggest that RM is a mechanism by which the visual system compresses information while preserving critical stimulus features.

This study was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under the Grant Number 122N748.

26.423 STRIKING NONUNIFORMITIES IN THE PERCEPTION OF RANDOMNESS

Michaela Bocheva¹, Brian Scholl¹; ¹Yale University

One of the foundational tasks of perception is to detect and characterize “structure” in our local environments—in the form of objects, events, and even abstract statistical regularities. And this can be explored indirectly by asking how people perceive randomness (that is, the lack of structure). Many studies over many decades have explored the ability to detect “whether” a sequence of items is random or not, in a categorical sense. Here, in contrast, we ask how people perceive different “degrees” of randomness, and we provide some

counterintuitive answers. People viewed coin-flip outcomes, one by one, and used a slider to report how random the sequence appeared. Different sequences had varying degrees of actual randomness, as determined by bits per symbol, and the probability of an outcome repeat. This allowed us to ask: Are people equally sensitive to all degrees of structure? Or might they be especially sensitive only to whether a sequence has any structure (or no structure at all) — or to whether a sequence is more or less structured than some average value? In fact, none of these were true. Instead, certain degrees of structure were far more discriminable from each other, even while equating objective differences. For example, when characterized in terms of average information, people were highly sensitive to the difference between 0.6 and 0.7 bits—while showing no ability to discriminate between other similarly adjacent values. This pattern was surprising even to observers: we collected continuous confidence ratings for each of the randomness judgments, and these diverged from the actual patterns of sensitivity in qualitative ways. In particular, people were confident for extreme degrees of structure, but *not* for extreme degrees of randomness. We conclude that people are sensitive to structure as a continuous variable in nonuniform ways that prioritize some degrees of structure over others.

26.424 WATCHING CHAOS UNFOLD: ASYMMETRIES IN ATTENTIONAL EFFORT WHEN MOVING TOWARDS ORDER VERSUS CHAOS

Skadi Gerkenmeier¹, Joan Danielle K. Ongchoco¹; ¹University of British Columbia

In perception, visual processing strives towards order, both in space (via object segmentation) and in time (via event segmentation). In physics, however, systems strive towards chaos, and the reverse—maintaining order—costs energy. Here we asked how this fundamental law of entropy manifests in visual processing. Might there be an asymmetry in the effort it takes to attend to order vs. chaos? Inspired by physical systems, we turned to a curious phenomenon: the pendulum wave. It consists of a row of pendulums, swinging with identical amplitudes but monotonically increasing frequencies. Launched from the same start point, the pendulums synchronize and desynchronize over time, resulting in momentary percepts of “order” (where the pendulums look like they are moving as a single wave) and of “chaos” (where the coherent global motion seems to disappear). We tracked attentional effort across continuously transitioning pendulum waves using a multiple object tracking task. Observers tracked “pendulums” (rendered as Gabor patches) in a pendulum wave animation that was transitioning from either order to chaos or chaos to order. Observers reported whether a probe appeared on/off a tracked pendulum. Because tracking performance decays over time, we presented sliding windows of the full transition and presented probes in the early, middle, or late parts of each window. We found that tracking performance decreased over time, but this was influenced by the progression and direction of the transition. In particular, the transition from order-to-chaos disrupted performance early in the

progression of the pendulum wave (compared to its chaos-to-order counterpart)—and chaos showed a stabilizing effect on performance throughout the rest of the transition. Thus, in perception, as in physics, the law of entropy seems to structure visual processing: while the loss of order can be disruptive, the resulting chaos may free up attentional resources, allowing for more flexible attention allocation.

Supported by a Natural Sciences and Engineering Research Foundation of Canada Discovery Grant

26.425 PERCEPTUAL SPACE FOR PLAIDS

Yi xu¹, Akihito Maruya², Qasim Zaidi¹; ¹SUNY college of optometry, ²Zuckerman Institute Columbia University

Plaids formed by the sums of two gratings have been used extensively in vision research, but the perceptual qualities of the large variety of possible plaids have not been studied. Many plaids are seen as the sum of two gratings, but others look like gratings with one or more phase shifts (orthogonal or oblique to the grating orientation), braided patterns, or as discontinuous or wavy gratings. Given that each grating is defined by four parameters (spatial frequency, orientation, phase and contrast), the generative space of plaids is continuous and 8-dimensional. We aim to use perceptual similarities to construct a more tractable space for plaids that would be useful for many purposes, using 37 plaids that covered all the categories above. The space could be topological consisting of discrete clusters, or hierarchically organized into a tree structure, or the similarities could be represented as distances in a multi-dimensional Euclidean, spherical or hyperbolic space. Four observers looked at each plaid as the reference surrounded by the 36 others and indicated the perceptual nearest neighbor (Tversky and Hutchinson, 1986). Observers' centrality parameters (2.45, 2.13, 2.34, 2.29) indicate a tree structure with around 4 branches, but with 37.4% concordance across observers. To test if a tree can be embedded into a space requires a complete similarity matrix. We used Waraich and Victor's (2024) method of ranking 8 similarities to efficiently estimate the matrix and compare the log likelihood of Euclidean model spaces (1-10 dimensions) with a random model as the lower bound and distances equal to empirical choice probabilities as the upper bound. The log-likelihood asymptoted at 4 dimensions and DBSCAN indicated a single cluster. A spherical space with minimal curvature provided a slightly better fit, but a hyperbolic space did not. The perceptual space of plaids is thus simpler than the generative space.

EY035085, EY035838

26.426 FEATURING MISBINDING DRIVEN BY COMMON HIGHER-LEVEL PERCEPTS RATHER THAN LOWER-LEVEL STIMULUS FEATURES

Sunny M Lee¹ (sunnylee@uchicago.edu), Steven K Shevell¹; ¹University of Chicago

Does feature binding to objects depend on higher-level object percepts? Percepts are mentally constructed representations that are

visually experienced, such as hues, motions, and three-dimensional objects, not physical stimulus properties they derive from. Similarity among physical stimulus features, however, typically is proposed to drive feature binding to objects. While binding by physical similarity is a plausible pre-percept mechanism, does higher-level perceptual similarity also contribute to feature binding? Aim: Distinguish between two levels of visual neural representation to determine if a higher-level perceptual representation can drive the conscious experience of color-motion feature misbinding (the incorrect perceptual pairing of a physical object color and physical motion direction for peripheral stimuli, induced by the physical color-motion pairing in the central stimulus). Method: Observers viewed dot kinematograms comprised of two sets of colored dots, one red and one green, and with dots moving vertically upwards or downwards. While fixating centrally, observers reported the perceived motion direction in the periphery for one set of colored dots. Each set of dots had one of two motion types: (1) Planar linear motion in which all dots moved at the same constant vertical speed or (2) Structure From Motion (SFM) that had dots accelerating/decelerating vertically to evoke a 3D percept of a rotating cylinder. The two sets of dots otherwise had identical feature values (size, density, luminance) and were matched in average speed. Results/Conclusion: Color-motion perceptual misbinding was induced in the periphery for the majority of the presentation time when both sets had the same motion type (either both planar or both SFM). When one set was planar and the other SFM, however, misbinding was significantly reduced despite the preponderance of matched stimulus features for all dots. This indicates that misbinding can be facilitated by similar higher-level perceptual representations of motion.

Supported by NIH EY-026618

26.427 NOVEL TAILORED NOISE ACCELERATES REVERSE CORRELATION EXPERIMENTS REVEALING DIFFERENCES IN OBJECT GROUPING CUES

Ethan Duwell¹, Adam Greenberg¹; ¹Medical College of Wisconsin

In psychophysical reverse correlation (PRC) experiments, subjects perform tasks on base images (BIs) with noise overlaid. After many trials, noise from correct and incorrect trials is computationally combined to generate classification images (CIs), revealing BI features critical for task performance. PRC's main limitation is efficiency, often requiring 5–10k trials for interpretable CIs. However, our prior simulations suggest PRC can be accelerated by optimizing the noise's spatial-frequency (SF) content to reflect the task BIs'. Here, we tested this in two object-grouping experiments during which subjects judged whether two rectangular objects were rotated 45° left or right. Objects were randomly occluded or un-occluded across trials. In the luminance-grouping (Lum) task, the objects differed in luminance from 0–255 around a gray (127.5) background. In the common-region (CR) task, grouping was determined by a thin rectangular boundary surrounding one of two equi-luminant objects. Boundary luminance varied 0–127.5 above/below gray. Each experiment contained 3750 trials using both white noise and noise

tailored to the BIs' SF spectra, in separate blocks. Eight subjects completed the Lum task; seven completed the CR task. As in our simulations, Lum and CR experiments using tailored noise produced interpretable CIs with features corresponding to underlying BI structures, even with analyses limited to <1k trials. Task-specific CIs yielded important differences: Lum-task CIs highlighted entire objects; whereas CR-task CIs preferentially highlighted occluded regions. Conversely, white-noise versions failed to produce interpretable CIs using all 3750 trials. Occlusion increased difficulty in all conditions. Furthermore, Lum-task psychometric function shifts indicate that tailored noise increased difficulty relative to white noise. However, several CR-task subjects showed the opposite pattern, likely reflecting greater high-SF overlap between CR cues and white noise. Overall, these results provide evidence that noise optimization in PRC can greatly improve efficiency while allowing for detailed interpretation of psychometrics and CIs.

This work was funded by NSF grant SBE 2122866 to A.G.

26.428 AN EXPERIMENTAL INVESTIGATION OF HOW STIMULUS SHAPE MODULATES PERCEPTUAL STABILITY IN THE SILHOUETTE ILLUSION

Kohei Ronden¹ (ronden.kohei.34e@st.kyoto-u.ac.jp), Kiyofumi Miyoshi¹, Shin'ya Nishida¹; ¹Kyoto University

In bistable perception, the interpretation of an ambiguous stimulus alternates spontaneously over time. Perceptual stability has been theoretically explained in terms of attractor “well depths” in an energy landscape, as well as model parameters such as input strength, adaptation, and noise. However, how perceptual stability is determined by physical stimulus properties remains unclear. Here, using the silhouette dancer illusion—where perceived rotation switches between clockwise (CW) and counterclockwise (CCW)—we explored how stimulus shape and representational format (silhouette vs. dot cloud) modulate perceptual stability. Forty-five participants viewed nine ambiguous rotating stimuli for three 60-second trials each and reported perceptual changes via keypress. All nine stimuli—five biological and four non-biological variants, each including one dot-cloud version—were constructed in Blender from a baseline silhouette reproducing the Spinning Dancer pose. Perceptual states were categorized as CW, CCW, Swing (an oscillatory motion between lateral endpoints), or other. Durations pooled across CW and CCW perception were estimated using a Gamma-family GLMM, and all stimulus pairs were compared with FDR correction. Compared to the baseline spinning-dancer stimulus, a symmetric pose and a cylinder-simplified dancer showed shorter dominance durations, whereas a inverted unfamiliar pose yielded longer durations. One non-biological stimulus (a chain of five connected cylinders) also showed significantly longer durations. A scrambled dot-cloud stimulus showed shorter durations than a structured dot-cloud dancer, while no significant difference emerged between the silhouette and dot-cloud dancer versions. PredNet next-frame prediction accuracies were computed for the seven silhouette stimuli, and a subset showed positive

correlations with dominance duration. The correlation suggests that shape complexity modulates predictability, which in turn may influence the depth of attractor wells and thereby regulate perceptual stability. These results suggest that structural complexity and familiarity, rather than silhouette format or biological relevance, drive differences in perceptual stability.

This work was supported by JSPS KAKENHI (JP24H00721).

26.429 FROM GIST TO DETAIL: ALPHA-BAND NETWORK SHIFTS ACROSS VISUAL SPATIAL FREQUENCIES

Michal Fux¹, Suayb Arslan², Abdul Rahman³, Pawan Sinha¹; ¹MIT, ²Boğaziçi University, ³Indian Institute of technology Madras

This study explores how the brain's communication patterns change when we look at images that vary in visual detail—specifically blurred, clear, and high-pass filtered pictures. We focused on activity in the brain's alpha band, a rhythmic signal linked to attention, coordination between brain regions, and the integration of visual information. By measuring how strongly different brain areas synchronized with one another, we examined how the brain adapts its processing strategies depending on how much coarse or fine detail an image contains. The results show a consistent pattern. When observers viewed blurred images, the brain displayed strong long-range communication, especially between frontal, parietal, and temporal regions. This suggests that when visual detail is limited, the brain recruits a broad network to fill in missing information and form an overall understanding of the scene. Clear images produced a more moderate level of connectivity, reflecting a balance between global interpretation and the straightforward processing of available detail. In contrast, sharp, high-detail images led to much weaker widespread communication. Instead, only a few localized connections—most notably between visual and prefrontal regions—increased, indicating a shift toward more detail-oriented, local processing. Overall, the study suggests that the brain flexibly changes its communication patterns depending on the type of visual information available. Blurred images encourage holistic, integrative processing; clear images produce a balanced response; and sharp images promote focused, detail-driven analysis. These findings offer insights into how the brain allocates attention and coordinates different regions to make sense of the visual world.

26.430 RECONSTRUCTING 2D IMAGES FROM BRAIN ACTIVITY EVOKED BY RANDOM-DOT STEREOGRAMS

Taiga Kurosawa¹ (kurosawa.taiga.66y@st.kyoto-u.ac.jp), Shuntaro C Aoki^{1,2}, Tanaka Misato^{1,2}, Yukiyasu Kamitani^{1,2}; ¹Kyoto University, ²Advanced Telecommunications Research Institute International

Random-dot stereograms (RDS) induce the perception of 3D structure through binocular disparity, despite containing no explicit shape information in the monocular 2D views. In this study, we investigate whether the neural representations generated by these stereoscopic inputs can be translated into 2D visual reconstructions. We employed a cross-modal decoding approach, first training a decoder on fMRI

signals collected while participants viewed natural 2D images. We then applied this decoder, trained solely on standard 2D visual features, to fMRI data recorded while participants viewed shapes defined exclusively by RDS disparity cues. We found that the decoder successfully reconstructed recognizable 2D images of the object shapes from the RDS-evoked brain activity. These reconstructions filtered out the high-frequency dot-pattern texture of the stimuli and instead visualized the holistic shape, exhibiting high similarity to reconstructions obtained from standard 2D luminance-defined stimuli. These results demonstrate that brain activity evoked by binocular depth cues can be decoded into 2D images, suggesting that the visual system transforms disparity information into a representation compatible with the neural coding of 2D pictorial features.

Supported by JSPS KAKENHI (JP25H00450, JP20H05705, JP20H05954), NEDO (JPNP20006), and JST CREST (JPMJCR22P3).

26.431 INVESTIGATING THE STABILITY OF THE ATTENTIONAL SPOTLIGHT BY MANIPULATING DISTRACTOR SALIENCE.

Joe Opdenaker¹, Chloe Paredes¹, Miranda Scolari¹; ¹Texas Tech University

Spatial cues determine the breadth of the attentional window, prompting observers to adopt either a narrowly focused or widely dispersed selection strategy. According to zoom-lens and gradient theories, this cue-driven adjustment shapes how visual information is prioritized and processed. Previous work shows that narrowly focused attention can make individuals susceptible to reorienting toward nearby ($<1^\circ$) salient distractors compared with diffuse deployment, calling into question the stability of an attentional spotlight. To test whether this reorienting is salience driven, we used a flanker task ($n = 25$) in which a focal or diffuse pre-stimulus cue preceded a target T and flanking distractor Ts that were either matched in contrast to the target (black or gray) or mismatched (high-contrast black targets with low-contrast gray distractors; or low-contrast gray targets with high-contrast black distractors). This design isolated the effect of salience on distractor suppression. Consistent with prior work, a significant interference effect was detected for uniform displays when a focal pre-cue was provided, but not following a diffuse cue. Furthermore, low-contrast distractors failed to interfere with high-contrast target processing, whereas high-contrast distractors generated strong interference effects when paired with a low-contrast target, regardless of cue condition. These results replicate prior work by demonstrating that attentional reorienting occurs in a crowded flanker task when attention is initially narrowly focused. This reorienting is likely driven by the relative salience of the distractors, rather than an inherent instability of the attentional spotlight.

Visual Memory: Objects, features

26.432 A REPRESENTATIONAL GEOMETRIC ACCOUNT OF WITHIN-CONCEPT VARIABILITY IN IMAGE MEMORABILITY

Hyewon Willow Han^{1,2}, Yalda Mohsenzadeh^{1,2}; ¹Western University, ²Vector Institute for Artificial Intelligence

What makes some images of the same object concept more memorable than others? Recent findings suggest that semantic features exert a stronger influence on the memorability of object images than perceptual features (Kramer et al., 2023). However, the factors driving variability in image memorability within the same object concept remain elusive. Here, we address this question through the lens of concept manifold geometry (Sorscher et al., 2022). Using a large-scale naturalistic object image collection with diverse concepts (Hebart et al., 2019), we first identify the representational space that best predicts memorability by linearly probing layers of deep neural network (DNN) vision models. We then characterize, for each object concept, the geometry of memorable and forgettable submanifolds within the overall concept manifold. Our analysis reveals that memorable exemplars occupy higher-dimensional submanifolds than forgettable exemplars, and that the distance between the submanifolds increases as the overall concept manifold has a larger radius but lower dimensionality. Furthermore, after accounting for between-concept variability, semantic and visual dimensions (Hebart et al., 2023; Contier et al., 2024), together with local manifold geometry, all make significant contributions to predicting image memorability. Lastly, by estimating within each concept the direction that best separates memorable from forgettable images, we show that these memorability directions are largely concept-specific and that variation in local dimensionality and radius along them shapes which images are remembered or forgotten. Together, these results suggest that image memorability within an object concept is governed by concept-specific manifold geometry rather than by a single global set of features.

26.433 CONTRIBUTION OF INTRINSIC MEMORABILITY AND SELF-RELEVANCE TO FACE MEMORY

Hidekazu Nagamura¹, Wilma A. Bainbridge², Shunichi Kasahara^{1,3}; ¹Sony Computer Science Laboratories, Inc., ²University of Chicago, ³Okinawa Institute of Science and Technology Graduate University

Understanding the factors that determine face memory is crucial for elucidating the encoding mechanisms of face perception. Previous research suggests that face memory is driven by both "intrinsic memorability," which is consistent across observers, and "self-relevance," which depends on observer attributes such as ethnicity (e.g., the other race effect; Meissner & Brigham, 2001). However, the relative contributions of these factors remain unclear, leaving optimal intervention strategies undefined regarding whether to prioritize

universally memorable features or to personalize stimuli for individuals. To answer this question, we quantitatively compared the contributions of intrinsic memorability and self-relevance. We generated a balanced face dataset (160 images: 4 ethnicities × 2 genders × 20 images) based on features estimated by the DeepFace neural network model to minimize confounds. We then measured memory performance using a continuous recognition task with a demographically balanced participant group (N = 120: 4 ethnicities × 2 genders × 15 participants). Participants viewed a continuous stream of faces (750 ms duration) and identified repeated images appearing after an average interval of approximately 1 minute. Results showed that while gender congruence between participant and image significantly improved memory performance, ethnicity congruence did not. To quantify the contribution of each factor, we employed a Bayesian mixed-effects logistic regression analysis. This approach allowed us to decompose the variance in memory performance into image-specific effects (intrinsic memorability), participant-image interaction effects (self-relevance), and participant-specific effects (individual memory ability). The impact of intrinsic memorability on memory performance was approximately three times greater than that of self-relevance and comparable to individual differences in participants' memory ability. These findings demonstrate that intrinsic image features are the dominant factor in face memory, outweighing the contribution of self-relevant demographic cues. This suggests that leveraging intrinsic visual information offers a more potent approach to memory augmentation than relying solely on demographic matching.

This work was supported by JST, PRESTO Grant Number JPMJPR2314.

26.434 SEMANTIC RELATIONSHIPS FACILITATE VISUAL WORKING MEMORY FOR OBJECT IDENTITY BUT MIGHT NOT ENHANCE THE DETAILS.

NITHIT SINGTOKUM^{1,2} (nithitsingtokum136@gmail.com), Chaipat Chunharas^{2,3,4}; ¹Department of physiology, Faculty of medicine, Chulalongkorn University, Bangkok, Thailand, ²Cognitive Clinical and Computational Neuroscience Center of Excellence, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand, ³Division of Neurology, Department of Medicine, Faculty of Medicine, King Chulalongkorn Memorial Hospital, Thai Red Cross Society, Bangkok, Thailand, ⁴Chula Neuroscience Center, King Chulalongkorn Memorial Hospital, Thai Red Cross Society, Bangkok, Thailand

Visual working memory (VWM) capacity is limited due to the constrained cognitive resources. To overcome these limits, VWM can leverage statistical regularities and compress similar items together. Studies of real-world objects have demonstrated that semantically related items improve memory capacity compared with unrelated items. However, it remains unclear whether it improves memory for object identity alone or also for fine-grained object details. Here, we investigated how resource availability modulates the benefit of semantic relatedness in VWM, and at which level of memory precision this benefit occurs. Participants performed a VWM task in which we

manipulated VWM load using set size (2, 4, 6 items) and encoding time (150, 500, 1000 ms) manipulation, along with changing the semantic relatedness between objects. Memory precision was assessed by requiring participants to identify the studied item shown in different states against two states of the foil object. We found that when objects are semantically related, participants can remember them better than when they have no semantic relationship ($F(1,29)=104.22$, $p<.001$). But this advantage emerged only when resources were depleted. We observed an interaction between the set size and semantic relatedness ($F(1.66,48.09)= 36.07$, $p< 0.001$), but no interaction between encoding time and semantic relatedness ($F(1.96,56.96)=0.27$, $p= 0.759$). However, we didn't observe the improvement in object state precision on trials where participants correctly identified the object. ($F(1, 27)= 0.50$, $p = 0.484$) In summary, semantic relationships improve VWM performance for object identity, particularly when resources are limited. And, these findings might correspond with the underlying mechanism observed in low-level feature chunking, where surpassing capacity limits results in diminished detail for individual representations. However, absence of an effect on state precision might be because of the 4-AFC response, which could not clearly define effect on object identity and state separately. Our following work will examine these aspects independently.

26.435 THE EFFECT OF CELEBRITY STATUS ON FACE MEMORABILITY

Esther Goldberg¹, Cambria Revsine², Wilma Bainbridge²; ¹University of California, Los Angeles, ²University of Chicago

Celebrities are known around the world because of their prevalence in various media platforms and advertisements. For that reason, it is common for celebrity faces to be used as stimuli in visual memory research since fame can represent a shared form of familiarity. However, it is possible that celebrity faces as a stimulus set include perceptual biases that may impact memory performance. For example, celebrities could fundamentally differ in the intrinsic memorability of their faces, separate from their repeated cultural exposure. In this study, we assess the memorability of famous people independent of their level of familiarity by testing memory for celebrity faces cross-culturally. We selected participants and stimuli from different cultural backgrounds; Indian participants ($n = 266$) viewed Brazilian celebrity and non-celebrity faces, while Brazilian participants ($n = 236$) viewed Indian celebrity and non-celebrity faces during a continuous recognition task. Across both cultures, celebrity faces were remembered significantly less accurately than non-celebrity faces, contradicting the intuition that celebrities have inherently more memorable faces. A separate attribute rating study revealed that, beyond celebrity status, only perceived confidence predicted memorability, where more confident-looking faces were remembered less often. Other attributes, including attractiveness and distinctiveness, were significantly different between celebrities and non-celebrities, but were not predictive of memorability. These findings

suggest that celebrities and non-celebrities differ significantly in their intrinsic memorability, and that memory effects that often have been attributed solely to familiarity may instead come from underlying differences in the memorability of the stimuli. This work challenges the assumption that celebrity faces can serve as neutral "familiar" stimuli in cognitive research and emphasizes the need to control for intrinsic memorability when selecting stimuli for memory studies.

26.436 TOPOLOGICAL VISUAL REPRESENTATION REVEALED THROUGH DRAWING

MAHWISH KITTUR¹, Sami Yousif²; ¹University of North Carolina at Chapel Hill, ²Ohio State University

From the gist of a narrative to the details of the visual scene, our memories are not veridical. Yet, the ways in which we misremember things are not random either. Rather, they can reveal something meaningful about the content that is represented in the first place: Essential information tends to be preserved, and incidental information tends to be lost. Careful examination of these visual memory errors can therefore reveal the underlying format of a mental representation. Here, we exploit this fact. By having observers remember and draw simple letter-like objects, we 'reverse engineer' the building blocks underlying their representation. Consistent with recent work, we find evidence of topological representation. In a first experiment, using strategically designed line drawings, we investigated how angular information is distorted. We found an unambiguous bias to misremember angles as closer to 90 degrees. We found a similar pattern for line lengths: All line lengths tended to be biased towards the average value. Despite these distortions, the overall topological form – consisting of the "T-junctions" and "holes" that composed the shapes – was remarkably well preserved. In fact, holes were virtually never lost, and we found that topologically relevant junctions (e.g., T-junctions) were maintained at a higher rate than visually similar but topologically irrelevant junctions (e.g., L-junctions). In other words, while the metric properties of the line drawings were robustly washed away, the overall topological structure was preserved in memory. To investigate the stability of these findings further, we repeated the angle and length experiments with children. In a sample of five to eight year olds, we found similar distortions and memory errors as we found in adults. These results provide a direct window into the shape of visuospatial representations, revealing the topological primitives underlying visual representation throughout the lifespan.

26.437 PROBING THE STRUCTURE OF WORKING MEMORY REPRESENTATIONS IN VISION AND AUDITION

Abigail Noyce¹, Raina Alam¹, Will (Yuhang) Li², Eli Bulger¹, Stephanie Bronfman^{1,3}, Kammiee Ardo¹, Barbara Shinn-Cunningham¹; ¹Carnegie Mellon University, ²University of Wisconsin, ³University of Pittsburgh

Working memory (WM) is a capacity-limited resource. Grouping simultaneously-presented features into coherent objects provides a scaffolding that supports WM efficacy, but less is known about how to-be-remembered items are grouped together over time. In hearing, the

organization of sound information into auditory objects is intrinsically temporal, dramatically impacting WM storage and retrieval. To ask whether the same is true for vision, we manipulated temporal grouping and WM retrieval demands in visual and auditory WM in separate experiments. In visual WM, the memoranda were a series of locations occupied by a stimulus disc. In the temporally grouped condition, the disc moved smoothly between locations; in the temporally ungrouped condition, it jumped discontinuously. In auditory WM, the memoranda were a series of sound stimuli. In the temporally grouped condition, these sounds were complex tones with identical timbre; in the temporally ungrouped condition, these were snippets of diverse environmental sounds. WM retrieval of individual items was probed via Sternberg-type old/new judgment of a single item; WM retrieval of the temporally-grouped object was probed via sequence change detection. Our preliminary data suggest that temporal grouping works differently in the two senses. Visual WM performance was similar regardless of temporal grouping; auditory WM performance replicated our prior result. WM capacity thus depends not only on stimulus features and task demands, but also on fundamental characteristics of the underlying perceptual process, which differ across sensory modalities.

Supported in part by ONR MURI N00014-19-12332; NIH R25 DC020922-02; CMU HURAY.

26.438 CROSS-MODAL RECRUITMENT OF SENSORIMOTOR CORTICES DURING A PURELY VISUAL WORKING MEMORY TASK

Leah J Ettensohn¹ (lettensohn@ucsd.edu), Sunyoung Park^{1,2}, John T Serences^{1,3}; ¹UCSD Psychology, ²Rutgers Center for Advanced Human Brain Imaging Research, ³UCSD Neurosciences Graduate Program

Working memory (WM) involves the temporary storage and manipulation of information across sensory domains. Research on unimodal haptic and visual WM shows that detailed mnemonic representations are sustained in sensory-encoding areas over short delays. Here, we used fMRI to examine object-selective response patterns in visual and sensorimotor cortex to test whether haptic representations are recruited during a purely visual WM task, particularly when observers are familiarized with the haptic properties of the objects. Object-selective responses during a visual delayed match-to-sample task were collected before and after 3 days of behavioral training. Objects were assigned to three sets: a visual training set without haptic feedback, a haptic training set without visual feedback, and an untrained control set. To control for modality and amount of experience, we used novel asymmetric greebles (Sheinberg & Tarr, 2009). Behavioral data from the purely visual WM task during the fMRI sessions showed improvements before and after training across all conditions, but we observed significantly greater gains after visual and haptic training compared to the untrained control set. Next, using multivariate analysis we found that visual object identity was reliably decoded in both visual and sensorimotor cortex. Comparing

gains in decoding accuracy before and after behavioral training, we did not observe significant gains in early visual cortex and lower-level object selective regions. However, visual and haptic training did produce significant gains in higher-level object and scene-selective regions of the ventral visual cortex, as well as regions of parietal and sensorimotor cortex. Collectively, these results suggest that experience modulates object representations in higher visual and sensorimotor cortex, and that haptic representations can be recruited cross-modally to support a purely visual WM task.

National Eye Institute R01EY025872

26.439 IRRELEVANCE-INDUCED AGNOSIA: REPORT FAILURES FOR SALIENT FEATURES OF VISIBLE, MEANINGFUL OBJECTS

Ariel Berlinger¹, Ariel Goldstein², Baruch Eitam³; ¹University of Haifa, ²The Hebrew University, ³University of Haifa, University of British Columbia

Much of the debate over sparse vs. rich accounts of consciousness hinges on whether report failures reflect limitations of conscious experience itself or of post-perceptual access mechanisms. Failures to report may reflect limits of experience, limits of access, or both. We adapt an experimental paradigm that pushes the sparse position by demonstrating systematic report failures for clearly visible, meaningful stimuli, under no capacity constraints (minimal set-size and memory load) while accounting for mind wandering and surprise. In two online experiments ($n=60$ and $n=47$), participants viewed flower-in-a-pot stimuli for 1,000ms. They reported one object's color (flower or pot) until a surprise probe tested knowledge of the irrelevant object's color. In the second experiment, the stimuli rotated and appeared at different locations, minimizing the effects of focal spatial attention and mind wandering. As expected, the relevant color was reported near-perfectly (Exp1: 99%; Exp2: 96%), indicating engagement and comprehension. Following repeated reporting of the task-relevant object's color, around mid-experiment, an unexpected probe was presented to quantify the effect of surprise. Surprise reduced the subsequent success in reporting the relevant object's color (Exp1: 67%, CI [55,79]; Exp2: 55% CI[41,70]). Critically, when unexpectedly probed for the irrelevant object's color, performance was substantially lower, reaching chance level (Exp1: 20% CI[10,30]; Exp2: 23% CI[11,36]; chance=25% 4-alternative-forced-choice). Reporting of the relevant color's object following the probe of the irrelevant object's colour, showed only the effect of surprise (Exp1: 67% CI[55,79]; Exp2: 74% CI[62, 87]). In consequent trials, after both colors were deemed relevant, performance rebounded (Exp1: 85% CI[76, 94]; Exp2: 73% CI[64,82]). We term this pattern Irrelevance-Induced Agnosia; task irrelevance alone can drive striking report failures for salient stimulus features beyond the effect of surprise, spatial attention, attentional disengagement, and arguably, conscious perception itself. These results constrain the use of report as a proxy for conscious perception.

This research was supported by The Israel Science Foundation (ISF) grant number 1714/23 and The Bi-national Science Foundation (BSF) grant number 2022/321.

26.440 SPATIAL ATTENTION'S ROLE IN ACTIVE MEMORY MAINTENANCE ACROSS SEQUENCES DEPENDS ON WHETHER SPACE IS THE REMEMBERED FEATURE

Sage Bendickson¹, Nicholas Schmitz, Ilke Kavusturan, Alexis McCraw, David Sutterer; ¹University of Tennessee Knoxville

Space and time are thought to play a priority role in working memory (WM) formation and maintenance, due to their role in feature binding and tendency for incidental encoding in memory. However, the extent that space and time operate complementarily or trade priority depending on task relevance remains an open question. Here we investigated if the relevance of spatial information changes the deployment of spatial attention, and its contribution to task-specific memory maintenance, when items are separated in space and time. We report two experiments that differed in whether location or color was the tested feature. In both experiments, we measured EEG alpha-band power while observers memorized the target feature of two sequentially presented colored dots. We then applied an inverted encoding model (IEM) to the topography of oscillatory alpha-band power (8 – 12 Hz) on the scalp to measure how long the location of each item was attended. In Experiment 1, observers performed a delayed color estimation task, and we found that patterns of alpha-band power only tracked the location of the most recently presented item. In Experiment 2, observers performed a delayed spatial estimation task, and we found that patterns of alpha-band power tracked each item's location and that the spatial representation of item 1 persisted significantly longer than in Experiment 1. Together, our results show that spatial attention only maintains the most recent item's location in a location-irrelevant task, but that multiple sequential locations are actively attended when location is the tested feature.

26.441 THE INTERPLAY BETWEEN REPRESENTATIONAL MOMENTUM AND REMEMBERED SIZE

Dominique Lopiccolo¹ (dlopicco@udel.edu), Alon Hafri¹; ¹University of Delaware

Representational momentum (RM) is a bias in visual memory whereby moving or changing objects are remembered as farther along their trajectory than they actually were. Most investigations of RM have involved rigid spatial transformations (e.g., rotation), but less is known about RM for nonrigid transformations. Finding RM in this context would expand our understanding of the extent and limits of anticipated change that are encoded in memory. In earlier work, we examined RM for spatial scaling, a nonrigid change in which an object resizes while maintaining its shape. Dynamic displays of expanding or contracting circles elicited RM, such that the circles were remembered as larger or smaller than their final size. Yet a striking asymmetry emerged: RM appeared consistently greater for shrinking than for growing. Here we

asked whether this asymmetry reflects an interaction with an independent “smaller bias” previously observed in visual memory, whereby static objects are remembered as smaller than they appeared. If such a bias exists, it would align with shrinking events, increasing the propensity to report the later (smaller) frame, but oppose growing events, decreasing that propensity. To test this, participants viewed two trial types: in dynamic trials, circles grew or shrank before being masked; in static trials, they remained stationary. Participants then completed a two-alternative forced-choice task, selecting which of two probes (one smaller, one larger) matched the last image they saw. We again observed a robust RM effect and reproduced the shrink-grow asymmetry. Crucially, static trials also revealed a reliable smaller bias, and after accounting for this bias, RM magnitude was equal across scaling directions. Together, these results show that the apparent asymmetry between growing and shrinking RM arises from the interaction of two distinct memory biases. More broadly, they reveal how multiple mechanisms jointly shape visual memory, even for simple geometric transformations.

Color, Light and Materials: Material perception

**SATURDAY, MAY 16, 2:45 – 6:45 PM,
PAVILION**

26.442 SEMANTIC DISTANCE PREDICTS ABILITY TO INTERPRET TEXTURE MEANING IN INFORMATION VISUALIZATIONS

Anna Chinni¹, Kushin Mukherjee², Karen B. Schloss¹; ¹University of Wisconsin - Madison, ²Stanford University

A central goal in research on information visualization is to identify generalizable principles and metrics that predict people's ability to interpret visualizations, transcending any one feature or chart type. Previous work within the domain of color semantics developed semantic distance as a metric to predict performance on assignment inference, the process by which people infer the mappings between perceptual features and concepts in visualizations. Semantic distance is computed from the probability of one assignment over the alternative(s). For color, semantic distance predicted accuracy in assignment inference, independent of perceptual distance and color-concept association strength (Schloss et al., 2021). In this study, we tested the generalizability of semantic distance as a predictor of interpretability by studying assignment inference for a different type of perceptual feature: visual texture. During the experiment, participants viewed a series of bar graphs, in which the textures of the bars represented different concepts (e.g., bubbly and stripey textures representing the concepts ocean and field). On each trial, participants were given a target concept (e.g., ocean) and they reported which textured bar mapped onto that concept. Across trials, the texture sets spanned a wide range of semantic distances and varied systematically

with respect to the concepts they represented. We found that semantic distance correlated strongly with accuracy in assignment inference (e.g., $r = .88$, $p < .001$ for the concepts ocean and field). A logistic mixed effects model showed that semantic distance predicted accuracy independent of perceptual distance (estimated by the Euclidean distance between textures in an embedding space constructed from triadic similarity judgements (Chinni et al., VSS-2025)) or association strength for the optimal texture choice. These results support the possibility that semantic distance is a generalizable metric for predicting performance in assignment inference, which can be used broadly to design visualizations that facilitate communication.

This work is supported in part by the National Science Foundation Grant BCS-2419493

26.443 HIGHLY REPRESENTATIONAL TEXTURES INTERFERE WITH GLOBAL ASSIGNMENT FOR INTERPRETING INFORMATION VISUALIZATIONS

Zoe S. Howard¹ (zoeshoward@gmail.com), Anna Chinni¹, Karen B. Schloss¹; ¹University of Wisconsin-Madison

To create easily interpretable information visualizations, it is useful to understand people's expectations of how visual features (e.g., texture, color) map onto concepts. Prior work on color showed that observers use assignment inference to infer color-concept mappings that maximize the "goodness" of all color-concept pairs in the scope of the encoding system (global assignment), rather than mapping colors to their strongest associate (local assignment). Global assignment led participants to assign concepts to weakly associated colors, despite more strongly associated candidates (Schloss et al., 2018). Studies of assignment inference for colors focused only on the colors and concepts in the scope of the encoding system. However, this scope may be too narrow when considering perceptual features that can be highly representational, such as textures (e.g., brick textures specifically resemble bricks). When textures strongly evoke concepts outside of the encoding system, observers may be less open to assigning textures to weakly associated concepts, resulting in local over global assignment when the two conflict. To test this possibility, we had participants interpret bar charts in which two textured bars represented two concepts ("encoding system"). On each trial, they reported which bar corresponded to a "target" concept (non-prompted concepts were "non-targets"). We focused on trials where global vs. local assignment would produce opposite responses. Overall, assignments were no more likely to be global than local, but the probability of global assignment increased as globally optimal textures were less representational of concepts outside of the encoding system ($p < .01$), more associated with the target ($p < .001$), and less associated with the non-target ($p < .001$). These results expand our understanding of factors that contribute to assignment inference—when visual features are highly representational, it is necessary to account for concepts outside of the encoding system to predict how people ascribe meaning to visual features in information visualizations.

NSF award BCS-2419493 to K.B.S

26.444 USING IMAGE-BASED METRICS TO PREDICT GLOSS PERCEPTION ACROSS DIFFERENT SCENES

Zoe R. Goll^{1,2} (zoe.goll@psychol.uni-giessen.de), Jacob R. Cheeseman^{1,2}, Roland W. Fleming^{1,2}; ¹Justus Liebig University Giessen, Germany, ²Center for Mind, Brain and Behavior, Universities of Marburg, Giessen, and Darmstadt

Material appearance depends strongly on distal scene factors like viewpoint, illumination, and object shape, causing discriminability between two surfaces to vary dramatically across scenes. The same surface reflectance can appear very different under certain viewing conditions, while two different materials may look perceptually equivalent under other conditions. Here, we investigate whether perceived equivalence between materials across such scene changes can be predicted using image-based metrics. We created a set of images that systematically varied viewpoint, illumination and object shape, rendering each scene with five levels of surface roughness. Within scenes, we calculated the predicted perceptual differences between neighbouring roughness levels using a visual difference predictor (VDP, Mantiuk et al., 2023). We averaged VDP values across each scene and selected three "high-VDP" scenes with the largest differences and three "low-VDP" scenes with the smallest. In this way, VDP was used to select scenes for which we expect high sensitivity to changes in surface roughness, when other scene factors are held constant. Here we are interested in whether those VDP values also predict asymmetric gloss judgements across varying scene conditions. Observers were asked to choose the object with the rougher material in a 2AFC task. A scene with a near-median VDP value was chosen as the comparison stimulus. Psychometric functions were estimated for the three "high-VDP" and three "low-VDP" scenes, and a control with symmetric scene parameters. We find that VDP well predicts biases in gloss perception across different scenes, even though it is based on measuring visibility of differences within particular scenes. Matches for "low-VDP" scenes were rougher than "high-VDP" scenes and exhibited greater variability across participants. These results suggest that image metrics can be a useful tool for estimating bias and sensitivity not just within scenes, providing a possible first step towards predicting material discriminability across scenes.

26.445 LAYER-WISE REPRESENTATION OF GLOSSINESS FEATURES IN THE DEEP NEURAL NETWORK FOR OBJECT RECOGNITION.

Hiroaki Kiyokawa¹ (hkiyokawa@mail.saitama-u.ac.jp), Kai Yoshida¹, Ichiro Kuriki¹; ¹Saitama University

Glossiness is one of the remarkable features in object surfaces that helps identifying the object material. However, the key image feature for the glossiness perception is still under debate. This study addressed this issue by utilizing hierarchical structure of a DNN model VGG19 (Simonyan & Zisserman, 2014) to elucidate the image features necessary for the glossiness. VGG19 is known for high similarity in the

hierarchical structure of object representations in human brains (Nonaka et al., 2020). To visualize image features in each layer, we employed style-transfer technique (Gatys et al., 2015) to transplant image features of glossy surface onto a non-glossy one, while manipulating which layers to transfer image features. Glossy object images were rendered using actual BRDFs (EPFL database) of metallic and non-metallic objects. Five layers after the pooling process in VGG19 model were manipulated under two experimental conditions: transferring features from only one layer (one-layer condition) and transferring features except using one of the five layers (one-layer exclude condition). Human observers rated glossiness of the style transferred images on a 9-point scale. As a result, perceived glossiness across all materials were the highest when image features in the third layer were used under the one-layer condition. The glossiness scores were also significantly higher for metallic materials when the fourth layer was used, and non-metallic materials was so when the second layer was used. In contrast, no significant difference was observed when only one layer was excluded. These results imply that glossiness representation is spread widely across layers. Additionally, considering that the shallower DNN layers represent more spatially localized information than deeper layers, the need of deeper layer for metallic objects and shallower layer for non-metallic objects in the style transfer suggest the importance of spatially distributed features to render glossiness in metallic surfaces than non-metallic ones.

26.446 MATERIAL PERCEPTUAL SPACE FOR TRANSLUCENT OBJECTS: A COMPARISON BETWEEN AUSTRALIAN AND JAPANESE OBSERVERS

Takehiro Nagai¹, Ruka Watanabe², Juno Kim³; ¹Institute of Science Tokyo, ²Institute of Science Tokyo, ³University of New South Wales

In English, translucency denotes a material property characterized by subsurface light scattering, as observed in wax. In contrast, Japanese does not have a single, commonly used lexical item that maps directly onto this technical concept. Such linguistic and cultural differences may influence the strategies used in everyday visual material judgments, and these influences may manifest as cross-cultural differences in the perceptual impressions derived from material appearance. The present study aimed to examine whether the characteristics of material appearance judgments for translucent objects differ across these two cultural groups. We tested 10 Australian and 10 Japanese participants. Stimuli consisted of 64 computer-generated images of translucent objects that varied in shape, surface roughness, scattering coefficient, albedo, and other parameters. In each trial, participants ranked four test stimuli with respect to their perceived similarity to a designated reference stimulus. The resulting rank-order data were treated as repeated paired comparisons and analyzed using nonmetric multidimensional scaling (NMDS) to reconstruct the perceptual space defined by the stimuli. Differences between cultural groups were quantified using Procrustes-based disparity measures. Although the overall structure of the

perceptual spaces was broadly similar between Australians and Japanese, a significant divergence emerged specifically in the perceptual organization of objects containing specular highlights. To account for trial-wise similarity responses, we further constructed models based on simple image features—mean luminance and contrast—as well as global object shape. For Australian observers, low-level visual features alone were insufficient; shape information contributed substantially to explaining their responses. In contrast, Japanese observers' judgments were more readily explained by simple image features without requiring shape-based predictors. These findings suggest that cultural factors, including linguistic differences, may lead to distinct strategies in material judgment, resulting in measurable variations in the perceptual organization of material appearance.

JSPS KAKENHI 21KK0203, 23H03484, 25K22816

26.447 ROLE OF EXTERNAL CONTOUR IN MATERIAL-INVARIANT SEPARATION OF DEFORMATION FROM DYNAMIC POSTURE CHANGE

Yung-Hao Yang¹, Shin'ya Nishida¹; ¹Cognitive Informatics Lab, Graduate School of Informatics, Kyoto University

To detect 3D non-rigid deformations of moving objects, the visual system must separate image motion caused by shape change from that caused by posture change. Although changing an object's material alters its internal shading pattern and flow, recent work (van Zuijlen et al., 2025 JOV) suggests that deformation detection remains stable across diverse optical materials. Whether this robustness relies on material-dependent motion cues (internal shading flow) or material-independent cues (e.g., external contour dynamics) remains debated. In Experiment 1, we measured deformation-detection thresholds for a rotating knot object rendered with four materials (dotted matte, glossy, mirror, and transparent), and with a silhouette version lacking internal flow. Silhouette performance matched that of the first three material conditions when a simple material-invariant cue (e.g., object height) was available in a side-viewing condition, but significantly worse than the three when this cue was reduced in a diagonal-viewing condition. This suggests that internal flow cues can help disentangle deformation from posture change. To see which internal flow cues are used, Experiment 2 measured deformation-detection performance for four original materials, iso-luminance contours (simulating internal flow of the material stimuli without shading), rims (contour outlines), and silhouettes. Only the diagonal viewing condition was used. Removing shading had only minor effects, and silhouette and rim thresholds were almost as low as those in the three material conditions. Experiment 2 indicates no effect of internal flows, unlike Experiment 1. A possible interpretation of this dissociation is that observers in the latter experiment may have learned to use potentially ambiguous external contour cues more effectively during extensive viewing of diverse versions of the same stimulus set. Overall, the results we obtained so far suggest that deformation perception may rely on internal flow cues,

but their contributions are evident only when external contour cues are less reliable.

This work has been supported by JSPS Kakenhi JP20H05957.

26.448 THE IMPACT OF TEMPORAL ATTITUDES ON PHYSICAL AND AFFECTIVE MATERIAL IMPRESSIONS

Kenta nakajima¹, Takehiro Nagai; ¹Institute of Science Tokyo

It remains unclear whether there are any differences in the stages of visual processing underlying impressions of different material attributes, such as glossiness or aesthetics. This study investigated how “temporal attitude”, the participants’ expectation of stimulus duration, modulates the neural dynamics of material perception, thereby revealing the underlying processing stages. Participants viewed 320 object images and rated either glossiness or aesthetics in separate blocks. Temporal attitude was manipulated between blocks by varying the proportion of short (100-ms) and long (1000-ms) trials. Each block consisted of either long trials only, or a mixture of long and short trials presented in equal proportion. EEG was recorded while participants performed the task, and decoding analyses based on linear support vector machine were conducted using sliding time windows on the raw EEG signals. The decoding results indicated qualitatively different patterns across attributes and temporal-attitude conditions. For glossiness, decoding accuracy tended to be higher when short presentations were mixed and participants expected shorter presentation durations. For aesthetics, decoding accuracy tended to be higher when only long presentations were included and participants expected longer viewing durations. A statistically significant interaction between attribute and temporal attitude was observed in the 200–300 ms time window, indicating differential impacts of temporal attitude on glossiness-related and aesthetics-related EEG signals. Overall, these findings suggest that implicit expectations about viewing duration alter the temporal dynamics of neural representations in attribute-specific ways. While the effects are modest, the present approach can be a useful framework for investigating temporal aspects of neural processing for different material impressions.

This work was supported by JST SPRING, Japan Grant Number JPMJSP2180 to Kenta Nakajima.

26.449 EFFECTS OF LIGHTING DISTRIBUTION ON IMPRESSIONS OF A HANDCRAFTED OBJECT: A COMPARISON OF VR SIMULATION QUALITY AND REAL-SPACE CONDITIONS

Yoko Mizokami¹, Keito Sato¹, Hiromi Sato¹; ¹Chiba University

In exhibitions of craftworks, faithfully conveying the subtle texture of an object is crucial to the viewing experience, making texture reproduction an important aspect of lighting design. In previous studies, we suggested that light distribution conditions influence impressions, such as the perceived weightiness of a traditional Japanese handcrafted pottery piece, a tea bowl. We also

demonstrated that although the impression of a tea bowl changes under different lighting conditions in Virtual Reality (VR) environments, the nature of this change differs from that in real space due to limitations in rendering accuracy. In this study, we used stimuli with varying levels of rendering quality to examine whether the impression gap between the two environments could be reduced. In the real-space experiment, tea bowls were illuminated under nine lighting conditions, combining three beam angles (8°, 16°, and 29°) with three lighting directions: 30° from the front, directly overhead, and 30° from the rear. In the VR experiment, we recreated similar lighting environments using Unity and presented them via an Oculus Quest 2 head-mounted display. Observers rated their impressions of the tea bowls under each lighting condition on a seven-point scale, using 12 adjective pairs: Glossiness, Brightness, Blackness, Smoothness, Depth, Softness, Lightness (or weightiness), Naturalness, Preference, Beauty, Flamboyance, and Luxury. As a result, higher-quality VR produced impressions more closely resembling those in real space. Impressions such as weightiness and flamboyance varied depending on the angle of the lighting. We attribute this to enhanced rendering accuracy in the VR environment, which improved the representation of gloss. Furthermore, luminance statistics such as skewness and kurtosis were found to influence impression formation in both environments. Taken together, these results provide insights that are helpful in controlling the impression of craftworks through light distribution and in improving texture representation.

This work was supported by JSPS KAKENHI JP19H04196, JP24K03024, and JST CRONOS JPMJCS25K6

26.450 OPTICAL MATERIAL DISCRIMINATION IN AUTISM SPECTRUM DISORDER

Masataka Sawayama^{1,2} (masa.sawayama@gmail.com), Takashi Itahashi³, Yoshinori Dobashi^{1,2}, Haruhisa Ohta³; ¹Graduate School of Information Science and Technology, Hokkaido University, Japan, ²Prometech CG Research, Japan, ³Medical Institute of Developmental Disabilities Research, Showa Medical University, Japan

Individuals with autism spectrum disorder (ASD) differ from typically developing (TD) individuals not only in social interaction and communication but also in atypical sensory responses, such as hypo-responsiveness to pain or hypersensitivity to sensory signals. These sensory characteristics suggest that the visual appearance in daily environments may differ substantially between ASD and TD individuals. However, previous studies on sensory and perceptual processing have focused mainly on early visual processing, such as contrast sensitivity or color discrimination. To clarify the characteristics of perceptual experience in ASD under more natural environments, the present study focused on material perception and compared optical material discrimination between ASD and TD individuals. Using physically based computer graphics rendering, we generated object images in which six optical material dimensions were systematically manipulated: gloss contrast, gloss distinctness-of-image, translucency, metallicity, transmittance, and highlight inconsistency

(cf. Sawayama et al., 2022). Discrimination sensitivity for each dimension was measured in ASD and TD groups using an odd-one-out task. Results showed that at the group level, no significant differences were observed between ASD and TD participants for any of the material dimensions. We further analyzed representational dissimilarity matrices (RDMs) across the six tasks within each group. In the TD group, the gloss contrast dimension showed relatively high similarity with the metallic task but low similarity with the translucency-related task. In contrast, the ASD group showed weaker similarity between the gloss contrast and metallic tasks, but stronger similarity between the gloss and translucency tasks compared to the TD group. These findings suggest that although the overall ability to discriminate material properties does not differ substantially between ASD and TD individuals, the underlying material representations that support discrimination vary considerably between the groups. This indicates that ASD individuals may structure and relate material attributes in a perceptual space distinct from that of TD individuals.

Color, Light and Materials: Affect, cognition

SATURDAY, MAY 16, 2:45 – 6:45 PM,
PAVILION

26.451 NEURAL REPRESENTATION OF PERCEPTUAL REALNESS REVEALED BY FMRI AND MEG

Taiki Orima^{1,2} (apple.lemon.melon000@gmail.com), Hiroshi Ban^{1,3};
¹Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology (NICT), ²Japan Society for the Promotion of Science, ³Department of Frontier Biosciences, the University of Osaka

The human visual system is remarkably sensitive to the realness of images: we can correctly distinguish real photographs from paintings or computer-generated images even if they look realistic. This ability is vital for us to make correct judgments in our daily lives. However, the neural representation of realness remains unclear. Here, to investigate where and when in the brain the realness of images is processed, we conducted fMRI and MEG experiments in which participants (16/13 for fMRI/MEG) viewed real and unreal images (6,900 each). Decoding analysis showed that the real/unreal condition could be significantly classified from all the pre-defined regions of interest (early, ventral, and dorsal visual areas). However, the decoding results from higher dorsal visual areas could not be fully accounted for by image features (image statistics and style features) that differentiate real from unreal images. Additional decoding analysis from time-resolved MEG source signals revealed that decoding accuracy peaked at around 150 ms after stimulus onset in early and ventral visual areas. Interestingly, decoding accuracy in dorsal visual areas peaked at around 250 ms. These results suggest that the unknown aspect of realness, which is implicitly coded in the human

brain, can be processed in higher dorsal visual areas. Next, to further investigate the neural and perceptual correlates of realness, we conducted an additional fMRI experiment using a parametric series of morphed images designed to modulate realness. Specifically, real and unreal images were morphed at 0/100, 10/90, ..., 100/0 ratios. We recorded fMRI data from 9 observers while they viewed those morphed images. Subsequent decoding analysis showed that BOLD signals from higher dorsal visual areas notably classified perceived realness, suggesting that these areas specifically process realness. Our results indicate that realness of images is processed independently of image features in higher dorsal visual areas and at 250 ms.

Supported by CiNet, and by JSPS KAKENHI JP24KJ2222, JP25K172220.

26.452 LEARNING BEHAVIORAL REPRESENTATIONS OF COLOR COMBINATIONS WITH SEQUENTIAL MODELING

Henry A.S. Lewinsohn¹ (henry.lewinsohn@nih.gov), Erin Goddard², Bevil R. Conway¹; ¹National Institutes of Health, ²University of New South Wales

Color preferences are important for a wide variety of perceptual behaviors. They underwrite individual choices about clothing, influence the development of technology, and impact many aspects of Western civilization. Although single-color preferences follow a universally predictable pattern and evolve consistently with age (Ling and Hurlbert, 2011), little is known about how people evaluate combinations of colors and the underlying factors that shape these preferences. To address this question, we created a novel web-based color game deployed with many same-sex monozygotic (MZ, N = 424) and dizygotic (DZ, N = 127) twin pairs. Participants completed two tasks: (1) they created 3x3 grids following prompts, “make grids that you do (don’t) like,” using a palette of 37 colors sampling color space (hue, chroma, tone) and including focal basic colors; and (2) they provided aesthetic ratings of both individual colors and grids created by others. Results of individual color preferences replicated prior cross-cultural findings (Schloss et al., 2018). Twin modeling showed that most variance in both single-color and color combination preferences are attributable to environmental factors, aligning with results from face-preference heritability studies (Germine et al., 2015). To characterize individual variance in decision-making behavior during grid construction, we applied sequential models originally developed for natural language processing. In addition to behavior quantification, the sequential models learned latent representations of both color and spatial position, allowing us to derive embeddings that represent how participants implicitly structure the task space. By modeling the entire grid sequence, we obtain image representations based on spatial and temporal data. This approach opens avenues for comparing representations learned from sequential behavior with those derived from supervised or contrastive visual learning methods on static images, and for integrating modern sequential models with behavioral quantification.

Supported by NIH IRP (1ZIAEY000558 to BRC), NSF (0918064 to BRC) and NIH (R01 EY023322 to BRC). Contributions of NIH authors are considered Works of the United States Government. The findings and conclusions are those of the authors and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

26.453 STATE-SPACE MODELING REVEALS SEPARABLE COGNITIVE COLOR AND SHAPE REPRESENTATIONS IN MACAQUES DESPITE LIFELONG SALIENCE OF COLOR-SHAPE CONJUNCTIONS

Spencer Loggia^{1,2} (spencer.loggia@nih.gov), Bevil Conway^{1,3,4},
¹National Eye Institute, ²Brown University, ³National Institute of Mental Health, ⁴University of Maryland, College Park

A longstanding debate is whether the brain employs separate representations for color and shape. People can contemplate an object's color independently of its shape, but this may be learned, reflecting language and experience controlling the colors of objects. Neurophysiological work has been invoked both to support conjoint "multiplexing" and separate, parallel processing of color and shape, yet neural tuning cannot be directly mapped onto cognitive representations. Here we asked whether macaque monkeys reared interacting with 2-D digital objects acquired separable or conjoined cognitive codes. Over four years, four animals performed a touchscreen foraging task. On each trial they viewed four objects drawn at random from 1,296 possibilities and were rewarded according to the one they touched. Objects were defined in a perceptually uniform 4-D space (color: u^* , v^* ; shape: spikiness, animacy), with every shape in every color. Each color-shape combination was assigned a reward value that varied smoothly over this space. We analyzed choices with a two-stage cognitive state-space model (cSSM) that (i) learns a diffeomorphic warp of color-shape space to capture perceptual non-uniformities and (ii) discovers category basis functions whose dynamics implement an interpretable learning rule. The cSSM shows that behavior is not explained by memorizing individual color-shape pairs, but by a low-dimensional category geometry that evolves over experience. This geometry is initially dominated by shape and progressively incorporates color (avg. neg. log-likelihood of behavior under category model $0.43 \pm .04$, memorization model $0.58 \pm .05$). At plateau, the category geometry shows separable color and shape bases (avg. covariance magnitude $0.08 \pm .06$). Models trained to maximize reward without considering behavior instead learn conjunctive color-shape categories (avg. covariance magnitude $0.34 \pm .11$). These results indicate that the primate brain computes separate color and shape representations that underlie flexible cognition.

Supported by NIH IRP (1ZIAEY000558 to BRC), NSF (0918064 to BRC) and NIH (R01 EY023322 to BRC). Contributions of NIH authors are considered Works of the United States Government. The findings and conclusions are those of the authors and do not necessarily reflect

the views of the NIH or the U.S. Department of Health and Human Services.

26.454 CONTINUOUS FEATURE-TUNING OF PRIMING BY MASKED AND UNMASKED STIMULI

João Pedro Parreira Rodrigues^{1,2,3}
(joaopedro99parreira@gmail.com), Zefan Zheng^{4,5}, Karla Matic^{2,6,7,8},
Lukas Kob^{9,10}, Marlo Paßler⁹, Alex Lepauvre^{4,11}, John-Dylan Haynes^{2,6,7,8}; ¹Charité – Universitätsmedizin Berlin, Einstein Center for Neurosciences Berlin, ²Humboldt-Universität zu Berlin, ³Bernstein Center for Computational Neuroscience Berlin, ⁴Research Group Neural Circuits, Consciousness and Cognition, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany, ⁵Goethe University Frankfurt, Germany, ⁶Max Planck School of Cognition, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, ⁷Berlin School of Mind and Brain, Humboldt-Universität zu Berlin Berlin, ⁸Center for Advanced Neuroimaging, Charité Universitätsmedizin Berlin, Corporate member of the Freie Universität Berlin and Humboldt-Universität zu Berlin, ⁹Otto-von-Guericke University Magdeburg, Germany, ¹⁰Institute of Philosophy, Czech Academy of the Sciences, Prague, Czechia, ¹¹Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands

Priming has been widely studied across visual features, yet most paradigms rely on discrete stimulus differences, leaving the continuous relationship between perceptual similarity and priming strength largely uncharacterized. In parallel, substantial progress has been made in reconstructing perceptual structures across modalities and relating them to neural activity. However, a fundamental comparison between conscious and unconscious structures has yet to be conducted. Here, we use continuous priming combined with metacontrast masking to address this gap. Observers viewed a briefly presented color prime followed by a color target and performed a speeded discrimination task. The stimulus onset asynchrony between the prime and target was varied to render the prime either conscious or unconscious, and reaction time for identifying the colored target served as the basis for deriving perceptual distances. When the prime was visible, reaction times increased continuously with the CIE distance between prime and target, yielding a smooth priming gradient across the color space. Priming-derived distances strongly correlated with those obtained from both CIE space and subjective similarity ratings ($r > 0.9$). In contrast, this chromatic structure was disrupted under masked priming, with similar reaction times observed across prime-target color pairs. Our findings reveal a distance-dependent tuning function only under conscious conditions: primes produce progressively stronger facilitation as their perceptual similarity to the target increases, whereas primes outside conscious access fail to modulate target processing. This dissociation demonstrates a fundamental divergence between conscious and unconscious chromatic structures. More broadly, this work introduces a general psychophysical framework for constructing and directly contrasting

perceptual structures across the continuum of conscious and unconscious states.

26.455 ALIGNMENT AND REFINEMENT: HOW LANGUAGE SHAPES COLOR CATEGORY REPRESENTATIONS IN ARTIFICIAL NEURAL NETWORKS

Alexander Gokan^{1,2}, Arthur Shapiro^{3,4}; ¹Department of Neuroscience, American University, ²Center for Behavior and Neuroscience, American University, ³Department of Psychology, American University, ⁴Department of Computer Science, American University

How do color categories based strictly on visual information differ from those informed by language? We hypothesize two complementary ways that simple categorization schemes develop into complex categorization schemes: “aligning” (existing categories converge to a shared location in perceptual space), and “refining” (amplifying and sharpening existing categorical patterns). We address this question with a representational similarity analysis (RSA) on 21 neural networks, some trained only on images and others on images and language. We measure representational alignment using a stimuli-normalized cosine similarity between activations extracted at every layer. Responses to pairs of stimuli (simple equiluminant patches on gray background) that differed from each other only in hue were compared across all angles on the CIELuv color circle to compute discriminability, from which we estimate the categorical prototypicality of hues. RSA analysis produces the following results 1.) The first layer of both language and non-language networks converge on categories centered on cyan, magenta, and yellow (attributable to optimal coding of input statistics), though residual networks show systematically different categories enabled by skip connections. 2.) Language provides immediate group consensus about color categories with minimal training, allowing models to take advantage of extended training to build more nuanced categories. 3.) Language stabilizes color representations across the models’ hierarchy, indicating more gradual transformation of information. 4.) language does not impose novel categorical structure, but instead amplifies minor categorical distinctions already present without language. Conclusion: RSA can disentangle two distinct processes (group alignment and categorical refinement) by which language facilitates color category development in neural networks. Our results show that language does not construct new categories, but rather stabilizes and amplifies existing, nonlinguistic categorical patterns. Lastly, we discuss how RSA can be extended to perform a variety of psychophysics experiments on neural networks, and demonstrate its applications to unique hues and optical illusions.

26.456 THE POSSIBLE ROLE OF REAL-WORLD COLOR DISTRIBUTION RARITY IN AESTHETIC PREFERENCE FOR ABSTRACT PAINTINGS: A MACHINE-LEARNING APPROACH

Ikuto Hanada¹ (hanada.i.78fd@m.isct.ac.jp), Takehiro Nagai¹; ¹Institute of Science Tokyo

Understanding how color distributions influence aesthetic preference for paintings is an important issue in vision research. Previous studies have suggested that specific color statistics and the similarity of color distributions to natural-scene statistics may contribute to painting preference, yet their findings remain inconclusive. In this study, we aimed to examine the relationship between color preference for abstract paintings and the rarity of color distributions in real-world scenes quantified using a variational autoencoder (VAE) trained on naturalistic images. We first trained a VAE on a large naturalistic image dataset, such that color distributions with lower rarity were represented closer to the origin of the latent space. Accordingly, rarity was defined as the Euclidean distance from the origin in the latent space. Using this metric, we selected natural images spanning a wide range of rarity levels and created experimental stimuli by reflecting their color distributions onto several abstract base patterns. Preference was measured using Thurstone’s pairwise comparison method. As a result, although there were slight differences across base patterns, a strong positive correlation was consistently observed between the rarity of color distributions and preference: abstract paintings with rarer color distributions tended to be preferred. Analyses of color statistics showed that higher rarity was associated with greater variance in luminance, chroma, and hue, and that abstract paintings with larger variance in these dimensions were more likely to be preferred. These findings suggest that the rarity of color distributions may serve as an integrative concept that encompasses various color-related factors known to influence painting preferences. Overall, this study suggests the possibility that the rarity of color distributions underlies color preference for abstract paintings and helps unify diverse findings in the literature on painting aesthetics.

This study was supported by JST SPRING, Japan Grant Number JPMJSP2180.

26.457 PREDICTING THE AFFECTIVE CONNOTATION OF COLORMAP DATA VISUALIZATIONS FROM ESTIMATED EMOTIONAL ASSOCIATIONS OF CONSTITUENT COLORS

Halle C. Braun¹, Kushin Mukherjee², Seth R. Gorelik³, Karen B. Schloss¹; ¹University of Wisconsin-Madison, ²Stanford University, ³Woodwell Climate Research Center

Research in affective visualization design aims to predict the affective connotation of information visualizations. Previous work quantified the effects of hue, saturation, and lightness on affective connotation of monochromatic colormaps (Braun et al., 2026), but this simple regression approach could not easily extend to complex, polychromatic colormaps. We propose a new approach, which predicts affective connotation of whole colormaps from affective connotation of constituent colors (e.g., 256 colors in standard color scales, like Viridis). Although it may seem unlikely that the “whole” could be well-predicted by the sum of its “parts,” prior evidence supported the feasibility of our approach on a smaller scale: color-emotion associations for color pairs were well-predicted by their two constituent colors (Ou et al., 2024). In this study, we first generated

colormap data visualizations from 40 color scales (256 constituent colors) and had participants ($n=80$) rate associations between each colormap and five emotions (angry, sad, happy, disgust, and fearful). To generate predictions for each constituent color ($c=256$) within each color scale ($s=40$) for each emotion ($e=5$), it was infeasible to collect human color-concept association ratings (51,200 color-emotion pairs). Instead, we estimated associations computationally using color space regression models (Schloss et al., 2018) from existing color-emotion association data for 71 colors systematically sampled across CIELAB color space (Mukherjee et al., VSS-2022). For all five emotions, the aggregated association estimates across all colors within each color scale were strongly correlated with participant color-emotion association ratings for whole colormap data visualizations (angry $r=0.88$, sad $r=0.84$, happy $r=0.61$, fearful $r=0.68$, disgust $r=0.73$). Thus, affective connotation of whole colormaps was well predicted by the sum of their parts, despite having estimated the parts computationally rather than through direct judgments. Our approach has the potential to scale to predict affective connotation of colors in virtually any kind of visualization.

NSF grant BCS-2419493 to K.B.S

26.458 COLOR-CONCEPT ASSOCIATION FORMATION FOR ABSTRACT CONCEPTS THROUGH HIGHER-ORDER LEARNING

Melissa A. Schoenlein¹ (mschoenl@highpoint.edu), Melina O. Mueller², Karen B. Schloss³; ¹High Point University, ²Vanderbilt University, ³University of Wisconsin-Madison

Previous research found that people form color-concept associations for novel concrete concepts from direct exposure to color-concept co-occurrences (first-order relations) (Schoenlein & Schloss, 2022). However, people also have associations for abstract concepts lacking directly observable colors. How are color-concept associations formed for such concepts? We addressed this question by investigating whether participants could form color-concept associations from indirect co-occurrences (higher-order relations). In Experiment 1, participants were exposed to concept-concept co-occurrences: Nerky/Lomby tails belonging to Filk/Slub alien species, respectively. Then, participants were exposed to color-concept co-occurrences via images of warm-colored Filks/cool-colored Slubs. Last, participants rated how much they associated Nerky/Lomby tails with warm/cool colors (second-order relations). Participants associated warm colors more with Nerky tails and cool colors with Lomby tails ($p<.001$), despite them never directly co-occurring. In Experiment 2 we evaluated if concept relation type influenced strength of second-order associations. For one group, Lomby/Nerky represented tails again (has-a-property relation); for another, Lomby/Nerky represented cities near the aliens (located-near relation). Both groups formed warm-Nerky/cool-Lomby associations ($ps<.001$), with no significant differences depending on relation type ($p=.123$). In Experiment 3 we tested if participants could learn second- and third-order relations using a similar task structure, while adding an additional concept-

concept pairing: emotions-movie genres, movie genres-alien species, and alien species-colors. Participants formed associations between abstract concepts and colors through second-order relations (movie genres-colors), but evidence did not support forming associations through third-order relations (emotions-colors). It is possible these results demonstrate limitations on forming higher-order color-concept associations. Alternatively, additional exposure may be needed for third-order associations to emerge. Together, these findings demonstrate that people use indirect co-occurrences to form color-concept associations, even from concept-concept relations varying in relatedness. This work furthers our understanding of how people form color-associations for concepts without directly observable colors, permitting future investigations into association formation for more abstract concepts.

This work was supported in part by the National Science Foundation [BCS-1945303].

26.459 NEUROCOLOR ATLAS: MAPPING EMOTIONAL COLOR SIGNATURES IN AGING AND EARLY DEMENTIA THROUGH REAL-TIME PSYCHOPHYSIOLOGICAL VISUALIZATION

MEHJABIN SHIRIN¹, Dr Shivaram Male Male², Dr. P Phani Krishna Krishna³; ¹University of Hyderabad, ²University of Hyderabad, ³Delhi University

Color remains one of the most resilient perceptual and affective modalities of the human brain, supported by distributed neural systems that often remain functional even as higher cognitive processes deteriorate. This project introduces the NeuroColor Atlas, a real-time psychophysiological framework designed to characterise how ageing and early-stage dementia alter the brain's emotional responses to color. The study investigates whether affective color signatures neural-physiological patterns linking emotion and perception remain stable or exhibit measurable reorganisation during cognitive decline. A total of $N = 10$ older adults participated, comprising healthy individuals and early-stage dementia patients. Participants viewed a standardised set of chromatic stimuli while multiple physiological indices linked to autonomic and neurocognitive processing skin conductance, heart-rate variability, pupil dilation (LC-NE system activity), and gaze dynamics (oculomotor-attentional control) were recorded. These streams were temporally aligned and processed using a custom algorithm designed to extract neural-affective indicators embedded in dynamic physiological patterns. Resulting signatures were mapped into CIELAB and HSL color spaces to generate individualised chromatic profiles forming the core of the NeuroColor Atlas. Findings revealed distinct neurophysiological shifts in emotional color processing across ageing and dementia. Healthy older adults showed relatively coherent affective profiles with moderate variability in arousal responses to saturation and luminance, consistent with expected age-related changes in autonomic regulation. In contrast, dementia participants displayed diminished hue differentiation, disrupted pupil-based arousal modulation, and broader

gaze dispersion patterns suggestive of altered locus coeruleus-amygdala interactions and reduced attentional stability. Notably, warm, high-saturation colors continued to evoke preserved emotional resonance, indicating that certain affective pathways may remain intact despite cortical-cognitive decline. These findings highlight real-time color-emotion mapping as a promising neurometric tool for detecting subtle affective-perceptual alterations in early dementia. The NeuroColor Atlas provides a foundational step toward developing neuroscience-informed visual environments, emotion-supportive therapeutic strategies, and biomarker-driven assessments of cognitive ageing.

I gratefully acknowledges the support of the International Society of Color Council (ISCC) Student Support Award, which contributed to the development and dissemination of this research.

Theory

SATURDAY, MAY 16, 2:45 – 6:45 PM,
PAVILION

26.460 TRACKING TRUST IN ARTIFICIAL INTELLIGENCE WITH EEG MEASURES OF VISUAL ATTENTION (N2PC) AND WORKING MEMORY (CDA)

Tobias Feldmann-Wüstefeld¹, Eva Wiese¹; ¹Technische Universität Berlin

Visual attention and working memory are limited resources, making cognitive offloading to AI a potential means to enhance human performance. However, such benefits depend on the degree to which users trust the system. Here, we introduce a novel approach to measure trust in AI using two EEG components: the N2pc, indexing attentional deployment, and the CDA, indexing working-memory load. In Experiment 1, we recorded the N2pc in a visual search task in which participants either performed the task themselves (solo trials) or monitored a simple algorithm, framed as AI, performing it (coop trials). In solo trials, participants reported the orientation of a target line. In coop trials, the AI suggested a response of varying accuracy, and participants confirmed or overruled it. Because N2pc amplitude reflects the amount of attention allocated to a target, offloading should decrease its amplitude. Indeed, the N2pc was smaller when participants worked with a highly reliable AI than with an unreliable one, indicating greater willingness to offload. N2pc amplitudes were highest in solo trials, suggesting that participants never fully relinquished attentional control to the AI. In Experiment 2, we recorded the CDA in a change detection task. Solo trials required monitoring one hemifield only. In coop trials, participants memorized items from one hemifield while the algorithm (again framed as AI) was assigned the other. The CDA reflects the asymmetry of encoded items; reduced asymmetry therefore indicates less offloading. CDA amplitudes were generally smaller in coop than solo trials, showing that participants

continued to encode items from both sides rather than relying fully on the AI. When AI reliability was low, CDA amplitude was even further reduced, indicating increased monitoring of the AI's side due to distrust. Together, these findings demonstrate that N2pc and CDA amplitudes provide sensitive, implicit neural markers of trust in AI.

This research was supported by the Alexander von Humboldt Foundation and the Deutsche Forschungsgemeinschaft (DFG)

26.461 CATEGORY-SELECTIVITY IN ANNS DIFFERS SYSTEMATICALLY FROM HUMAN FMRI RESPONSES

Alish Dipani^{1,2}, N. Apurva Ratan Murty^{1,2}; ¹Cognition and Brain Science, School of Psychology, Georgia Tech, Atlanta, GA 30332, ²Center of Excellence in Computational Cognition, Georgia Tech, Atlanta, GA 30332

Category-selectivity for stimuli such as faces, scenes, bodies, and words is among the most striking and reproducible findings in vision neuroscience. With the emergence of artificial neural networks (ANNs), category-selective responses have also been reported in these models. Here, we show that category-selectivity in ANNs differs in meaningful and systematic ways from that observed in the human visual cortex. To this end, we first identified category-selective units across 35 ANN models using a standard human fMRI localizer. All pretrained models, but not untrained ones, exhibited robust category-selective units. Next, we found that responses ($N = 515$ stimuli from the NSD dataset) in the human category-selective brain regions ($N = 9$) were highly similar across individuals ($N = 8$, median univariate $R = 0.71$, median multivariate $R = 0.47$). This indicates that the structure of category-selective responses is highly consistent across people. How similar are the responses of category-selective ANN units to the same stimuli? Across all regions, no model reached the brain-brain ceiling in either univariate (median $R = 0.56$) or multivariate tests (median $R = 0.21$). Lesion tests further showed that the category-selective ANNs were neither necessary nor sufficient to predict responses in human category-selective brain regions. Finally, stimulus-level analyses revealed systematic mismatches between ANN and brain category-selectivity. Specifically, ANN units failed to reproduce robust neural response patterns, including comparable FFA responses to human and animal faces, suppression of PPA responses in scenes containing people, and the weak sensitivity of EBA responses to clothing. Taken together, these results show that although ANNs contain category-selective units, the structure of category-selectivity they express differs fundamentally from that observed in the human brain. Our findings highlight the need to go beyond ANN unit localization and compare the full representational structure of category-selective responses between ANNs and brains.

This work is supported by a CoCo fellowship by the School of Psychology, Georgia Tech, to AD; the NIH Pathway to Independence Award (R00EY032603), NSF Nexus (Allocation number: SOC250049), and a startup grant from Georgia Tech to NARM.

26.462 HUMANS VS. AI: ASSESSING THE SIMILARITY OF TASK-SPECIFIC SCENE DESCRIPTIONS FROM HUMANS AND GENERATIVE AI

Skylar Stadhard¹, Gillian Rosenberg¹, Bruce Hansen², Michelle Greene¹; ¹Barnard College, ²Colgate University

Multimodal generative AI can provide rich scene descriptions, but we have not yet established how similar AI-generated descriptions are to humans'. To what extent can generative AI learn a broad array of scene semantics from statistical distributions of text and images? Are some tasks more difficult to mimic? We created a broad battery of 15 tasks, divided into five groups: general knowledge, affordances, sensory experiences, affective experiences, and mental simulation. Human participants and 15 multimodal AI programs provided descriptions for 40 images on each task. We embedded text outputs using three sentence embedders, and we computed cosine distances between human and AI embedding vectors. Additionally, we assessed descriptions with traditional NLP metrics. Human descriptions were shorter than AI-generated counterparts (14.4 versus 28.3 words, $p < 0.0001$), but richer and more variable: they had higher lexical entropy (3.6 bits versus 3.4, $p < 0.0001$) and type-to-token ratio (0.89 versus 0.84, $p < 0.0001$), indicating humans used greater word variety. Cosine distances varied across models and tasks. Open-weight AI systems performed worse than closed-weight systems (mean cosine distances: 0.28 versus 0.21, $p < 0.0001$). Critically, we observed significant differences across tasks, with general knowledge about scene and object categories being easiest for AI (mean: 0.22) and assessments of scene affordances being hardest (mean: 0.29). This was somewhat surprising, given human sensitivity to scene affordances (Greene et al., 2016). One possibility is that humans produce compressed affordance descriptions due to their sensitivity, and AI consequently learns less about them. Contrary to commonly-held stereotypes of AI, affective tasks were comparatively easy for generative AI systems (mean: 0.23). AI can increasingly approximate human descriptions for many tasks, even tasks that require inferences that transcend mere image content, such as predicting future scene appearance.

NSF 2522311

26.463 FEEDBACK PROMOTES EFFICIENT-CODING WHILE REGULATING BIAS IN RECURRENT NEURAL NETWORKS

Holly Kular¹, Robert Kim^{2,3}, John Serences¹, Nuttida Rungratsameetaweemana³; ¹University of California San Diego, ²Cedars-Sinai Medical Center, ³Columbia University

Studies of human decision-making demonstrate that environmental regularities, such as the nonuniform image statistics of natural scenes, can be exploited to improve efficiency (termed 'efficient-coding'). Consistent with efficient coding, previous work has shown that the distribution of tuning preferences in early sensory areas changes during development to track the distribution of features in the

environment. In addition to these hard-coded changes, top-down feedback alters selectivity in early sensory areas to bias processing in favor of more likely stimuli when expectations change rapidly with task context. However, top-down feedback might also evolve over development to complement hard-coded changes in sensory areas to help balance efficient coding with the need to sometimes process infrequently encountered stimuli. To better understand the role of top-down feedback in exploiting naturally occurring statistical regularities, we trained a 3-layer hierarchical continuous-time recurrent neural network (ctRNN) to reproduce one of six possible inputs under biased conditions (stimulus 1 more probable than stimuli 2-6). Across all hidden layers, more information was encoded about high-probability stimuli, consistent with the efficient-coding framework. Importantly, reducing feedback from the final hidden layer of trained models selectively magnified representations of high-probability stimuli, at the expense of low-probability stimuli, across all layers. Thus, in addition to enhancing the selectivity of responses to expected stimuli, these results suggest that feedback pathways evolve to protect the processing of low-probability stimuli by regulating the impact of biased input statistics.

This work was funded by the National Eye Institute award (RO1-EY025872).

26.464 A REGRESSION THEORY OF DISSOCIATION BETWEEN BRAIN PREDICTION AND CONTROL

Binxu Wang¹, Jacob Prince¹, Thomas Fel¹, Akshay Jagadeesh^{1,2}, George Alvarez¹, Cengiz Pehlevan¹, Talia Konkle¹; ¹Harvard University, ²OpenAI

Deep neural network models can achieve high accuracy in predicting visual neuronal responses to natural images. Yet, when these same models are used to generate images to control neuronal activity, their success varies dramatically. Recent systematic evaluation work (Prince, Wang, Fel, Jagadeesh, et al.) revealed three surprising findings: (1) strong predictive performance does not ensure controllability; (2) models that fail at control tend to produce high-spatial-frequency modulations that are less effective for neurons; and (3) some models that cannot drive neurons themselves can still predict which stimuli generated by other models will be effective. We introduce a theoretical explanation based on a simplified linear regression framework. Prediction depends on the inner product $\mathbf{w}^{\text{top}} \cdot \mathbf{x}$ between regression weights and natural image features, whereas neural control requires moving along the weight direction $\mathbf{w}$ itself. Because natural images exhibit highly anisotropic feature covariance—with high-frequency components contributing low variance—regression tend to amplify these poorly constrained directions. This allows large changes in $\mathbf{w}$ with minimal impact on prediction, dissociating encoding accuracy from controllability. When ground-truth neural tuning lies primarily within the natural image manifold, learned weights can acquire substantial off-manifold components. Image synthesis then pushes stimuli toward unnatural directions that are ineffective in real neurons, although these

same models can still evaluate the success of stimuli produced by better-aligned models, which are more on manifold. Finally, applying random matrix theory, we predict how sample size, feature dimensionality, and neuronal noise level determine which weight modes are most error-prone, and thus when control will fail. Together, this framework explains why good encoders can be poor controllers, and motivates new approaches for on-manifold feature accentuation and robust closed-loop neural control.

This work has been made possible in part by a gift from the Chan Zuckerberg Initiative Foundation to establish the Kempner Institute for the Study of Natural and Artificial Intelligence. B.W. is funded by the Kempner Research Fellowship.

26.465 BENCHMARKING PERFORMANCE METRICS FOR CONTINUOUS PSYCHOPHYSICS ACROSS OCULOMOTOR AND HAND MOTOR TRACKING MODALITIES

Aleksandrs Koselevs¹, Veronica Pisu¹, Ömer Yildiran^{2,3}, Chloe Lam¹, Saivydas Villani¹, Pascal Mamassian², Dominik Straub^{4,5}, Constantin Rothkopf⁴, Guido Maiello¹; ¹University of Southampton, ²Laboratoire des Systèmes Perceptifs, DEC, ENS, PSL University, Paris, France, ³Department of Psychology, New York University, New York, USA, ⁴Institute of Psychology, Centre for Cognitive Science, Technical University of Darmstadt, Darmstadt, Germany, ⁵Department of Engineering, Cambridge University, Cambridge, UK

Continuous psychophysics relies on tracking behaviour to estimate visual functions (Bonnen et al., 2015). Poorer tracking reflects greater perceptual uncertainty: harder-to-see targets are tracked less faithfully. But which is the best way to quantify tracking? Performance can be quantified from the peak, lag, and width of the cross-correlation between target and tracking velocities. As performance worsens, peak cross-correlation decreases, occurs at longer lags, and the cross-correlation broadens. However, these measures are not purely perceptual: they also reflect motor and cognitive factors, and are strongly intercorrelated, limiting their usefulness as independent descriptors. Fitting a Kalman filter to the tracking data instead yields a single observation-noise parameter, but its relationship to cross-correlation measures remains unclear. Here, we compared cross-correlation measures to the Kalman-derived parameter across five datasets: the Bonnen et al. dataset and four in-house datasets (pooled N=27). Participants tracked Gaussian-blob targets via mouse or eye tracking, and we manipulated target visibility and hand performance (dominant vs non-dominant hand). As expected, cross-correlation peak, lag, and variability were strongly intercorrelated, although interestingly, both the magnitude and direction of intercorrelations varied across datasets (range of correlations: $r=[-.81, .70]$). Nevertheless, across all datasets pooled together, the Kalman-derived observation-noise parameter consistently correlated with each cross-correlation metric (peak-noise: $r=-0.37$; lag-noise: $r=0.20$; variability-noise: $r=0.17$, all $ps < 0.05$), suggesting that it could serve as a single summary index of variation across these metrics. Together, these findings show that the Kalman-derived observation-noise estimate

provides a unified descriptor of tracking performance across datasets and tracking modalities, although it cannot by itself disentangle visual, motor, and cognitive contributions. The ultimate goal of this project is to apply inverse optimal control (Straub & Rothkopf, 2022) to decompose these components and provide more interpretable, mechanism-level estimates of the processes underlying tracking behaviour.

This research was funded by Fight for Sight (Grant Reference: RESSGA2302, awarded to GM) and Wessex Medical Research (Grant Reference: AF04, awarded to GM).

26.466 HOW DOES YOUR BRAIN KNOW WHAT'S REAL? DOMAIN ADAPTATION AND FMRI APPROACHES TO REVEALING THE NEURAL MECHANISMS OF REALITY MONITORING

Ali Ekhlasi¹ (aeekhiasi@uci.edu), Emil Olsson¹, Michaela Klimova², Xueyi Huang², Angela Shen¹, Nadine Dijkstra³, Jorge Morales², Megan A. K Peters¹; ¹University of California, Irvine, ²Northeastern University, ³University College London

The fundamental question of how the brain distinguishes internally generated experiences (imagination) from external reality (perception) is central to vision science. Higher-Order (HO) theories of consciousness propose that this distinction relies on a Reality Monitoring (RM) signal that evaluates first-order visual content. However, leading HO models disagree on the nature of this signal. Perceptual Reality Monitoring (PRM) (Lau, 2019, 2022) posits a dedicated RM signal. The Higher-Order State Space (HOSS) framework (Fleming, 2020; Dijkstra & Fleming, 2023) proposes that vividness of the internal representation serves as the primary RM cue. Rich HO theories, such as Higher-Order Representation of a Representation (HOROR; Brown, 2015), suggest that HO representations contain content dimensions themselves. To arbitrate among these models, and as part of an adversarial collaboration (TWCF-2021-22032), we introduce a novel analytical framework using Domain Adaptation (DA), a machine-learning technique that statistically aligns fMRI activation patterns across perception and imagination. Using two publicly available fMRI datasets in which participants perceived and imagined objects (Margolles et al., 2024; Horikawa & Kamitani, 2017), we used DA-enhanced searchlights to discover brain regions that encode content similarly across perception and imagination. Critically, we also identified regions whose brain activity patterns were divergent between both states, but strongly benefited from DA's statistical alignment—showing improved alignment of brain activity distribution but not improved content decoding. This analysis allows us to find regions that may specifically differentiate between reality and imagination regardless of the content being represented. Our method suggests that these “pure RM” regions are located in the dorsomedial prefrontal cortex (dmPFC), dorsal parietal cortex, and the left ventral occipital cortex. These findings provide the basis for directly testing our core hypotheses: whether a dedicated, source-specific RM dimension (as predicted by PRM)

exists independently of the visual content's vividness and features, and which brain regions may support such representations.

This work was supported by the Templeton World Charity Foundation, Inc. (funder DOI 501100011730) under the grant <https://doi.org/10.54224/22032> to (JM, MAKP & ND) and in part by the Canadian Institute for Advanced Research (CIFAR, Fellowship in Brain, Mind, & Consciousness; to MAKP).

26.467 A NEW IMAGE-COMPUTABLE APPROACH TO SENSORY-PERCEPTUAL ESTIMATION AND DISCRIMINATION

Johannes Burge¹; ¹University of Pennsylvania

A fundamental problem of perception is to estimate behaviorally relevant properties of the environment, and the organism's relation to it, from highly variable and uncertain stimuli. Here, we report a new class of image-computable ideal observers—spectral ideal observers—that optimally solve such problems. Spectral ideal observers are distinguished from more traditional ideal observers because they do not assume target stimuli with a stereotyped spatial structure; rather, they process stimuli with random spatial structure as perceptual systems must in natural viewing. Further, they operate on measurements of the real and imaginary coefficients of a stimulus' Fourier transform—rather than on amplitude and phase, or on pixels—, by characterizing the joint probability distributions of these coefficients for different states of the latent variable of interest. For filtered (e.g. 1/f) noise stimuli—which share statistical properties with natural images—the coefficients are conditionally Gaussian; for natural images and sounds, appropriate normalization Gaussianizes them. From these distributions, the posterior over the latent variable, estimates, and/or the optimal Bayesian-theoretic decision variable can be computed. The results provide principled performance predictions, and reveal the receptive fields and energy-model-like neural computations that optimally support performance. We develop a spectral ideal observer for estimating focus error and pupil size from signals in the cone responses to small patches (0.5deg) of images blurred by the optics of individual human eyes. Estimates are accurate and precise. Hence, image-based estimates of focus error (and pupil size) could be used by circuits that drive human accommodation. We discuss the design of experiments that will test novel predictions of this theory. We additionally develop spectral ideal observers for other basic tasks in visual and auditory science: 3D surface attitude (i.e. slant & tilt) estimation, binocular disparity estimation with different optics in the two eyes, and binaural sound localization (elevation & azimuth).

This work was supported by the National Eye Institute and the Office of Behavioral and Social Sciences Research, National Institutes of Health Grant R01-EY028571 to J.B.

26.468 A SIGNAL DETECTION FRAMEWORK FOR INTERPRETING BEHAVIORAL SIMILARITY

Yu (Eric) Qian¹ (ericqian@utexas.edu), Wilson S. Geisler¹, Xue-Xin Wei¹; ¹University of Texas at Austin

Introduction: Analyzing behavioral similarity between observers is fundamental in perceptual and behavioral sciences. One recent application involves assessing how well the behavior of computational models (e.g., deep neural networks) predict that of animals or humans. While a number of methods were proposed (e.g., Cohen's kappa, Gwet's AC1 and Krippendorff's alpha), how to properly interpret these metrics remains debated. **Methods:** We proposed a principled framework based on signal detection theory (STD) to study behavioral similarity. In particular, we leverage a recent generalization of STD that models the decision of two observers simultaneously in binary choice tasks. This framework models the joint distribution of the decision variables of the two observers, and their similarity is captured by decision variable correlation (DVC). The decision of each observer is determined by the decision variables and criterion. By varying DVC and other key variables (such as d-prime, criterion of the observers), we can investigate how other popular behavioral metrics depend on key variables in STD. **Results:** We find that estimating DVC from behavior is reliable from experimentally relevant sample sizes. Investigating popular metrics (e.g., Cohen's kappa), we find that these measures are monotonic in DVC but have complex relationships with other variables in STD. They are sensitive to d-prime, criterion of individual observers, and class imbalance. For example, Cohen's kappa reduces systematically when the difference of the accuracy between the two observers increases. In conclusion, we show that DVC is a robust and principled measure of behavioral similarity when information about the ground truth class is available. **Conclusion:** Under the STD framework, the commonly reported behavioral agreement metrics are co-determined by DVC, d-prime, criterion and class imbalance. Thus, DVC and other STD variables together constitute a more faithful and comprehensive representation of the underlying decision making processes.

This research is funded by NIH R01NS133924 and a Sloan Research Fellowship (to X.-X. Wei).

26.469 WHY “CONVERGENT EVIDENCE” AND “REPRESENTATIONAL SIMILARITY ANALYSIS” FAIL AS METHODOLOGICAL PRINCIPLES IN VISION SCIENCE

Fernando Ramirez¹; ¹Congressional Center for Computational Neuroscience (CCCN), Maryland, USA

Appeals to “convergent evidence” are widely used in vision science to justify inferences that integrate results across paradigms, measurement techniques, and species. In contemporary cognitive and systems neuroscience, particularly in studies using multivariate analyses of neural population activity, this strategy is often treated as principled rather than heuristic, notably through “Representational Similarity Analysis” (RSA). Here, I argue that convergent evidence is logically incoherent as a general criterion of evidential validity, and this incoherence linked to foundational assumptions underlying RSA. RSA

is often motivated by geometric intuitions related to the Johnson–Lindenstrauss (JL) lemma, which guarantees approximate preservation of Euclidean distances under random linear projections of a fixed set of points. This is taken to license abstraction from measurement and direct comparison of dissimilarity structure across methods. I show that this inference fails on logical grounds: neuroscientific measurements are neither random nor generic projections, and the multivariate patterns analyzed with RSA are constructed by measurement, preprocessing, and normalization operations. Using previously published neuroimaging and electrophysiological datasets together with novel computational modeling and analytical work, I derive necessary conditions under which RSA-based comparisons are valid and illustrate how violations arise in practice. Representational structure is not invariant to translations or rotations in multivariate space when signal imbalances and measurement-scale differences are present. Moreover, empirically observed pattern spaces are typically radial rather than isotropic, so vector length carries representationally relevant information. Under these conditions, RSA outputs that appear similar can support logically incompatible conclusions about the underlying neural code. These failures are compounded by dependence on signal-to-noise regime: different but plausible noise assumptions can flip representational inferences. I conclude that neither convergent evidence nor RSA can serve as general methodological principles in vision science. Valid cross-method inference requires explicit forward models specifying how hypothesized neural representations, noise, and measurement processes jointly generate empirical observations.

26.470 MULTIARRANGEMENT: A PLUG & PLAY GEOMETRIC DATA COLLECTION PACKAGE FOR VIDEO STIMULI

Umur Yıldız^{1,2} (umur.yildiz@bilkent.edu.tr), Burcu Ayşen Ürgen^{1,2}; ¹Bilkent University, Ankara, Turkey, ²Aysel Sabuncu Brain Research Center and National Magnetic Resonance Research Center (UMRAM), Bilkent University, Ankara, Turkey

Naturalistic videos are increasingly used in vision science to probe how the visual system represents events, but standard behavioral methods still rely on sequential ratings or exhaustive pairwise comparisons of individual clips. These approaches are temporally inefficient for large stimulus sets and capture little contextual information across stimuli. While the multi-arrangement method (Kriegeskorte & Mur, 2012) offers a scalable alternative, practical implementation options are limited. Existing tools typically focus on static images, require payment, or demand substantial custom implementation, leaving no free, off-the-shelf software option. To address this gap, we present Multiarrangement, a Python toolkit designed to efficiently implement the multi-arrangement method for dynamic visual stimuli. During the procedure, participants arrange subsets of videos within a circular arena such that spatial distances correspond to perceived dissimilarity. These partial arrangements are then aggregated to estimate a full representational dissimilarity matrix

(RDM) reflecting the high-dimensional geometry of the stimulus space. The toolkit includes two trial schedulers: a fixed set-cover strategy that guarantees at least one co-occurrence for every unordered pair while keeping trial size small and that does not require an initial full-stimulus trial, and an adaptive lift-the-weakest strategy that preferentially samples poorly constrained regions of the similarity space. We validate the approach with two participants on a 58-item set of human action videos, comparing multi-arrangement dissimilarities against those from traditional one-by-one comparison tasks. Multiarrangement yields stable similarity structures that closely track those obtained from pairwise ratings (mean Pearson $r = 0.77$, concordance correlation = 0.68), while requiring fewer explicit comparisons and showing a mild compression of the dissimilarity range. The method offers a practical way to obtain dense, behaviorally grounded similarity data for large sets of naturalistic videos with minimal programming effort, providing a useful bridge between human judgments, computational models, and neural measures in vision science.

Multisensory Processing: Recalibration, temporal

SATURDAY, MAY 16, 2:45 – 6:45 PM, PAVILION

26.471 THE CROSS-MODALITY TRANSFER OF CALIBRATION EFFECT

Yanfei Yu¹ (yuyanfei31@gmail.com), Urvi Maheshwari², David Barner³; ¹University of California, San Diego, ²University of California, San Diego, ³University of California, San Diego

Humans can represent large quantities imprecisely (Cordes et al., 2001; Gibbon, 1977), but it is debated whether this reflects a single amodal system or multiple modality-specific capacities. The Approximate Number System (ANS) account proposes a modality-independent number sense, with similar acuity across visual, auditory, and tactile formats (Barth et al., 2003; Tokita & Ishiguchi, 2016). In contrast, the Sensory Integration System (SIS) theory argues that numerical estimates arise indirectly from modality-specific computations of correlated magnitudes, such as area or duration (Gebuis et al., 2016). The present study tested these views by asking whether feedback-based calibration of number estimation in one modality (vision or audition) transfers to the other. We employed a 2 (Inducer: overestimated vs. underestimated) $\times$ 2 (Modality: visual vs. auditory) $\times$ 2 (Session: pre- vs. post-calibration) mixed design, with inducer as a between-subjects factor. Adults ($N = 61$ and 60 in Experiment 1 and 2) estimated quantities of dots (10–97) and tones (10–37) before and after calibration. During calibration, participants were told that the inducer displays contained 42–48 dots in Experiment 1 and 20–28 tones in Experiment 2, although the actual stimuli contained either 3/4 this N (overestimated inducer) or 4/3 this N (underestimated inducer). Experiment 1 showed increased visual and

auditory estimates after calibration in both inducer conditions, consistent with a general priming effect of large digits. However, only visual estimation showed an inducer x session interaction, suggesting a within-modality calibration effect that did not transfer to audition. In Experiment 2, auditory estimates increased after calibration only in the overestimated condition, and we observed transfer to visual estimation for values near the inducer set size, compatible with a modality-independent number representation argued by ANS. Ongoing work is exploring the possibility that the strong priming effect in Experiment 1 may have masked cross-modal calibration, producing weaker transfer.

26.472 AUDIOVISUAL TEMPORAL RECALIBRATION EXERTS AN INFLUENCE ON SOUND-INDUCED FLASH ILLUSION

Yeeun Kang¹, Chai-Youn Kim¹; ¹Korea University

Prolonged exposure to audiovisual temporal asynchrony induces recalibration, where the point of subjective simultaneity shifts toward the adapted lag (Fujisaki et al., 2004; Vroomen et al., 2004). Previous studies primarily used explicit temporal judgment tasks, leaving unclear whether recalibration extends to implicit audiovisual integration. We investigated whether audiovisual lag adaptation modulates the temporal binding window (TBW) measured through the sound-induced flash illusion (SIFI; Shams et al., 2000), where a single flash paired with multiple beeps can be perceived as multiple flashes. Eleven participants completed SIFI trials before and after lag adaptation, judging whether they perceived one or two flashes. Each trial presented a white circle under fixation with auditory beeps in congruent (one-flash one-beep [1F1B], two-flashes two-beeps [2F2B]) and illusory (one-flash two-beeps [1F2B]) conditions. In 1F2B and 2F2B trials, stimulus onset asynchronies (SOAs) between a flash-beep pair and additional component (a beep or another flash-beep pair, respectively) varied within ± 300 ms. Adaptation sessions exposed participants to either auditory-leading (-235ms) or visual-leading (+235ms) asynchrony using identical audiovisual stimuli as SIFI. Participants performed an oddball detection task during three-minute initial adaptation, with post-adaptation SIFI blocks interleaved with top-up adaptation. We analyzed illusion rates in 1F2B trials to derive perceptual sensitivity (d') and TBW (Foss-Feig et al., 2010). While not significant, post-adaptation TBWs shifted approximately 50ms toward each adapted lag (Baseline: -150 to 100ms; Visual-first: -50 to 150ms; Auditory-first: -100 to 100ms). Visual-first adaptation yielded significantly higher d' at -300 and -100ms SOAs than auditory-first ($ps=0.040, 0.032$). The magnitude of sensitivity changed from baseline ($\Delta d'$) was significantly greater after visual-first than auditory-first adaptation at these SOAs ($ps=0.041, 0.033$). These findings reveal that audiovisual temporal recalibration can be detected via SIFI through changes in perceptual sensitivity. Notably, auditory-leading exposure induced stronger recalibration than visual-leading exposure, indicating asymmetric plasticity in temporal processing.

Supported by NRF-2023R1A2C2007289

26.473 PERCEPTUAL CONFIDENCE FOR AUDITORY-INDUCED VISUAL POSITION BIASES

Alejandro Lorca Vyhmeister¹, Pascal Mamassian^{1,2}; ¹École Normale Supérieure Paris, ²Centre national de la recherche scientifique (CNRS)

To efficiently interact with the world, we constantly integrate different sensory sources and engage in metaperceptual processes to judge the validity of the resulting multisensory percepts. However, most research in metaperception has focused on unisensory stimuli. Here, we are interested in the metaperception of multisensory percepts. We designed an audiovisual paradigm that biased the perceived position of a visual flash by manipulating the timing of an auditory sound. Visual stimuli were three Gaussian blobs presented successively at three vertically aligned positions. Auditory stimuli were pure tones. All flashes and tones lasted 33ms. The first and third visual and auditory stimuli were synchronized and separated by 467ms. The second flash was presented at one of five possible positions near the spatial midpoint, and always halfway in time (233ms from the beginning). The time of the second tone could start early, on time, or late relative to the second flash (at 167, 233, or 300ms from beginning). A no-sound condition was also added as a baseline. Participants reported the perceived position of the second flash with the computer mouse. After their perceptual decision, participants judged how confident they were in the accuracy of their perceptual report using a continuous scale. As expected, reported positions of the second flash followed physical positions, with some regression to the mean of the five possible positions, and some additional biases when the tone was presented early. Importantly, early tones created a spatial bias towards the first flash, and reversely for late tones. For each flash position condition, confidence judgments were higher for perceptual reports closer to the median position for that condition, thereby revealing metacognitive sensitivity. Finally, no differences were observed in metacognitive sensitivity across all sound conditions. These results suggest that participants cannot detect visual spatial biases induced by a sound in our experiment.

MSCA Doctoral Network "CODE" (grant number: 101119647)

26.474 THE ROLE OF AUDIOVISUAL ASYNCHRONY IN DRIVING CROSS-MODAL PREDICTIONS

Merve Kınıklıoğlu¹ (mervekinikloglu@gmail.com), Daniel Kaiser^{1,2}; ¹Neural Computation Group, Department of Mathematics and Computer Science, Physics, Geography, Justus Liebig University Giessen, Germany, ²Center for Mind, Brain and Behavior (CMBB), Universities of Marburg, Giessen, and Darmstadt, Germany

Perception is an active process, with the brain continuously anticipating sensory input. While predictive mechanisms have been well studied within single modalities, real-world perception relies on cross-modal predictions, where information from one modality shapes expectations in another. Here, we introduce controlled temporal asynchrony between auditory and visual streams to test whether

leading auditory inputs predictively modulate the neural processing of upcoming visual inputs. In an fMRI study, participants (N = 31) viewed sixteen 10-second naturalistic videos of everyday cooking actions (e.g., chopping, peeling, blending, pouring, frying, grating). Audio tracks were shifted by three seconds to temporally precede (“leading” condition) or trail (“lagging” condition) the visual input, or remained synchronized (“synchronous” condition). Participants monitored audiovisual congruency by detecting occasional mismatched sound–video pairs. Pairwise decoding of the sixteen video stimuli within each ROI was performed to assess the discriminability of stimulus-specific activity patterns across conditions. Leading sounds produced higher decoding accuracy than lagging sounds in scene-selective PPA and object-selective LO, indicating that predictive auditory cues enhance visual representations of upcoming events across visual regions. In contrast, decoding accuracy in multisensory regions (IPS and IFG) tended to be higher in the lagging than in the leading and synchronous conditions, consistent with greater demands for temporal conflict monitoring. A whole-brain comparison between the synchronous vs. asynchronous conditions revealed only partially overlapping regions, suggesting that predictive effects under asynchronous conditions are spatially dissociable from integration effects. Together, these findings demonstrate how anticipatory signals from one modality predictively shape representations in another modality.

This work was supported by the Deutsche Forschungsgemeinschaft (German Research Foundation, DFG) under Germany's Excellence Strategy (EXC 3066/1 “The Adaptive Mind”, Project No. 533717223)

26.475 CORRELATION DETECTION AS A STIMULUS COMPUTABLE ACCOUNT FOR AUDIOVISUAL PERCEPTION, CAUSAL INFERENCE, AND SALIENCY MAPS IN MAMMALS

Cesare Parise¹; ¹University of Liverpool

Despite recent progress in multisensory research, the absence of stimulus-computable perceptual models fundamentally limits our understanding of how the brain extracts and combines task-relevant cues from the continuous flow of natural multisensory stimuli. Here, we introduce an image- and sound-computable population model for audiovisual perception, based on biologically plausible units that detect spatiotemporal correlations across auditory and visual streams. In a large-scale simulation spanning 69 psychophysical, eye-tracking, and pharmacological experiments, our model replicates human, monkey, and rat behaviour in response to diverse audiovisual stimuli with an average correlation exceeding 0.97. Despite relying on as few as 0–4 free parameters, our model provides an end-to-end account of audiovisual integration in mammals—from individual pixels and audio samples to behavioural responses. Remarkably, the population response to natural audiovisual scenes generates saliency maps that predict spontaneous gaze direction, Bayesian causal inference, and a variety of previously reported multisensory illusions. This study demonstrates that the integration of audiovisual stimuli, regardless of their spatiotemporal complexity, can be accounted for in terms of

elementary joint analyses of luminance and sound level. Beyond advancing our understanding of the computational principles underlying multisensory integration in mammals, this model provides a bio-inspired, general-purpose solution for multimodal machine perception.

26.476 NOT THE COMMON CAUSE FOR THIS TIME: NO SIGN OF CAUSAL INFERENCE IN AUDIOVISUAL DURATION INTEGRATION

Omer F. Yildiran¹ (ofy204@nyu.edu), Long Ni^{1,2}, Michael S. Landy^{1,2}; ¹Department of Psychology, New York University, ²Center for Neural Science, New York University

How does the brain integrate conflicting temporal information across sensory modalities? Using small conflicts (± 50 ms), Hartcher-O'Brien et al. (2014) showed that observers integrate audiovisual duration cues optimally. Does causal inference lead to a breakdown of audiovisual integration when duration conflicts are large? Two levels of auditory signal-to-noise ratio (SNR) were used. We confirmed in a unimodal 2-IFC task that lower auditory SNR increased auditory duration discrimination threshold. Cross-modal duration discrimination revealed a consistent bias: auditory intervals were perceived as longer than physically matched visual intervals. This bias was taken into account in the main bimodal experiment. In the bimodal task, participants compared the auditory durations of two bimodal stimuli. Test stimulus: the two cues were consistent (same apparent duration). Standard stimulus: the two cues differed by seven conflict durations (range: ± 250 ms). Auditory duration percepts shifted systematically toward the visual duration, especially when auditory uncertainty was high. The shift was proportional to the amount of cue conflict, inconsistent with the predictions of causal inference. We compared fits of several models. A heuristic model in which the observer probabilistically switches between the visual and auditory cues was preferred for most participants. However, AIC differences across models were small. Simulations revealed why the models were difficult to distinguish: Given the observed sensory noise, competing models can only be discriminated using unreasonably large conflicts. Within the reasonable conflict range used here, all models produced largely overlapping, near-linear patterns of shift as a function of the degree of cue conflict. In conclusion, observers do not rely on causal inference when judging durations under our conditions. The heuristic, cue-switching is favored, but the large uncertainty of duration estimates fundamentally limits our ability to discriminate computational models in this domain.

NIH EY08266, NYUAD Center for Brain and Health, funded by Tamkeen under NYU Abu Dhabi Research Institute award CG012

26.477 MULTISENSORY EXPLORATION OF VISUAL MOTION PERCEPTION

Danica Barron¹, Ralph Hale², Troy Smith³; ¹University of North Georgia

Smooth pursuit eye movements are optimized for rigid motion, where spatial relationships remain stable. Non-rigid motion, which continuously deforms, impairs pursuit by reducing gain and increasing positional errors. Koerfer et al. (2024) found that non-rigid motion is incapable of being tracked with smooth pursuit. Preliminary evidence from Barron et al. (2025) suggests a trend toward auditory cues reducing corrective saccade errors during non-rigid motion. Although pursuit is typically considered a predominantly visual process, multisensory theories suggest that redundant auditory or tactile cues could enhance target localization when visual coherence is degraded. We tested whether synchronized audio and tactile cues influence pursuit during non-rigid motion tracking. Participants tracked a vortex-style target across five conditions: rigid visual motion (baseline), non-rigid visual motion alone, and non-rigid motion paired with spatially mapped auditory cues, vibrotactile feedback, or combined audio-tactile cues. Primary measures included pursuit gain, absolute pursuit speed, and post-saccadic positional error. We predicted that non-rigid motion would reliably impair pursuit relative to rigid controls, and that multisensory cues would impact tracking behavior. Auditory cues may provide supplementary velocity information through continuous modulation, whereas tactile feedback may enhance body-centered localization. The combined condition tests whether multisensory redundancy yields additive improvements or interference. Patterns indicating improvements in gain or reduced corrective saccades would support a role for cross-modal influences on oculomotor tracking, whereas persistent deficits would reinforce visual dominance in oculomotor control. This study evaluates whether smooth pursuit's traditional visual-centric characterization holds under conditions of unstable motion structure, demonstrating how multisensory signals impact tracking behavior.

SUNDAY MORNING POSTERS IN BANYAN BREEZEWAY

Color, Light and Materials: Adaptation, contrast, lightness, brightness

**SUNDAY, MAY 17, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY**

33.301 CHROMATIC ADAPTATION SYSTEMATICALLY RESHAPES HUMAN COLOR DISCRIMINATION THRESHOLDS

Cameron Zhang¹, Kara Wang¹, Jason Chow², Phillip Guan², Alex H. Williams³, David H. Brainard¹, Fangfang Hong¹; ¹Department of Psychology, University of Pennsylvania, ²Reality Lab Research, Meta, ³Center of Neural Science, New York University

Chromatic adaptation to a steady background has a profound effect on color discrimination thresholds. Previous studies that examined this

effect typically restricted threshold measurements to cases in which the reference stimulus matched the adapting chromaticity, to a few discrete references, or along a limited set of color-perturbation directions. It remains unknown how chromatic adaptation reshapes full iso-discrimination contours when the reference varies continuously throughout color space. Here, we investigate this question across an isoluminant plane by using adaptive trial placement combined with a semi-parametric Wishart Process Psychophysical Model (WPPM) to characterize thresholds at all references and color-perturbation directions. We studied two adaptation conditions with matched luminance: a neutral gray and a blue daylight (~12,000 K) background. On each trial, participants (N = 2) viewed three matte, blobby objects for 1s, and identified the odd one out based on surface color. The two adaptation conditions were interleaved two runs per session, with condition order randomized within each session. Participants adapted to the background for 120s before the first trial of each run. Each participant completed ~7,500 trials per condition, and we fitted the WPPM separately for each. The results reveal clear changes in the field of iso-discrimination contours across the adapting backgrounds. Under gray adaptation, the smallest contour (highest sensitivity) is centered on the reference whose chromaticity matches the gray background. Under blue adaptation, the smallest contour shifts towards the background, while locations that were most sensitive under gray adaptation have increased thresholds. In both cases, the major axes of the contours tend to point toward the adapting background. These findings indicate that changes in adapting background systematically modulate color discrimination thresholds. This dataset provides a rich benchmark for evaluating mechanistic models of color vision and hypotheses about efficient coding.

33.302 ASYMMETRIES IN THE STRENGTH OF CHROMATIC INDUCTION ALONG THE S+ AND S- HALF- AXES OF MB/DKL COLOR SPACE

Michael E. Rudd¹ (mrdudd@unr.edu), Sean O'Neil¹, Natasha Staneva¹;
¹University of Nevada, Reno

At VSS 2025, we demonstrated an asymmetry in chromatic induction strength along the two half-axes of the LM dimension of MB/DKL color space. Here, we demonstrate a similar asymmetry along the S axis. Two achromatic disks (1.16 deg diam, 30 cd/m²) were presented on opposite sides of a Display++ monitor (Cambridge Research Systems, Kent, UK), each surrounded by an achromatic annulus (60 cd/m²). The annuli had widths 0.19 and 1.78 deg. The display side containing the thinner annulus was randomized across trials. The luminance of the chromatic background was fixed at 45 cd/m². Across trials, the chromatic contrast of the background was randomly varied from 80% S- to 80% S+ in 17 equal steps. Four participants adjusted the S content of the one of the two disks to match the disks in hue. 50% of the time, the match disk was surrounded by the thin annulus, and 50% of the time it was surrounded by the thick annulus. Changes in the S content of the background induced a contrasting hue change in both the disks and annuli. These hue changes were stronger on the side

with the thinner annulus. We wondered if, when changing saturation of the background, the rate of change in the match settings would differ across the background ranges that were either S+ or S- relative to the achromatic point. If so, this might imply the existence of separate underlying half-wave rectified S+ and S- mechanisms characterized by different gains and producing different chromatic induction strengths. This prediction was confirmed, but the direction of the asymmetry varied across observers. In combination with the results of our previous study, these results support a generalization of an edge integration theory of lightness (Rudd, 2010; Rudd & Shareef, 2025) to the chromatic dimensions of color space.

Supported by NIGMS P30 GM145646

33.303 POPULATION-LEVEL REVERSE CORRELATION REVEALS CONTEXT EFFECTS ON THE NEURAL REPRESENTATION OF LUMINANCE

Tom P. Franken¹, Fatemeh Didehvar¹; ¹Washington University in St. Louis

The lightness of a surface is strongly influenced by scene context. Prior studies found that individual neurons in early visual cortex can exhibit context effects similar to perception, but such effects are often heterogeneous. It remains therefore poorly understood how context affects neural representation at the population level. To tackle this question, we used Neuropixels to record from columnar populations of neurons in visual cortical Area V4 in a macaque. During these recordings the animal fixated on scenes tiled into adjacent surface elements. The classical receptive fields (cRFs) of the neurons were centered within one element of fixed luminance. Random luminance noise was independently added to the other elements, and we presented >10000 such unique scenes per session (FixedLocalLum). In the same sessions we also recorded neural activity to scenes in which the luminance in the cRF was varied (VarLocalLum). To reveal context effects on the representation of luminance at the population level, we developed a between-condition decoding approach. We trained random forest decoders to decode the luminance in the cRFs, from spike counts of all neurons in a session to the VarLocalLum scenes. We then tested these decoders on the FixedLocalLum data. For the 1000 FixedLocalLum scenes with respectively the highest and the lowest decoded luminances, we averaged the pixel value element-wise and subtracted these averages. This revealed the contextual elements that significantly influenced the decoded luminance from a column (permutation test $p < 0.05$). We find that these elements often form non-uniform patterns, similar to reverse correlation patterns that we identified in individual units. Our data suggest net, non-uniform context effects on the representation of local luminance in columnar populations of V4 neurons. More broadly, our method represents a novel approach based on reverse correlation that can be deployed to study context effects in large simultaneously recorded populations.

This research was supported by the National Eye Institute of the National Institutes of Health under Award Numbers R00EY031795 and

R21EY036566, and the Small Grants Program from the McDonnell Center for Systems Neuroscience at Washington University in St. Louis.

33.304 BRIGHTNESS INDUCTION EMBEDS A RECOVERABLE, SYSTEMATIC INFLUENCE OF DISK LUMINANCE IN DISK-ANNULUS STIMULI

Osman Kavcar¹ (benkavcar@gmail.com), Michael E. Rudd^{1,2}, Christabel Arthur¹, Michael A. Crognale^{1,2}; ¹University of Nevada, Reno, ²Center for Integrative Neuroscience

According to Rudd's edge-integration model with contrast gain control, the achromatic appearance of a match disk varies with the target disk's luminance and its contrast polarity relative to the annulus in a simple disk / disk-annulus display. This dependence appears in the coefficients of the achromatic color-appearance matching functions (ACAMFs), obtained by fitting a second-order regression model to the match-disk settings at each sampled annulus luminance. Thus, the brightness-matching behavior in a disk-annulus stimulus should carry a hidden imprint of the target disk luminance under which the match was made, even though that luminance is not explicitly encoded in the match settings themselves. We tested whether this imprint is recoverable from behavior alone by training a smooth forward model exclusively on data collected at a single target disk luminance (T18). For each contrast-polarity condition, a Gaussian Process Regression (GPR) model was fit using annulus luminance, annulus radius, and (in a separate version) participant identity as predictors of log match-disk settings. These models captured the characteristic induction function for T18 and were subsequently frozen as fixed reference manifolds. New datasets collected at T11, T14, T22, and T28 were then passed through the T18 models, and trial-wise residuals were computed and plotted as a function of log annulus luminance. Across all polarity conditions, residual curves corresponding to different target luminances formed orderly, non-overlapping "fans," maintaining a consistent vertical ordering ($T11 < T14 < T22 < T28$) and preserving curvature within polarity. Including participant identity did not alter this structure. These properties demonstrate that target disk luminance systematically shifts the induction function in a stable, recoverable manner. Thus, brightness-matching behavior embeds a discernible imprint of the disk luminance under which observers perform the task, even when the model has no explicit access to that variable.

Research reported in this presentation was supported by National Institute of General Medical Sciences of the National Institutes of Health under grant numbers [P30 GM145646 and/or P20 GM103650]

33.305 THE RESISTANCE OF VERNIER HYPERACUITY TO VARIATIONS IN INTENSITY AND WAVELENGTH.

Yaw Buabeng¹ (yaw.buabeng@uga.edu), Billy Hammond¹; ¹University of Georgia

This study assessed vernier performance across wavelengths and at different broadband intensity levels to determine whether hyperacuity

depends primarily on photopic drive or varies systematically with spectral composition. Twenty healthy young adults (mean age = 24.8 $\pm$ 2.3 years) with normal, corrected vision were tested in an aperture defined vernier task. Narrowband stimuli (420–660 nm) and a broadband intensities were produced by a 1000-W Xenon arc lamp with interference filters. Thresholds, defined as the minimum detectable misalignments between two visual targets, were recorded using a digital micrometer. Vernier thresholds did not vary significantly with wavelength ($F(6,98) = 0.628$, $p = 0.71$). Similarly, vernier thresholds did not vary significantly with the different broadband intensities ($F(5,24) = 0.26$, $p = 0.93$). The stability of vernier acuity across these conditions indicates that hyperacuity performance is independent of spectral composition or photopic drive. This consistency supports the established view that vernier acuity is a hyperacuity mediated primarily by cortical neural mechanisms rather than by retinal sampling or luminance-dependent factors.

33.306 THE EFFECT OF EXPOSURE TIME ON ACHROMATIC AND CHROMATIC MUNKER-WHITE ILLUSION: INTRASUBJECT COMPARISON

Sae Kaneko¹, Koharu Sugano¹; ¹Hokkaido University

The perceived color of a target embedded in achromatic or chromatic gratings is affected by the flanking and/or coaxial bar (Munker, 1970; White, 1979). The chromatic and achromatic versions were reported separately, but here we refer to both effects as Munker-White illusion. The exact mechanisms of the effects, or whether the effects share common mechanisms, are unknown. The effects of exposure time on several brightness/lightness illusions have been reported (e.g., Kaneko & Gilchrist, 2020). However, the reported effect on the Munker-White illusion is unclear (Robinson & de Sa, 2008). In this study, we tested the effect of exposure time on achromatic and chromatic Munker-White illusions to explore the temporal dynamics of the illusions and to examine the possibility of shared mechanisms between the two phenomena. Subjects viewed a stimulus made with paper in a custom-made viewing box through a camera shutter, which regulated each exposure time (5 conditions, 15 ms to 907 ms). The task was to report the closest match of two identical targets embedded in a stimulus using a chart with Munsell chips (achromatic or chromatic). In achromatic versions, the neutral gray targets are either on a black bar or on a white bar of a background grating. In a chromatic version, the blue targets are either on a green bar or on a lavender bar. Results from two experiments showed that the achromatic Munker-White illusion was stronger for shorter exposure times. In contrast, we found no effect of exposure time on the chromatic version of the phenomenon. The strength of two illusions within a subject, both collapsed across exposure time conditions and within each condition, showed no significant correlation. The differences in temporal dynamics and the lack of intrasubject correlation suggest that the achromatic and chromatic Munker-White illusion, however similar they appear, may stem from separate mechanisms.

This work was supported by JSPS KAKENHI Grant number JP24H0070225.

33.307 HOW DOES THE DIMENSIONALITY OF CHROMATIC DISTRIBUTION AFFECT INDIVIDUAL DIFFERENCES IN COLOR PERCEPTION?

Jin Hirano¹, Masataka Sawayama^{2,3}, Kiyofumi Miyoshi¹, Shin'ya Nishida¹; ¹Kyoto University, ²Hokkaido University, ³Prometech CG Research, Japan

Large individual differences in color perception, as observed with #TheDress and #TheShoe, can be attributed to the ambiguity in specifying the surface–illuminant relationship arising from the limited chromatic information in images. One factor that may increase the ambiguity in these images, and thereby increase individual differences, is that they have approximately one-dimensional chromaticity distributions (elongated along the daylight locus or along two complementary colors). According to this idea, making the chromaticity distribution two-dimensional will reduce inter-observer variability in perceived colors. To test this, we psychophysically examined how the dimensionality of the chromaticity distribution affects inter-observer variations in categorical color judgments. We created computer-rendered cube stimuli presented against a black background. Each face of the cube contained a 6×6 grid of tiled surface colors. We manipulated the dimensionality of the stimulus chromaticity distributions by changing the number of surface colors used across the grid tiles. Stimuli with one and two colors formed one-dimensional distributions, while those with four colors and multiple colors exhibited two-dimensional distributions. To systematically vary color appearance, we shifted the chromaticity distribution along a color axis in the CIE L*u*v* space. The experimental procedure followed Hirano et al. (VSS2025). For each participant, we estimated the color switching point at which each participant's perception switched between alternative interpretations (e.g., “white and gold” to “blue and black”). Inter-observer variations in switching points were substantially larger for the two-, four-, and multi-color conditions than for the one-color condition. Contrary to the expectation, the two-, four-, and multi-color conditions yielded comparable levels of variability. The color distribution along the daylight-locus axis tended to show slightly greater variability than that along the other axes. This result suggests that the one-dimensional chromaticity distribution is not a determining factor in explaining inter-observer variability in color perception.

This work was supported by JST SPRING (JPMJSP2110) to JH and JSPS KAKENHI425 (JP20H05957, JP24H00721) to SN.

33.308 THE VISION SCIENCE OF MEDICAL MONITOR CHARACTERISTICS

Nancy O. Mahfouz¹, Aliza Virji¹, Parker Flanagan¹, Jana Radwan¹, Stephen Waite¹, Robert G. Alexander¹; ¹Department of Psychology & Counseling, New York Institute of Technology, USA

Radiologic diagnosis relies heavily on a clinician's ability to notice faint visual cues, yet image interpretation is ultimately limited by both the biology of the human visual system and the quality of the display used to view images. Medical-grade monitors (MGMs) are designed to optimize this final stage of the imaging pipeline, but the vision science principles behind their specifications are not always made explicit. This project highlights the perceptual foundations of ACR–AAPM–SIIM standards for electronic medical imaging displays and explains how these standards support accurate clinical interpretation and decision making. We outline how luminance, contrast sensitivity, spatial resolution, color perception, and visual adaptation shape what radiologists can—and cannot—see on a display. Because human contrast sensitivity is nonlinear, monitors must be calibrated using the DICOM Grayscale Standard Display Function to ensure that small differences in grayscale values remain perceptible. High luminance ratios and stable minimum luminance levels protect low-contrast lesion detection, while pixel pitch and display size must align with the limits of foveal acuity and the useful visual field. Consistent luminance across the full screen is also essential, as peripheral vision plays a key role in guiding radiologists' search patterns. Overall, this project demonstrates that display standards in radiology are grounded in the core fundamental properties of human vision. Understanding these perceptual constraints can help inform the selection of clinical displays and support ongoing efforts to improve diagnostic reliability through vision-science-aligned technology design.

This work was supported by a New York Institute of Technology Teaching and Learning with Technology grant and by the National Institute of Health (Award R16GM159810 to RGA), which provided support for student training and research activities that contributed to this project.

Perceptual Organization: Ensembles

SUNDAY, MAY 17, 8:30 AM – 12:30 PM,

BANYAN BREEZEWAY

33.309 TEMPORAL ENSEMBLE PERCEPTION IN DERMATOLOGICAL JUDGMENTS

Tiffany Tai¹, Ning-Ting Hsu¹, Haley Frey², Katrina A. Wolters², Shao-Min (Sean) Hung³, Po-Jang (Brown) Hsieh¹, David Whitney²; ¹National Taiwan University, ²University of California, Berkeley, ³Waseda Institute for Advanced Study, Waseda University

Dermatologists often scan across many lesions in rapid succession, forming quick impressions that guide which areas require closer examination. Such efficiency suggests that clinicians may depend on temporal ensemble perception—the ability to extract a quick overall sense of a feature from a sequence of stimuli rather than from detailed analysis of each item. Although this mechanism is well established for basic visual features, far less is known about whether observers can

compute meaningful summaries for high-level categories that lack strict, explicit rules, such as skin lesion malignancy. Demonstrating temporal ensemble processing for malignancy would clarify how broad-area screening is achieved under tight time constraints and reveal both the strengths and limits of this strategy. This study examines whether observers form an ensemble malignancy from a brief, sequential stream of skin lesion images and whether their judgments reflect integration across all items rather than reliance on a randomly sampled subset. Participants (N = 5) viewed sequences of 1, 2, 4, or 6 lesions (600 ms total), sampled from image sets with known consensus malignancy ratings. After each sequence, they rated the average malignancy. Observers' ratings in the whole-set condition (set size = 6) strongly correlated with the consensus average malignancy (mean Fisher's $z = 0.68$), indicating reliable extraction of overall malignancy despite extremely limited exposure. Correlations increased systematically with set size, climbing from 1 to 6 images in a pattern well captured by a linear model ($R^2 = 0.93$). This improvement shows that observers integrated information across multiple images rather than basing judgments on a single randomly sampled lesion. These findings provide the first evidence that viewers can perceive average malignancy from rapidly presented skin lesions, suggesting that temporal summary mechanisms may contribute to efficient screening in dermatology and offering a foundation for understanding how such processes might aid—or bias—clinical decision-making.

supported in part by NIH R01CA236793

33.310 NEURAL MARKERS OF PROBABILISTIC PERCEPTION

Anton Lukashovich¹ (anl22@hi.is), Sabrina Hansman-Roth¹; ¹University of Iceland

Previous Feature Distribution Learning (FDL) paradigms demonstrated that observers implicitly represent distractor feature distributions to optimize visual perception and behavior. These findings furthermore, reveal that internal representation are much richer than merely encoding summary statistics. However, the neural mechanisms underlying the encoding of feature probabilities remain unclear. Here, we investigated whether encoded distractor probabilities modulate the N2pc event-related potential (ERP) component, an index of attentional selection, focusing on amplitude differences in a modified FDL task. Observers (N=5 in pilot) viewed arrays of colored diamonds (36) and searched for an oddly-colored target, reporting the location of the target in lower or upper half of the set. During learning trials distractor distributions were kept constant to ensure sufficient learning of the distractor probabilities. Subsequent test trials introduced role reversals: previous distractor features became the target, with its probability varying. EEG data were recorded via a 32-channel system, with epochs time-locked to stimulus onset. Behavioral results validated that observers implicitly encoded the distractor probabilities: search times for targets closer to the mean of the previous distractor distribution were slower than search times

for targets further away from the mean of the previous distractor distribution. Visual inspection of EEG data indicated N2pc amplitude differences tied to distractor probabilities: attenuated amplitudes for high-probability reversals contralaterally, implying stronger suppression, versus larger amplitudes for low-probability ones, suggesting reduced filtering. Latency effects were subtle, meriting expanded testing. These preliminary findings bolster probabilistic encoding in vision, linking the learning of feature probabilities to neural signatures. Ongoing work will quantify latency shifts and scale up sample of participants. Work is supported by the Icelandic Centre for Research (Rannis): 239774-052 to SHR.

33.311 MODELING RESPONSES AND RESPONSE TIMES IN ENSEMBLE PERCEPTION

Jinhyeok Jeong¹ (jinhyeok.jeong@vanderbilt.edu), Giwon Bahg¹, Thomas J. Palmeri¹; ¹Vanderbilt University

Ensemble perception is the ability to summarize a collection of objects in a visual array. Most work in ensemble perception has focused on response probabilities. Response times (RTs) have received little attention despite their importance for fully understanding computational mechanisms. We developed a computational modeling framework that provides a process-level account of both response probabilities and RTs in ensemble perception tasks. Our framework allows systematic tests of competing hypotheses regarding how individual objects are perceptually encoded, how ensembles are represented, and how decisions are made. Decisions are modeled as a similarity-based random walk accumulation of evidence, driven by comparisons between a test object and a stored ensemble representation. Given particular assumptions about perceptual encoding and ensemble representations, alternative models can provide alternative mechanistic accounts of how decisions unfold over time, which naturally predicts both response probabilities and RTs. We focus here on three alternative models differing in assumptions about ensemble representation: prototype, summary statistic (mean and variance), and exemplar models. We first evaluated these models by fitting these to an existing dataset in which participants judged the mean orientation of an ensemble array. All three models accounted well for joint distributions of response probabilities and RTs across 7 test level conditions using a single parameter set, demonstrating the viability of the modeling framework. To further differentiate the models, we conducted a new ensemble mean discrimination experiment that factorially manipulated ensemble variance, set size, and test levels, yielding a total of 42 conditions. Summary statistics and exemplar models predicted the major trends in response probabilities and RTs across all conditions, whereas prototype models failed to capture variance effects. These results support a view that ensemble representations are richer than a single abstract mean and illustrate how RTs as well as response probabilities can be accounted for by models of ensemble perception.

33.312 THE IMPACT OF RECENT SACCADDES & TIMING ON ENSEMBLE PERCEPTION

Inchara Manjunatha^{1,2}, Kevin Ortego³, Viola Störmer³, Sven Ohl¹, Martin Rolfs^{1,2}; ¹Humboldt-Universität zu Berlin, ²Berlin School of Mind and Brain, ³Dartmouth College

People can quickly process summary statistical information across individual features. These ensemble representations have been shown to be biased towards items at the center of gaze suggesting that ongoing eye movements influence the way we perceive ensembles. Here, we examined the time course of this foveal bias following instructed saccades or prolonged fixations. We presented a stimulus consisting of oriented lines for 200 ms and participants reported their mean orientation under two viewing contexts: after an instructed saccade or while maintaining fixation. When saccades were required, the stimulus appeared gaze-contingently with a stimulus onset asynchrony (SOA) of 0, 50, 100, or 200 ms after online saccade detection. In the fixation condition, a temporal cue marked the upcoming stimulus onset using the same SOAs. Out of four quadrants in the ensemble, three contained orientations drawn from the same uniform distribution (standard quadrants), while the remaining quadrant contained orientations drawn from another uniform distribution with a mean shifted by 37.5° relative to the standard quadrant distribution (outlier quadrant). Unbeknownst to the participants, across an equal number of interleaved trials, the standard or the outlier quadrant was at the center of gaze. We expected reports of the ensemble mean to be biased toward the foveated quadrant, and for the magnitude of this bias to decrease with longer SOAs. Indeed, the foveated quadrant biased reports of the ensemble mean indicating that orientations across the visual field are not pooled uniformly. The bias was reduced at the 200 ms SOA relative to 0 ms, but showed no modulation by a recent saccade. Thus, while a longer delay between fixation and ensemble onset attenuated the foveal bias, similar performance across saccade and fixation trials suggests that orientations at the saccade target do not further influence ensemble perception beyond prolonged fixation.

33.313 ENSEMBLE PERCEPTION: WHERE, WHY, AND HOW?

Shaul Hochstein¹, Noam Khayat¹, Marina Pavlovskaya¹, Haya Abtan¹, Daniella Koyfman¹, Merav Ahissar¹; ¹Hebrew University, Jerusalem

Since Dan Ariely (2002) reignited interest in ensemble perception, also known as scene summary statistics perception, hundreds of papers have been published on this subject. It is now well established that ensemble perception includes low and high level parameters, objects, and scenes, and that observers perceive ensemble statistics including average, range, skewness, and distribution. Here we present new findings concerning the broad extent, neural computational process, cerebral site(s), and function of ensemble perception. Surprisingly, we found that ensemble means are represented in mid-level cortical areas specializing in object perception, including centrally the Lateral Occipital Complex, leading to the suggestion that ensembles are perceived as single objects rather than as groups of objects: a forest is not a group of trees, a traffic jam is not a bunch of cars, and a VSS

conference is not a bunch of scientists, lectures and posters. Rather, they are, indeed, a forest, a traffic jam, and a VSS conference. Considering the very limited capacity of short term working memory, such aggregation and redefinition is essential for perception, computation, and memory. Yet, the process of aggregating ensembles without first perceiving each individual needs clarification. I shall show that direct population encoding solves the issue elegantly and simply. In summary, we now have a better handle on answering the three essential questions regarding ensemble perception: Where in the visual brain are these summary statistics computed; why is this important for perception and memory; and, how is this computation performed.

grant from ISF Israel Science Foundation

33.314 NUMBER DISCRIMINATION IS BETTER WHEN COLOR SUBSETS REFLECT OVERALL DIFFERENCE

Ella Beschta Westfall¹ (ebesch1@swarthmore.edu), Berenice Benitez-Carlos¹, Frank Durgin¹; ¹Swarthmore College

The comparison of visual number is often measured using mixtures of black and white dots in two separate patches (to control for luminance). Because mixtures typically contain equal numbers of each color, it is unclear whether observers base their comparisons on a single subset, or on the collection as a whole. We investigated this in two experiments. In an initial experiment, we compared performance on blocks of trials where 50/50 mixtures of black and white dots were used, with performance on blocks where the percentage of black dots in each field was randomly assigned in the range from 30% to 70%. Across a total of 30 participants tested at three numerosities (14, 40, and 114) this manipulation did not significantly affect the precision of number discrimination. However, when we split the variable-percentage trials by the median ratio of the percentages of black dots in the two fields (RoP), we found overall evidence of significantly higher Weber fractions for trials where the RoP differed substantially between the two patches (Cohen's $D = 0.66$). In a second experiment, we directly manipulated the RoP with 30 participants. Weber fractions were higher when there were large differences in RoP, than when the ratios of black and white dots were similar in the two dot fields, $F(1, 29) = 24.2$, $p < .001$, $\eta^2 = 0.14$, Cohen's $D = 0.73$. More importantly, despite that the black and white dots were of equal luminance contrast with the background, re-analysis confirmed that participant responses were strongly biased by the relative numbers of black dots in the patches, but not by the relative number of white dots. This was true across all conditions of the experiment. These observations show that color subsets play a significant role in number comparison tasks even when reliance on subsets impairs performance.

Swarthmore College

33.315 AVERAGE ORIENTATION PERCEPTION EXPLOITS MID-LEVEL GROUPING CUES

Elijah Schwartz¹ (elibschant@gmail.com), Gwen Holley¹, Jason Haberman¹; ¹Rhodes College

Humans readily perceive the summary statistics from an ensemble, whether composed of oriented gabor patches or faces varying in expression. A pervasive question in ensemble perception research centers on understanding the hierarchical level at which statistical representation occurs. Seminal work suggests multiple levels of ensemble representation (e.g., Haberman, Brady, and Alvarez, 2015), but it is unknown whether the level of representation for a particular ensemble may be influenced by the strength of perceptual grouping cues. Here, we leveraged mid-level perceptual structures (i.e., Kanizsa figures/illusory contours) to explore whether implicit grouping cues might confer an advantage in average orientation judgments. Participants viewed either individual Kanizsa triangles or ensembles of four Kanizsa triangles constructed using traditional 'pac-man' shapes (i.e., inducers). These figures are powerful grouping cues, drawing upon dedicated neural mechanisms within area V2 and beyond. When the inducers are rotated, the vertices of the triangles remain but the illusory contours perceptually vanish. We asked participants to adjust a test triangle to match the orientation of the individual or the average orientation of the ensemble, when the inducers were rotated to either give rise to illusory contours or not. Overall performance was significantly worse when illusory contours were not visible. Most striking, however, was the significant interaction between stimulus type (Kanizsa or no Kanizsa) and set type (individual or ensemble) — participants were significantly worse in perceiving the average orientation of ensembles with no illusory contours compared to the orientation of a single triangle with no illusory contours. These results suggest ensemble perception efficiently leverages grouping cues to generate a more precise average representation, and also point to a higher-level of representation than might be expected, given that illusory contours do not emerge until extrastriate regions.

Rhodes College

33.316 NUMBER IS NOT SPECIAL: COMPARING THE RELATIVE SALIENCE OF VARIOUS VISUAL ENSEMBLES

Gabriel C. L. Waterhouse¹, Sami R. Yousif²; ¹University of North Carolina at Chapel Hill, ²The Ohio State University

Imagine looking out at a forest. What might you see? Visual masses of green, the number of trees, the various sizes of the leaves. Is one of these dimensions more salient than the others? According to a popular view, numerical information is uniquely prioritized in perception, being extracted and utilized in preference to other dimensions. This "number bias" is more than a perceptual preference: It has been argued that this proclivity for numerical information is directly responsible for the cultural evolution of mathematical knowledge. Here, we question this orthodoxy. In eleven experiments and six unique paradigms, we demonstrate that multiple other visual ensembles are prioritized over number by both adults and children. We first used a categorization task to contrast number and average size.

Participants categorized displays using abstract symbols, learning each symbol corresponded to either “less” or “more” in both dimensions. When the dimensions were later pitted against each other, participants overwhelmingly categorized based on average size rather than number. In another set of experiments, participants were asked to place a line to divide an 8x8 stimulus space (each dimension corresponding to a different ensemble) into two distinct groups. Across three versions of this task, with various dimensions contrasted, participants decisively failed to use number, instead dividing by average size or cumulative area. We then put this “size bias” to the test, using several unique paradigms to measure the relative salience of average size and number. Across all experiments, we found that both adults and children preferentially matched, distinguished, and rapidly identified stimuli based on average size rather than number. Collectively, these results undermine claims about a universal number bias. While numerical knowledge is indeed a crucial cultural artifact, this knowledge may not be rooted in perceptual biases.

33.317 CAN YOU COUNT THIS? PERCEIVING AFFORDANCES FOR MENTAL ACTIONS

Lana Milman¹ (lmilman1@jh.edu), Ian Phillips¹, Chaz Firestone¹;
¹Johns Hopkins University

We can often see how easy a physical action is to execute, such as climbing stairs or grasping an object. Is the same true for mental actions? Can we also see how easy it is to execute a cognitive operation, before actually carrying it out? Whereas the perception of physical affordances has a long history in cognitive science, the hypothesis that we perceive “mental affordances” (McClelland, 2020) has received less empirical attention. Here, four experiments explore this hypothesis for the mental affordance “countability”—how easily an array of objects can be precisely counted. In Experiments 1-2, subjects briefly (500ms or 2500ms) saw two “cookies” (circles) containing a number of “M&Ms” (dots), varying in size (large/small/mixed), color (single/mixed), opacity (full/partial), and grouping (clustered/dispersed). Subjects selected whichever cookie seemed easier to count (causing it to reappear onscreen), and then counted that cookie’s M&Ms. In Experiments 3-4, subjects counted a single cookie’s M&Ms, without indicating preference. Results showed that even a 500ms preview was sufficient to form strong preferences for countability, with subjects preferring fully opaque to partially opaque M&Ms, clustered M&Ms to dispersed M&Ms, and large M&Ms over mixed and small M&Ms. However, these impressions were only partially accurate. Subjects performed better with opaque M&Ms, and large as opposed to mixed and small M&Ms, in line with their preferences. But performance was better for dispersed M&Ms than clustered M&Ms—contrary to their preference. Moreover, despite expressing no preference for single over mixed-color M&Ms, subjects performed better with single colors. These results suggest that naïve observers can rapidly form impressions of a mental affordance and use it to guide behavior—but that, just like with physical affordances, we may be imperfectly calibrated to our actual capabilities.

33.318 ENSEMBLE SIZE PERCEPTION ACROSS DEPTH IN REAL VS. VIRTUAL ENVIRONMENTS

Katrín Fjólá Aspelund¹, William Gaudreau¹, Burke Zimmer¹, Ömer Dağlar Tanrikulu¹; ¹University of New Hampshire

The present study examined how viewing distance influences single and ensemble size perception in real-world (RW) and virtual reality (VR) environments. Previous work using 2D displays or stereoscopes has shown that ensemble judgments follow perceived size after size constancy, rather than retinal size. Here, we asked how this process operates with more natural stimuli and richer depth cues. In the RW experiment, participants viewed handmade paper cubes and completed two size judgment tasks. In the single condition, one cube of variable size appeared at one of three depth planes (0.5, 1, or 2 m), and participants matched its physical size by adjusting a probe cube at 1 m. In the ensemble condition, eight cubes of varying sizes appeared at a single depth plane, and participants matched their perceived average size. In the VR experiment, different groups performed single and ensemble 2AFC tasks in a virtual supermarket, judging which of two flour packs (or two groups of eight packs) at 2.5 and 5 m contained more flour. Points of subjective equality (PSEs) were estimated with the staircase method. In the RW, observers underestimated near cubes and overestimated far cubes, for both single and ensemble displays, even though far cubes projected smaller retinal images. This pattern replicates earlier screen-based findings and suggests that a biased internal depth map feeds both individual and ensemble size constancy. Ensemble judgments tended to be more accurate than single-object judgments, consistent with improved depth estimates when multiple items are present. In VR, PSEs were again larger for the far plane in both tasks, with slightly larger overestimation for single than ensemble displays. Together, the findings indicate that ensemble size perception relies on the same 3D size–distance machinery as single-object perception and that pooling across items can partially mitigate, but does not eliminate, overconstancy.

Perceptual Organization: Individual differences, aesthetics

**SUNDAY, MAY 17, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY**

33.319 BEAUTY BOOSTS WILLINGNESS TO PAY

Youjing Luo^{1,2} (yj12563@nyu.edu), Anna Bruns¹, Olivia Cheung³, Denis Pelli¹; ¹New York University, ²New York University Abu Dhabi, ³United Arab Emirates University

How valuable is beauty? The link between perceived beauty and monetary value is an underexplored aspect of psychology and behavioral economics. In this study, we investigated this relationship using an incentive compatible image-auction paradigm, an adaptation

of the Becker–DeGroot–Marschak (BDM) procedure that reveals a participant's willingness to pay for an item. Participants ($N = 107$) received \$1, which they used to bid on 16 images. They then rated the beauty of each image on a 0–100 scale and completed questionnaires measuring aesthetic responsiveness (AReA) and risk attitude. We tested i) whether beauty predicts the likelihood of placing a bid, and, ii) conditional on bidding, how beauty affects the bid value. First, using a generalized linear mixed-effects model with random intercepts for participants and images, we found that beauty increases the likelihood of placing a bid. Each 10-point increase in perceived beauty more than doubles the odds of placing a bid. Second, using a linear mixed-effects model on the log-transformed bid values, we found a 12% increase in value for each 10-point increase in beauty. Further moderation analyses showed that individual differences in aesthetic responsiveness and risk attitude moderated the effect of beauty on bidding likelihood, but not on bid value. For people high in aesthetic responsiveness, beauty had a stronger effect on decision to bid. For people high in risk-seeking, beauty had a weaker effect on decision to bid. Our work links beauty to real monetary decisions, and we find that beauty boosts value: People pay more for what they find beautiful.

33.320 THE TEMPORAL DYNAMICS OF ART-STYLE PERCEPTION

Philipp Flieger¹, Daniel Kaiser^{1,2,3,4}; ¹Neural Computation Group, Department of Mathematics and Computer Science, Physics, Geography, Justus Liebig University Giessen, Germany, ²Cluster of Excellence “The Adaptive Mind”, Universities of Giessen, Marburg, and Darmstadt, Germany, ³Center for Mind, Brain and Behavior (CMBB), Universities of Giessen, Marburg, and Darmstadt, Germany, ⁴Center for Applied Computer Science and Data Science (ZAD), Justus Liebig University Giessen, Germany

Looking at a photograph or van Gogh's rendition of a night sky, we have no problem telling apart style from content. Yet, how the brain represents style and content has not been investigated. Here, we created stylized renditions of 20 natural objects and scenes in seven different artistic styles (3D-render, cubism, expressionism, drawing, pixel art, renaissance, watercolor) using Stable Diffusion XL. In an EEG experiment, participants ($N = 25$) viewed the stylized images, and the original photographs used to create them, while performing a simple catch-trial detection task. Representational similarity analysis revealed robust neural representations of both art style and content. As expected, we observed higher pairwise decoding accuracies between visually distinct art styles (e.g., 3D-render vs. expressionism), compared to more similar ones. Moreover, neural representations of art style peaked earlier (~100ms post-stimulus) than representations of content (~150–200ms post-stimulus), suggesting that rapidly emerging perceptual codes differentiate art styles. When regressing art-style and content predictors, respectively, onto the neural data while accounting for hierarchical deep neural network (DNN) features and diverse image statistics (e.g., saturation, symmetry) we found that color features were robustly related to style coding while DNN features

better explained content coding. Accounting for early- to mid-level DNN features abolished early coding of content while leaving style representations at similar time points intact, and removing the variance explained by high-level features abolished early coding of both art style and content. Conversely, controlling for the variance explained by various simple image statistics left content representations largely unaffected. In summary, our results suggest that neural representations of art style and content emerge with distinct temporal dynamics and, crucially, differ in terms of format.

33.321 SCHIZOTYPY MODULATES THE EXTERNAL-INTERNAL NOISE TRADE-OFF IN SYMMETRY-BASED STOCHASTIC RESONANCE

Roberta Cessa^{1,2} (roberta.cessa@studenti.unipd.it), Martina Passaggi¹, Deniz Demirkapı¹, Md Jobayer¹, Nadja Kosovic¹, Esma Dilara Yavuz¹, Luca Battaglini¹, Giulio Contemori¹, Marco Bertamini¹; ¹University of Padua, ²University of Rochester

Visual perception is shaped by noise: external noise (variability in physical stimuli) and internal noise (spontaneous neural fluctuations within the brain's processing systems). Stochastic resonance captures how these two sources interact, showing that adding moderate external noise enhances detectability of subthreshold signals. Internal noise varies across individuals, and people within the schizophrenia spectrum exhibit elevated neural variability. Whether similar alterations are already present in subclinical schizotypy, and whether they influence global perceptual organization, remains unresolved. We addressed this question through two experiments on symmetry perception. In a first threshold-based study, we measured the symmetry signal strength required for reliable detection across individuals, providing an index of perceptual organization. We then examined whether trait differences also influence how the visual system handles noise by testing the stochastic-resonance profile for symmetry detection. In this experiment, 100 participants completed a two-interval forced-choice task in which one random-dot pattern contained vertical symmetry and the other was random. Symmetry coherence was fixed at 65%, while dot density varied across 14 levels to manipulate external noise. All participants completed the O-LIFE and CAPS questionnaires. Global schizotypy scores showed mild positive correlation with symmetry thresholds, indicating reduced perceptual organization as scores increase (global: $r=0.15$, $p<0.001$, UE: $r=0.25$, $p<0.001$). Accuracy produced the expected inverted-U, as captured by a quadratic generalized linear-mixed model. However, higher schizotypy, particularly Unusual Experiences, produced a flatter curve and an earlier peak (global: $d=0.22$, $p=0.017$; UE: $d=0.33$, $p<0.001$), showing that performance improved only under lower noise conditions. Together, the elevated symmetry thresholds and altered stochastic resonance profiles demonstrate that schizotypy, particularly Unusual Experiences, impairs both perceptual organization and the optimal internal-external noise balance that supports it. These findings make symmetry perception a sensitive behavioral tool for discerning

noise-related perceptual anomalies that differentiate symptom dimensions across the schizophrenia spectrum.

This work is part of the “InDi-ViDi” project, funded by the BIAL Foundation Grants Programme 2024/25 (Application No. 389/2024). Roberta Cessa is currently supported by a Fulbright Visiting Researcher Scholarship in the United States.

33.322 GLOBAL AND LOCAL PROCESSING DIFFERENCES WITHIN ENSEMBLE PERCEPTION AS A FUNCTION OF CULTURAL VARIABLES

Gaomong Lo¹ (glo3@csuchico.edu), Chia-huei Tseng², Bo Yeong Won¹; ¹California State University, Chico, ²Tohoku University

Prior research on cultural differences in perception has shown that individuals from Eastern cultures tend to rely more on global processing, whereas those from Western cultures rely more on local processing, reflecting holistic versus analytic cognitive styles (Masuda & Nisbett, 2001; Nisbett et al., 2001). However, most studies have compared cultures at the group level and have rarely examined how individual differences—such as levels of individualism and collectivism or exposure to other cultures—may influence perceptual processing within a culture. Our study explored whether individual differences in collectivistic versus individualistic orientations influence perceptual processing within a culture and whether cultural differences extend to subconscious perceptual abilities, specifically ensemble judgments. To investigate these questions, participants completed a within-subjects size-judgment task under global and local processing conditions and also completed the Racial Contact Questionnaire (Hancock & Rhodes, 2008) and the Individualism and Collectivism Scale (Triandis & Gelfand, 1998). Visual displays were created using randomized circle stimuli adapted from Cha et al. (2022). Based on previous literature, we hypothesized that individuals endorsing more collectivistic values would show higher global processing accuracy, whereas those endorsing more individualistic values would excel in local processing. Contrary to these predictions, the results showed the opposite trend: participants with stronger collectivistic orientations tend to show lower global accuracy and higher local accuracy. Similarly, participants with greater exposure to Asian cultures tend to perform worse on the global task and better on the local task. Although unexpected, these findings suggest that cultural values or exposure alone may not fully account for perceptual differences previously reported. Cultural influences may also manifest differently in conscious versus subconscious perceptual processes. To clarify these patterns, we plan to directly compare Japanese and American participants in future work.

This research was supported by the Student Award for Research and Creativity (SARC-C), awarded to Gaomong Lo by California State University, Chico.

33.323 TOWARDS EXPLORING AND MAPPING INDIVIDUAL DIFFERENCES TO HIERARCHICAL LEVELS OF VISUALIZATION COMPREHENSION

Tapendra Pandey¹ (pandey@ou.edu), Ghulam Jilani Quadri¹; ¹University of Oklahoma

Individual differences in education, profession, visualization literacy, and familiarity shape natural and uncued interpretations of visualization, called high-level comprehension (HLC) (Quadri & Szafir, 2022; Quadri et al., 2023). However, how individual differences influence perception and comprehension of visualization remains unclear. Our empirical study offers a broader perspective on how they can guide and predict HLC. We propose a three-level hierarchical framework categorizing information depth of HLC: Basic Graph Knowledge(BGK; reading graphical elements), Conceptual Graph Understanding(CGU; capturing intended message), and Contextual and Interpretative Insights(CII; combining inference and reasoning), defined based on classifications provided by (Friel et al., 2001; Lundgard & Satyanarayan, 2022) and maps them to individual cognitive profiles. We recruited 20 participants from diverse backgrounds (9 professional domains, 4 education levels, 4 visualization familiarity groups) who completed two studies: (a) miniVLAT test (Pandey & Ottley, 2023) yielding literacy scores (mean = 8.2, SD = 1.7), and (b) main study focused on eliciting their understanding across 16 visualizations (visualization(4)× data(2)). Using open coding, two independent coders categorized 320 responses from (b) into three levels, ensuring high inter-rater reliability, revealing 15% BGK, 68% CGU, and 17% CII level responses. We calculated weighted comprehension-level scores, with 5%, 65%, and 30% of participants dominating at each level. Given the diversity of responses, participants reported nine statistical task results (Amar & Stasko, 2005) and often added more. High school students with frequent exposure to visualizations mapped to the same levels as students with STEM advanced degrees. Similarly, participants with identical literacy test scores showed varying levels of HLC. Individual differences showed complex interactions, and findings suggest standard literacy tests are insufficient to determine comprehension levels; i.e., interactions among all four shapes comprehension. Our resulting framework enables visualization designers to create visualizations that account for diverse cognitive profiles that emerge from a combination of individual differences.

Scene Perception: Neural mechanisms

**SUNDAY, MAY 17, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY**

33.324 INTEGRATION PROCESSES IN SCENE PERCEPTION: AN EEG FREQUENCY-TAGGING STUDY

Ahmet Ali Yaman¹ (ahmet.yaman@sabanciuniv.edu), Daniel Kaiser^{2,3,4,5}, Nihan Alp¹; ¹Sabanci University, Istanbul, ²Neural

Computation Group, Justus Liebig University Giessen, ³Center for Applied Computer Science and Data Science (ZAD), Justus Liebig University Giessen, ⁴Center for Mind, Brain and Behavior (CMBB), Universities of Giessen, Marburg, and Darmstadt, ⁵Cluster of Excellence “The Adaptive Mind”, Universities of Giessen, Marburg, and Darmstadt

Understanding how humans perceive scenes is essential for uncovering how the visual system interprets our surroundings. Prior works suggest that coherent scenes are processed more efficiently than distorted ones because their typical structure and content guide top-down predictions. Here, we aim to establish an objective neural marker that quantifies the integration of scene parts into a coherent whole. We utilize an EEG frequency-tagging paradigm to measure Steady-State Visual Evoked Potentials (SSVEPs). Twenty-five participants viewed 10-second flickering scene images. Images were presented in a 2×2 grid, with the top-left/bottom-right and top-right/bottom-left segments, flickering at different fundamental frequencies (6 Hz and 7.5 Hz). Critically, we manipulated the composition of the scene by re-combining parts from different scenes across the top-left/bottom-right and top-right/bottom-left segments. Scene parts could be (in)congruent in their semantic content (stemming from the same or different categories) or spatial structure (showing at their typical or atypical positions in the grid). We then analyzed fundamental frequencies, reflecting the processing of individual parts of a scene, and intermodulation (IM; e.g., 1.5 Hz) frequencies, reflecting nonlinear components that emerge when the brain integrates individual parts into coherently organized scenes. Our analysis focused on Signal-to-Noise Ratios (SNRs) of fundamental and IM frequencies from posterior-occipital electrodes. SNR values at fundamental frequencies did not differ across conditions, indicating that part-based computation was similar across conditions. In contrast, IM frequencies revealed a significant difference between conditions. Intact scenes elicited significantly stronger SNRs than all distorted variations except those with only categorical content distortion. Notably, the strongest decrease in IM frequency SNRs occurred in the spatially distorted conditions. These results suggest that the IM specifically reflects the binding of scene parts, a process heavily constrained by spatial structure. More generally, our findings demonstrate that SSVEP-based IM frequencies provide a robust and objective marker of scene integration.

N.A. and A.A.Y. are supported by the Scientific and Technological Research Council of Turkey (TUBITAK 1001) under the Grant Number 122K922. D.K. is supported by an ERC Starting Grant (PEP, ERC-2022-STG 101076057) and by the DFG (EXC 3066/1 “The Adaptive Mind”, project number 533717223).

33.325 MOMENTS-EEG: A LARGE-SCALE EEG DATASET OF NATURALISTIC AUDIOVISUAL EVENT PERCEPTION

Zitong Lu¹, Shuning Tang², Dongwei Li²; ¹Massachusetts Institute of Technology, ²Beijing Normal University

Understanding how the human brain processes dynamic audiovisual events requires datasets that combine naturalistic stimulation, high temporal resolution, and systematic experimental manipulations. Here we introduce Moments-EEG, the first large-scale, high temporal resolution human EEG dataset of naturalistic video events. Across participants, we collected 64-channel scalp EEG during free-viewing of short (3-second) naturalistic video clips drawn from the Audiovisual Moments in Time dataset. The stimulus set comprised 896 training videos (16 event types × 56 video per type) presented three times, and to a test set of 64 videos (16 types × 4 videos per type) repeated 40 times under three conditions: (1) intact audiovisual videos; (2) intact visuals paired with scrambled meaningless audio, and (3) scrambled meaningless videos with intact audio. Each training clip was repeated three times and each test clip was repeated 40 times, yielding 10,368 trials per participant. Moments-EEG complements existing short-video fMRI datasets by capturing neural dynamics of audiovisual event perception at millisecond precision across the whole brain. Moments-EEG enables several key scientific advances. First, this dataset serves as a benchmark for video-to-EEG encoding and EEG-to-video decoding models of complex events, supporting rigorous evaluation of how computational features align with neural dynamics. Second, the high trial counts for naturalistic audiovisual events allow precise characterization of the temporal evolution of information processing in dynamic scenes. Third, the audiovisual manipulations isolate modality-specific processing. Comparing intact versus scrambled conditions disentangles the individually and jointly contributions of auditory and visual inputs. Forth, Moments-EEG provides a testbed for evaluating the temporal alignment between artificial neural networks and the human brain under stimulation with rich annotations. Overall, by combining naturalistic stimuli, multisensory perturbations, repeated-measurements, and behavioral annotations, Moments-EEG offers a rich resource for study human event perception, multisensory integration, and advancing temporally resolved brain-model alignments.

National Natural Science Foundation of China (32400863)

33.326 EARLY TASK-SPECIFIC SCENE SEMANTICS REVEALED THROUGH ERP ENCODING

Vivian Gao¹, Gillian Rosenberg¹, Bruce Hansen², Michelle Greene¹; ¹Barnard College, ²Colgate University

Visual scene categorization is rapid and automatic, yet it is not the endpoint of understanding. Scene meaning is contextual and depends on an observer’s goals (a realtor and a burglar may “see” the same home differently). Here, we employ EEG encoding with variance partitioning to map the time course of goal-dependent scene semantics. Observers provided verbal descriptions of 40 scenes according to one of four prompts: affordances (what can you do?), affect (what would you feel?), mental simulation (what happens next?), or multisensory experience (what would you hear, smell, or taste?). All descriptions were encoded using MPNet (Song et al., 2020). We computed shared scene semantics as the averaged embedding for all

descriptions for each image across tasks. Task-specific representations were then computed as the residual between each task's embedding and the global mean. In our experiment, 128-channel EEG was recorded while observers (N=30) viewed scenes and performed an orthogonal border-color change detection task. We first predicted brain activity from shared image features; then, we predicted the residual brain activity from task-specific features. Thus, unique R^2 reflects only the additional variance explained by task-specific semantic framing. Overall, the affordances and mental simulation features were reflected earlier in the epoch (111 - 144 ms) and explained over twice as much unique variance as the affective and multisensory features, consistent with affordance and mental simulation's importance for scene understanding (Greene et al., 2016; Battaglia et al., 2013). Notably, we observed task-specific semantic activity for all prompts before 200 ms, even though several required inference (emotion, prediction, multisensory experience) rather than purely perceptual processing, implying an earlier emergence of task-specific semantics than suggested by previous literature (e.g. Kutas & Federmeier, 2011).

NSF 2522311

33.327 FAMILIARITY SYNCHRONIZES BRAIN RHYTHMS DURING NATURAL VIEWING

Kira N Noad¹, Daniel Kaiser^{1,2,3,4}; ¹Neural Computation Group, Department of Mathematics and Computer Science, Physics, Geography, Justus Liebig University Giessen, 35392 Giessen, Germany, ²Center for Applied Computer Science and Data Science (ZAD), Justus Liebig University Giessen, 35392 Giessen, Germany, ³Center for Mind, Brain and Behavior (CMBB), Philipps University Marburg, Justus Liebig University Giessen, and Technical University Darmstadt, 35032 Marburg, Germany, ⁴Cluster of Excellence "The Adaptive Mind", Justus Liebig University Giessen, Philipps University Marburg, and Technical University Darmstadt, 35394 Giessen, Germany

Naturalistic stimuli, such as movies, evoke reliable neural responses across individuals in fMRI. These shared neural dynamics vary with experience, with individuals familiar with a movie sharing more similar fMRI response dynamics. However, it remains unclear how time-varying rhythmic brain signals synchronise between individuals during natural viewing, and whether these brain rhythms are modulated by familiarity. To address this, we used a natural viewing paradigm in which participants viewed clips from the TV show Game of Thrones (GoT) while brain responses were recorded using EEG. Time-varying power in each frequency band (alpha, beta, delta, gamma, theta) across the whole GoT movie was compared between participants using intersubject correlation analysis. There was significant similarity between participants in all EEG frequency bands, demonstrating that naturalistic viewing aligns brain rhythms across individuals. Next, we explored whether familiarity modulated brain rhythms by comparing the intersubject correlations in participants who were familiar with GoT to participants who were unfamiliar with GoT. We found that the

familiar group exhibited greater intersubject synchronisation in time-varying beta rhythms, compared to the unfamiliar group. Finally, we explored whether the time course of these rhythmic responses related to the time course of BOLD fMRI responses during viewing of the same GoT movie. Here, we found correlations for familiar participants between beta power and BOLD responses in a network of brain regions previously found to be involved in coding familiarity. No comparable associations were found for unfamiliar participants. Together, these findings demonstrate that brain rhythms synchronise across individuals during natural viewing. They further suggest that familiarity enhances representations across a network of familiarity-coding regions through synchronized beta-frequency dynamics.

33.328 HOW DOES THE HUMAN BRAIN SELECT AND COMBINE FEATURES OVER TIME TO SUPPORT GOAL-ORIENTED BEHAVIOR?

Bruce Hansen¹, Michelle Greene², Audrey Kris¹; ¹Colgate University, ²Barnard College, Columbia University

Vision supports multiple behavioral goals, so the neural code must flexibly weight different features over time. The visual system selects a cascade of features for scene categorization (Greene & Hansen, 2020), and task demands alter the prioritization of those features (Hansen et al., 2025). However, when and how goal-relevant features are spatially selected remains an open question. To address that, we built a novel brain-guided convolutional neural network (CNN) with dedicated channels that were guided by the neural variance explained by each of eight features. Neural data (via 128-channel EEG) were collected from participants who viewed 78 scenes while performing either a path navigation task or a seating location task. The output layer was designed to classify each input image according to the observer's task while viewing it. To successfully predict an observer's task, the CNN combined image features with task-specific neural data. In each convolutional layer, eight independent sets of nodes were added for dedicated processing of features ranging from low-level wavelet and texture models to intermediate object and affordance models to high-level task-specific sentence embeddings. The network achieved 89.7% accuracy in task prediction. Using deconvolution, we projected feature-specific activations onto image locations to estimate each feature's regional contribution by task. We compared feature-specific activation maps with the behavioral maps, in which participants indicated the relevant regions for walking and sitting in each image. Later network layers (neural data > 250ms post-stimulus onset) best aligned with behavior. Critically, task-specific features contributed most: the navigation task was best captured by object-based features, sentence embeddings from a navigation task, and general description sentence embeddings, while texture and affordance features, combined with sentence embeddings from a sitting task, best disentangled the feature space when participants performed the sitting task. These findings demonstrate goal-dependent, time-evolving neural representations spatially reweight features to support behavior.

James S. McDonnell Foundation grant (220020430) to BCH; National Science Foundation grant (2522311/2) to MRG and BCH.

33.329 EARLY DEVELOPMENT OF LEFT/RIGHT INFORMATION IN THE SUPERIOR PARIETAL LOBULE

Rebecca J. Rennert¹, Daniel D. Dilks¹; ¹Emory University

Young children can navigate their immediately visible surroundings years before the occipital place area (OPA) – a cortical region that supports “visually-guided navigation” in adults – becomes functional. But how? Here we hypothesize that the superior parietal lobule (SPL), a newly identified cortical region also implicated in visually-guided navigation in adults, develops earlier and may therefore support these early navigation abilities. Using functional magnetic resonance imaging (fMRI) adaptation in 5-year-old children, we tested whether SPL – unlike OPA – already represents left/right information critical for visually-guided navigation. Indeed, SPL represented this information, indicating that it develops earlier and may scaffold visually-guided navigation before OPA. These findings suggest a developmental cascade across parietal regions supporting visually-guided navigation and raise new questions about how these regions influence each other as the brain matures.

33.330 DELAYED V1 CONNECTIVITY EXPLAINS THE LATE DEVELOPMENT OF OPA

Yaelan Jung¹ (yaelan.jung@emory.edu), Rebecca Rennert², Hee Kyung Yoon³, Daniel Dilks⁴; ¹Emory University

Recent pediatric neuroimaging work shows that the occipital place area (OPA) supports visually guided navigation only around age 8. Why so late? Here, we tested whether delayed connectivity between OPA and primary visual cortex (V1) explains this trajectory. To do so, we measured OPA–V1 connectivity in 5- and 8-year-olds and adults, dividing V1 along its posterior–anterior axis to isolate foveal to peripheral representations. Using resting-state fMRI (in children and adults) and diffusion-weighted imaging (in adults), we found that adult OPA is selectively connected to peripheral V1 – unlike the parahippocampal place area (PPA), and retrosplenial complex (RSC), which are not involved in visually-guided navigation – consistent with behavioral evidence that visually-guided navigation relies most heavily on peripheral vision. In children, this OPA–peripheral V1 connectivity was absent at age 5 and emerged only by age 8, paralleling OPA’s prolonged functional development. By contrast, PPA and RSC already showed adult-like V1 connectivity by age 5, consistent with their earlier functional development. Together, these findings suggest that developing V1 connectivity scaffolds OPA’s late-emerging function, shedding light on how functional specialization arises in the developing brain.

33.331 PLACE MEMORY AREAS TRACK PREDICTIONS ACROSS VIEWS IN REAL-WORLD SCENES

Anna Mynick¹, Adam Steel², Caroline Robertson¹; ¹Dartmouth College, ²University of Illinois Urbana-Champaign

Memory facilitates seamless perception in immersive environments (Mynick et al., 2025), setting up predictions of upcoming scene views that lie beyond our immediate field of view. Yet, how this process of memory-based scene prediction occurs in the brain is unknown. Here, we asked how scene predictions and their consequences are represented across the scene processing hierarchy: (i) the occipital pole, (ii) the scene perception areas (“SPAs”: PPA, OPA), (iii) the place memory areas (“PMAs”: VPMA, LPMA) (Steel et al., 2021), which sit immediately anterior to each of the SPAs, (iv) the medial place area (MPA), which is both mnemonic and perceptual (Epstein et al., 2007), and (v) the hippocampus. Participants (N=16) first learned an immersive scene in head-mounted VR. On each trial of the subsequent fMRI task, we presented a sequence of three images depicting 30° views from the studied scene. During the third view (“outcome”), participants reported whether the sequence represented a spatially accurate sweep across the environment. On valid trials (57%), all three scene views were consecutive and adjacent, allowing the outcome to be predicted. On invalid and omission trials, the first two views were consecutive and adjacent, but the predicted outcome view was replaced by either an incorrect view (invalid trials), or a blank grey square (omission trials), violating the prediction. We observed robust evidence for predictions and violation responses across the scene processing hierarchy, with different patterns for perceptual (SPAs) and mnemonic areas (PMAs). Specifically, mnemonic areas (hippocampus, MPA, and the PMAs) showed representations of predicted-but-omitted scene views (all $p < .05$), while the SPAs and occipital pole did not. Further, the MPA and the PMAs showed evidence for prediction errors, with greater univariate activity for invalid versus valid trials ($p < .01$). Taken together, these results suggest that the hippocampus and PMAs support memory-based scene predictions in real-world places.

33.332 OBJECT AND SCENE CONTRIBUTIONS TO NEURAL REPRESENTATIONS IN NATURAL IMAGES

Suhyun Kim¹, Hojin Jang¹; ¹Korea University, Seoul, South Korea

Object and scene understanding are fundamental to human vision. Substantial evidence from cognitive neuroscience shows that objects and scenes are processed in partially distinct regions of the visual cortex, suggesting that they rely on different underlying representational mechanisms. Yet real-world environments always contain objects embedded within scenes. Growing work reveals systematic, bidirectional influences between scene context and object processing, hinting at compositional integration rather than strict modularity. In this exploratory study, we investigate how object-derived and scene-derived information jointly contribute to neural representations of complex natural images. Leveraging the large-scale Natural Scenes Dataset (NSD), we investigate whether brain activity patterns are better accounted for by an object-based representational strategy, defined by constituent objects and their relational structure, or by a gist-based strategy that rapidly encodes global layout and scene-level structure. Applying representational similarity analysis

across high-level visual cortex with each of the theoretical models, we find that object-based features more strongly reflect the response structure in object-selective regions (LOC, FFA, EBA), whereas both strategies perform comparably in scene-selective regions (PPA, OPA), suggesting complementary contributions to scene understanding. We further compare human neural representational structure to convolutional neural networks sharing an identical architecture but trained to perform either object or scene categorization. Interestingly, scene-trained models exhibit lower similarity to representations derived from the gist-based strategy and higher similarity to those derived from the object-based strategy, relative to object-trained models. Together, these preliminary findings suggest that scene-selective cortical regions encode natural scenes by integrating object structure and global layout. Furthermore, scene-trained convolutional networks, despite their optimization for scene categorization, appear to rely on object-based information for scene representation.

This work was supported by the National Research Foundation of Korea grant funded by the Korea government (RS-2024-00451866).

33.333 OBJECT-BASED INFORMATION PREDICTS DELAYED NEURAL DECODING OF SCENES

Sundari Ruth¹, Gillian Rosenberg¹, Michelle Greene¹; ¹Barnard College

Although scene understanding is rapid and possibly automatic, there is also substantial evidence that some scenes are processed faster than others (Greene et al., 2015; Caddigan et al., 2017). This work examines the hypothesis that scenes with higher information content are slower to process. Prior work shows that visually information-rich scenes are detected and categorized more slowly and less accurately than visually simpler scenes. In contrast, semantically rich scenes show an early detection advantage but a later categorization disadvantage (Aronson et al., 2025). Here, we examine the role of object-based information, as objects contain both visual and semantic information. We began by comprehensively labeling all objects in each of 67,000 scenes using multimodal algorithms. To reduce labeling noise, we used two pipelines: one combined the Segment Anything Model (SAM, Kirillov et al., 2023) with GroundingDINO (Liu et al., 2024), and the other used Detectron2 (Kirillov et al., 2020). Both pipelines output the proportion of pixel area encompassed by each of ~1300 objects in each scene. We computed the object-based entropy for both pipelines and averaged the results. We selected the 20 highest-entropy and the 20 lowest-entropy images for our EEG experiment. Each image was shown to participants 30 times in a randomized order for 500 ms, while observers performed a border color change detection task and recorded a 128-channel EEG. We conducted a whole-brain decoding analysis using a linear support vector machine to assess the ability to read out image identity at each time point and across conditions. Strikingly, we found that scenes with low object information were decoded very early (maximum accuracy: 112 ms post-stimulus onset), while high object information scenes were decoded at a significant delay (maximum accuracy: 284 ms).

This suggests that greater object variety is associated with slower scene processing.

NSF CAREER 2240815

33.334 A 7T FMRI DATASET OF SYNTHETIC IMAGES FOR OUT-OF-DISTRIBUTION MODELING OF VISION

Kendrick Kay¹, Radoslaw M. Cichy², Thomas Naselaris¹, Alessandro T. Gifford²; ¹University of Minnesota, ²Freie Universität Berlin

Large-scale visual neural datasets such as the Natural Scenes Dataset (NSD) are boosting computational neuroscience research by enabling models of the brain with performances beyond what was possible just a decade ago. However, because the stimuli of these datasets typically live within a common naturalistic visual distribution, they do not allow for strict out-of-distribution (OOD) generalization tests which are crucial for the development of more robust models. Here, we address this limitation by releasing NSD-synthetic, a dataset consisting of 7T fMRI responses from the same eight NSD participants for 284 carefully controlled synthetic images. We show that NSD-synthetic's fMRI responses reliably encode stimulus-related information and are OOD with respect to NSD. Furthermore, we provide a proof of principle that OOD generalization tests on NSD-synthetic reveal differences between models of the brain that are not detected with the original NSD data; we demonstrate that the degree of OOD (quantified as the distance between a set of responses and the training data used for modeling) is predictive of the magnitude of model failures; and we show that less strict OOD generalization tests can be usefully applied even within the domain of naturalistic stimuli. These results showcase how NSD-synthetic enables OOD generalization tests that facilitate the development of more robust models of visual processing and the formulation of more accurate theories of human vision.

Collection of the NSD dataset was supported by NSF IIS-1822683 and NSF IIS-1822929. This work was supported by NIH grant R01EY034118, German Research Council (DFG) grants (CI 241/1-3, CI 241/1-7, INST 272/297-1), and European Research Council (ERC) starting grant (ERC-StG-2018-803370).

Scene Perception: Spatiotemporal factors

**SUNDAY, MAY 17, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY**

33.335 WHEN DO REAL ROAD HAZARDS BECOME PREDICTABLE? DRIVER PREDICTIONS OF HAZARD LOCATION IN DYNAMIC ROAD SCENES

Jiali Song¹, Benjamin Wolfe¹; ¹University of Toronto Mississauga

Prediction is a fundamental visual process and is essential for safe driving. Successfully avoiding a moose on the road requires knowing where it is and where it's going. Several theories of road hazard perception propose that avoiding collisions requires accurately predicting them several seconds in advance. We tested this claim and examined the extent of agreement among drivers. Twenty-four licensed drivers (mean age 19.7 years, and >1 year of experience) viewed 360ms excerpts from 174 dashcam videos that each contained a collision or near-collision. Four video excerpts were created from each video to manipulate excerpt time relative to hazard onset (moment of the first visible indication of an impending collision). Excerpt time was manipulated in four sequential blocks (negative times represent time before hazard onset): -2000ms, -500ms, 0ms, and just before the driver in the video responded. After viewing each excerpt, participants were asked to report the location of the most likely hazard in the scene by clicking in an onscreen rectangle where the video had been shown. Prediction accuracy was defined as the proportion of clicks that corresponded with the location of the collision. Prediction accuracy increased monotonically with later excerpt time, with the largest increase between -2000ms and -500ms, suggesting that this temporal window is key for accurate predictions. We indexed agreement for each excerpt using average pair-wise distance between all clicks, which decreased monotonically, with the largest decrease between -2000ms and -500ms. Click locations were moderately correlated with fixation locations collected previously in a separate sample of 30 drivers who watched the same videos (mean Pearson's $r = 0.52$), suggesting that drivers look at locations that are predicted to contain hazards. These results demonstrate that on-road predictions rely on information that appears 2000ms - 500ms before hazard onset, are moderately consistent across drivers, and guide search strategy.

This work was funded by a National Science and Engineering Research Council grant awarded to Benjamin Wolfe.

33.336 SUDDEN VERSUS GRADUAL SCENE ONSETS YIELD DISTINCT EEG DECODING DYNAMICS

Ilker Duyamaz¹, Micha Engeser^{1,2}, Daniel Kaiser^{1,2,3,4}, ¹Justus Liebig University Giessen, Germany, ²Center for Mind, Brain and Behavior (CMBB), Universities of Giessen, Marburg and Darmstadt, ³Center for Applied Computer Science and Data Science (ZAD), Justus Liebig University Giessen, Marburg and Darmstadt, ⁴Cluster of Excellence "The Adaptive Mind", Universities of Giessen, Marburg and Darmstadt

Multivariate analyses of M/EEG data commonly rely on aligning neural signals to stimulus onsets. Such approaches typically show high decoding accuracy during the transient onset response (within the first 500 ms), which then drops to a lower, more sustained level. In this study, we investigated time-resolved EEG decoding for natural scene processing under conditions where scenes gradually enter the visual field without a discrete onset. To this end, we generated video sequences in which one scene category (e.g., a beach) smoothly transformed into a different category (e.g., a forest). This was implemented by blending images from two categories into a composite

and moving a square aperture across it so that the visible content continuously shifted from one scene to the other. When decoding the category of the first and second scene in these videos from the EEG signal, we observed a striking contrast between decoding for the first and second scene. For the first scene, we found robust decoding peaking around 70 ms after onset with a discernible peak structure. In contrast, the second scene category yielded much lower decoding with no clear peak structure. Aligning the emergence of category-diagnostic information for the second scene using deep neural network features did not improve decoding or recover a peak structure. Classifiers trained on the first scene generalized to the second, but with a broad and diffuse temporal pattern, suggesting the second scene did not follow the hierarchical temporal cascade evident for the first scene. These results show that sudden versus gradual onsets produce distinct temporal dynamics, implying that results from onset-based decoding studies do not readily generalize to free-flowing natural stimuli.

This work was supported by the European Research Council (ERC) Starting Grant (PEP, ERC-2022-STG-101076057), under Germany's Excellence Strategy (EXC 3066/1, "The Adaptive Mind", project no. 533717223), and by the Deutsche Forschungsgemeinschaft (DFG) grants KA4683/5-1 and KA4683/6-1.

33.337 WHEN THE HEADS MATTER- HEAD AND GAZE EFFECTS ON ATTENTION IN REAL-WORLD SCENES

Carmela Gottesman¹, Meagan Ziemann¹; ¹University of South Carolina Salkehatchie

Research on the effects of gaze on spatial attention distribution typically use isolated faces or eyes without a background scene. Some previous research, using pictures of people in real-world scenes showed that the orientation of people in those scenes affects viewers' attention distribution. However, the body orientation to the left of the right was sufficient to facilitate a search task in that direction. Head or gaze orientation didn't make a difference unless the search target was in the line of sight. In this study, we examined if the same search would be affected by the face orientation if the body was not shown in the pictures. Disembodied heads were presented on the same real-world scenes used before. Heads of the same people were presented either facing forward or to the side- gazing to the right or to the left. Each head was presented on the same scene they were in the previous study, but they were enlarged to fill more of the picture space. Participants were tasked with finding a probe- small cross superimposed either on the right or on the left part of the scene. Participants were faster at detecting the probe when the head was oriented towards the relevant part of the scene than when it was oriented in the opposite direction. When the head was oriented forward there was no significant difference in search times to the right or the left and there was no overall difference in response time to forward versus sideways facing heads. Thus, we found that with the body absent, the head took on a more important attention directing role in a real-world scene, replicating results for gaze effects without scenes

present. Taken together these studies reinforce the importance of body orientation, although head orientation can make a difference as well.

33.338 EVALUATING SPATIO-TEMPORAL FUSION OF EEG AND FMRI WITH IEEG

Peter Brotherwood¹ (peter.brotherwood@umontreal.ca), Dora Hermes², Jacob S. Prince³, Kendrick Kay⁴, Frederic Gosselin¹, Ian Charest¹; ¹Université de Montréal, ²Mayo Clinic, ³Harvard University, ⁴University of Minnesota

Understanding how visual information unfolds across space and time in the human brain is limited by the fact that no single neuroimaging method provides precise spatial and temporal resolution. To overcome this, we recently introduced a data-driven method that learns how moment-to-moment EEG activity maps onto fMRI response patterns (Charest et al., 2025). Using the Natural Scenes Dataset (NSD; Allen et al., 2022) and the NSD-EEG (Brotherwood et al., 2024), we trained fractional-ridge regression (Rokem & Kay, 2020) models that, at each EEG time point, learn a set of linear weights predicting vertex-wise fMRI beta patterns from electrode amplitudes in response to the same images. The resulting time series of prediction performance recovers the known spatio-temporal hierarchy of the visual cortex: prediction performance in early visual regions (V1/V2) peaks first at ~60–80 ms, while in high-level visual regions performance peaks around 300–400 ms. To assess the accuracy of fusion-derived cortical activity estimates we compared it with broadband intracranial EEG (iEEG) responses recorded in participants viewing the same NSD images (Huang et al., 2024). We averaged both datasets across stimuli to focus on visual responsiveness, and correlated each electrode's broadband time course with the predicted time course of every cortical vertex. Correlation values decreased systematically with geodesic distance: electrodes nearest a vertex showed the strongest similarity to that vertex's predicted time course (peak vertex in LOC: $R \approx 0.8$), and similarity declined with distance (slope ≈ -0.54). This spatial fall-off indicates that the fusion successfully identifies the local cortical generators most aligned with the iEEG measurements. These findings suggest that our novel EEG-fMRI fusion offers anatomically precise, millisecond-resolved cortical activity estimates whose spatial organisation is now validated directly against intracranial electrophysiology. This opens a new avenue for studying the dynamics of visual processing across the human cortex, overcoming longstanding spatial–temporal trade-offs.

Undergraduate Just-In-Time 2

SUNDAY, MAY 17, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY

33.339 INVESTIGATING ACTIVE SUPPRESSION OF HIGH-CALORIE FOOD STIMULI IN INDIVIDUALS WITH ANOREXIC TENDENCIES

Yixuan Yin¹ (yixuan.yin@stonybrook.edu), Abe Leite, Gregory J. Zelinsky; ¹Stony Brook University

Attention bias in anorexia nervosa has yielded contradictory findings, with both avoidance of high-calorie foods (e.g., Horndasch et al., 2020) and heightened attentional capture (e.g., Neimeijer et al., 2017). This inconsistency reflects an unclear underlying cognitive mechanism. Drawing on Gaspelin et al.'s (2015) Signal Suppression Hypothesis, that salient distractors are actively suppressed below baseline, this project examines whether individuals who scored higher on the Eating Disorder Examination Questionnaire (EDE-Q) actively suppress attention to high-calorie food cues, potentially explaining previously mixed findings. Participants completed a dual-task paradigm combining visual search trials with probe trials. On each trial, six stimuli were presented in a ring. In Block 1, replicating the classic signal suppression effect, participants searched for a predefined target (green diamond) and indicated whether a black dot on the target appeared on the left or the right. Probe trials required recalling letters briefly presented on each stimulus. In Block 2, targets varied across trials, and stimuli were objects. The critical manipulation in this block was the identity of each object: each trial included one high-calorie food, two low-calorie foods, and three neutral objects. The search target was always a neutral object. The dependent measure was probe recall accuracy, which allowed us to measure the suppression of attention to high-calorie food items below baseline. Data collection is ongoing (current $N = 17$). Although the current sample is underpowered for definitive statistical inference, preliminary analyses reveal a large linear trend that higher EDE-Q scores predict reduced probe recall accuracy for high-calorie food stimuli relative to neutral stimuli ($R^2 = 0.23$, $p = .054$). This pattern suggests that individuals with greater anorexic tendencies exhibit selectively suppressed processing of high-calorie cues. Together, these findings provide preliminary evidence for active attentional suppression of high-calorie food cues as an underlying mechanism explaining previously mixed results.

33.340 DEPRESSIVE TENDENCIES CORRELATE WITH BETTER MEMORY FOR NON-SOCIAL SCENES

Donna Dean¹, Abe Leite, Gregory Zelinsky; ¹Stony Brook University

How is a person's ability to understand and remember non-social scenes influenced by their psychological traits? In the initial encoding session, 44 participants viewed 150 non-social scenes (from a 300-image subset of the Visual Genome dataset) for a controlled number of fixations (1-10) and were prompted to describe what they saw in as much detail as they could. In the retrieval session (2-21 days later), participants viewed a partially overlapping set of 150 scenes for 250ms each, indicating which scenes they remembered with a button-press following each trial. Finally, they completed a battery of psychometric questionnaires including IPIP (personality), AQ (autistic traits), PHQ-9 (depression), and GAD-7 (anxiety). We performed a forward stepwise

regression analysis to assess which traits significantly predicted performance. During encoding, participants who scored high on the AQ attention-to-detail subscale ($p < .005$) and IPIP agreeableness ($p < .05$) mentioned more nouns in their descriptions, indicating finer-grained scene understanding. During the retrieval session, participants who had stronger depressive tendencies ($r = 0.51$, $p < .005$) and IPIP openness-to-experience ($p < .05$) had better memory performance (hits - false alarms) during the scene recall task. The openness-to-experience result was expected since prior research shows that openness can lead to improved long-term memory performance in older adults (Stephan et al., 2020). The depression finding is more surprising. It is known that individuals who meet the DSM-5 criteria for Major Depressive Disorder show decreased memory for positive events, intensified memory for negative events, and impaired memory recollection overall (Dillon & Pizzagalli, 2018); does this mean that the neutral non-social scenes in our dataset are processed differently than the positive and negative stimuli used in prior work? In future work, we will correlate these measures with the eye-tracking data in the encoding session to seek a mechanistic explanation for these intriguing results.

Stony Brook University

33.341 WALKING TOWARDS THE UNKNOWN: VISUAL EVENT SEGMENTATION SHAPES CREATIVE THOUGHT

Grace Lesko¹, Joan Danielle K. Ongchoco¹; ¹University of British Columbia

Of all the cognitive processes, creative thinking seems maximally different from visual processing. Perception is encapsulated and constrained to the available structure in the environment, whereas creativity takes input (and inspiration) from everywhere, varies across individuals, and seems unconstrained—i.e., we cannot predict when an insight will come. Here we demonstrate that creative thinking may be more intimately tied to the visual and temporal structure of experience than previously thought. Subjects completed a “Divergent Association Test” (or DAT) over the course of a “virtual walk” through a 3D rendered space. At different points of the walk, they would be prompted to provide a common noun that was maximally “unrelated” from a previous noun (with the important constraint that it could not just be an item in the room). In the DAT, greater semantic dissimilarity between nouns predicts more creative thinking. We then monitored how subjects’ creativity “ebbed and flowed” over the course of the walk. Critically, all subjects walked towards a doorway at the second half of the walk—a manipulation known to mark an event boundary in time. Half of the subjects walked towards a doorway where they had a clear view of the other room (“future-is-known”) and the other half, a view into a black void (“future-is-unknown”). Subjects’ DAT scores in the second half showed a boost when the “future was unknown”, but not when the “future was known”. This result held true when subjects approached three doorways (instead of a black void), suggesting that the effect may come from the inability to anticipate what will happen next, not just the absence of visual input. Thus, visual event structure

can fuel divergent thought—leading us to spontaneously explore and sample from a wider space of possibility when we do not know what’s about to come.

33.342 USING GENERATIVE MODELS TO PROBE SCENE REPRESENTATIONS IN VISUAL SENSORY MEMORY

David R. Song¹ (david2004song@gmail.com), Ritik Raina¹, Abe Leite¹, Gregory J. Zelinsky¹; ¹Stony Brook University

In our lab’s prior work, we built Seen2Scene, a novel fixation-to-image latent diffusion model capable of generating metameric images using blurry peripheral input and high-resolution foveal information. Our prior work validated this model in the context of working memory, where a same–different task across an 8 second delay revealed scene metamers – pairs of images that, though different, were judged to be the same. In this poster, we extend that paradigm to the study of sensory memory scene representations. In a guided-saccade change-detection task on 300 Visual Genome images, participants monitored for scene changes while following a red moving ring around the image. The ring followed a predetermined scanpath generated from an empirical fixation density map for the image, and each saccade’s amplitude was at least 10 degrees of visual angle to ensure saccadic suppression. In change trials, the scene was replaced by a generation between fixations 4-5, 7-8, or 10-11. Potential metamers were generated by Seen2Scene, conditioned with foveal information at the pre-and-post-change locations, as well as the fixated locations of the followed scanpath. Catch trials contained no change. The primary outcome was scene-change miss rate, which we see as a proxy of sensory metamers – generated scenes that produce understandings with equivalent sensory memory representations. We found that miss rates rose with fixation index, rising from 34% at 5 fixations to 45% at 11 fixations, indicating that generations conditioned on more fixations had higher metamerism rates. These findings support the utility of Seen2Scene generations as probes into sensory memory scene representations, and the effect of fixation count suggests that fixation history may contribute some information to sensory representations. Future work will include feature analyses of metameric scene elements, as well as new scanpath conditions to assess whether and how sensory memory judgments are specifically influenced by fixation history.

33.343 COMPARING FUNCTIONAL TOPOGRAPHY IN FACE BLIND INDIVIDUALS AND NEUROTYPICALS DURING VISUAL RECALL OF FAMILIAR PEOPLE AND PLACES

Alexandra Haralanova¹ (alexharalanova@gmail.com), Alexis Kidder¹, Christopher Baker¹; ¹National Institutes of Health, National Institute of Mental Health, Bethesda, MD, USA

Developmental prosopagnosia (DP) is a condition characterized by severe impairment in face recognition ability despite typical early visual processing and the absence of any known brain lesions (McConachie, 1976). The underlying causes of atypical face recognition are unclear,

though theories include problems with perceptual encoding or memory recall deficits. Prior neuroimaging work has primarily focused on perceptual encoding and visual-selectivity, reporting the presence of the typical face-selective regions in DP participants, but with reduced selectivity. Here, we focused on memory recall for people and places. In neurotypicals (NT), we have previously described alternating regions along the medial parietal cortex (MPC) wall which preferably responded to self-generated recall of people or places (Silson, Steel, et al. 2019). For this study, we asked NT ($n = 13$) and DP ($n = 7$) participants to recall and visualize personally familiar people and places from memory without any visual input besides name prompts. The people and places that the participants recalled were selected by the participants prior to the scan. We also functionally localized visually-selective regions, including those responsive to faces, bodies, objects, and scenes. Consistent with prior reports, we could localize the major visually face-selective regions in both NT and DP participants, although the DP participants often showed more activation in frontal cortex. We were also able to find the expected alternating patterns of activation along MPC for recall of people and places in most participants, as well as strong activation of anterior scene-selective cortex for recall of places. We are systematically comparing these activation patterns within and across groups, quantifying the relationship between activation magnitude and self-reported vividness of imagery for the people and places. These results indicate that the topography of functional activation is similar in DP to NT participants for both visual encoding and recall of faces.

33.344 SPATIAL MEMORY IN STRUCTURED ENVIRONMENTS

Anthony Herzog¹ (herzog.127@osu.edu), Kyra Warmuth¹, Sami R. Yousif¹; ¹Ohio State University

Visual memory is biased. Even when remembering something as simple as the location of a dot within a circle, people make small but systematic errors — errors that reveal the underlying processes that structure spatial memory. Two prominent accounts attempt to explain these biases. On one account, spatial memory is described using a two-system framework in which a coarse, categorical representation of space interacts with a fine-grained, metric representation, with biases arising primarily from the categorical system (Huttenlocher, Hedges, & Duncan, 1991). More recently, an alternative account has proposed that spatial memory is biased towards regions of high perceptual acuity (Langlois et al., 2021). Here, we present evidence that challenges key predictions of both accounts. Across seven experiments, participants completed variants of a classic visuospatial memory task in which they viewed a single dot inside of a shape and later recalled its location. The task was performed in square environments with various salient structures (e.g., grid lines, landmarks). Three patterns emerged. First, categorical spatial biases proved remarkably stable in some cases: The presence of gridlines, for instance, did not eliminate such biases, with participants continuing to recall dots closer to the centers of their quadrants. Second, and in

contrast, local structural changes produced global distortions. The introduction of a small landmark in one quadrant of the space dramatically reshaped error patterns across the entire environment, even disrupting categorical biases in the opposite quadrant. Third, regions of high perceptual acuity produced repulsive rather than attractive distortions (contradicting recent accounts). Together, these findings reveal several highly general regularities in spatial memory that are not predicted by existing theories, suggesting that environmental structure interacts with spatial representations in complex and unintuitive ways.

Special thanks to the Spatial Cognition Lab at the Ohio State University for all the support they've given.

33.345 EXAMINATION OF INDIVIDUAL DIFFERENCES IN SIZE PERCEPTION OF DISTANT OBJECTS

Juliet Horenziak¹, Keithan Ducre¹, Peter Xu¹, Kalanit Grill-Spector¹; ¹Stanford University

Recent research from our lab discovered a new psychological phenomenon — perceptual magnification — wherein observers perceive small and distant objects as larger. Surprisingly, we found that perceptual magnification is substantially reduced when viewing the same scene monocularly. However, it is unknown whether this phenomenon differs for observers with impaired depth perception. Here, we tested perceptual magnification in participants with varying degrees of stereovision. 31 participants completed size estimation tasks (20 female, 11 male). Stereoacuity was assessed using the Randot Stereo Test. 17 participants were stereotypical (≤ 70 arcseconds), 6 had reduced stereovision (71-500 arcseconds), and 8 were stereoblind (> 500 arcseconds). Participants made size estimates of objects while seated in Stanford's Main Quad using reference probes placed on a table in front of them. They first completed a perceptual matching task in which they binocularly viewed 8 target objects at different distances [15.20–151.89 m], visual angles [0.38–7.15°], and elevations [−3.183°–9.43°]. For each object, they were asked to select the probe that best matched the size of the target in their visual field. Then, they repeated the experiment, now holding the probes up to each object and adjusting the chosen probe until a size match was found. The second experiment was done monocularly. We find that all participants, stereoblind, stereoreduced, and stereotypical, magnify the size of distant objects. Perceptual magnification is greater for smaller, more distant, and less elevated objects in the binocular case. In the monocular case, perceptual magnification is greater for smaller, more distant, and more elevated objects, but the effects of distance and elevation are overall smaller in magnitude. These results suggest that stereoacuity impacts the size perception of distant objects, even at distances of up to 150 m, typically not considered to use stereovision.

Stanford Vice Provost for Undergraduate Education Small Grants

33.346 LEARNED ASSOCIATIONS BETWEEN COLOR AND INFORMATION PREDICTION IN THE TRAINING

PHASE OF AN EXPERIENCE-DRIVEN ATTENTIONAL CAPTURE EXPERIMENT

Leon Hu¹ (leon.hu@trincoll.edu), Alenka Doyle¹, Emily Gonzales¹, Michael Grubb¹; ¹Trinity College - Hartford, CT

Attentional capture has traditionally been viewed as either a top-down (goal-driven) or bottom-up (stimulus-driven) process. However, recent work suggests a third type: attentional capture can be driven by associations with prior experiences. In recently published work, we demonstrated increased attentional capture by task-irrelevant, non-salient distractor colors associated with a history of providing instrumental (relative to redundant) information. The current study investigates whether such information-driven attentional capture occurs when stimuli have a history of predicting information versus providing it directly as in our previous work. Here, we report data from the training phase of our new experiment: on each trial, participants first reported a stimulus color (red or green), then received an endogenous cue (high info/low info) probabilistically tied to the preceding color. Participants then identified the orientation of a uniquely oriented Gabor patch out of four options; high info cues indicated the exact location of the target Gabor, and low info cues indicated two possible locations. Participants completed 4 blocks of 60 trials. In the training phase, we aim to determine whether a) task-relevant information is used to improve task performance, and b) whether the information-color associations are learned. Data from the first 40 participants provides evidence of improved Gabor task accuracy and RTs for high vs. low info cues. Additionally, by the last training phase block, RTs in the color discrimination task were significantly faster for high- vs. low-information-predicting colors, with no speed-accuracy trade-off, supported by multiple statistical analyses. These results suggest that the high info cues helped improve overall performance and, given the faster RTs for high-info-predicting colors, that participants gradually learn associations between color and information prediction throughout the experiment's training phase.

Supported by NSF-2141860 CAREER Award to MAG

33.347 BACKGROUND COMPLEXITY AND EMOTION RECOGNITION IN VIDEO CONFERENCING

Jane Kalla¹, Yelda Semizer¹; ¹New Jersey Institute of Technology

Virtual communication has become a part of our daily life, allowing us to connect with others across locations around the world. While video conferencing tools facilitate remote communication, they may also present perceptual challenges, especially in cases where visual features are complex or degraded in quality. For example, allowing users to control their virtual backgrounds may increase user satisfaction; however, the resulting perceptually rich displays may introduce additional demands on the perceptual system. In this study, we examined the extent to which the visual complexity of attendees' backgrounds influences the ability to accurately judge the average emotion of a group. Using a controlled set of faces with happy and

neutral expressions, we generated a series of composite images in which the mean and standard deviation of emotional expressions of face groups were systematically varied. The visual complexity of attendees' backgrounds was also manipulated. Using these stimuli in a 2-interval forced choice task, we asked participants to judge the average emotion of a group of individuals in a simulated video conferencing display. Preliminary results suggest that visually complex backgrounds impair the ability to extract average emotional expression from groups, making it more difficult to process perceptually relevant information.

33.348 THE ROLE OF INFORMATION PROCESSING STYLE IN IMAGE MEMORABILITY

Haripriya Kemiseti¹ (haripriyakemiseti21@gmail.com), Yelda Semizer¹; ¹New Jersey Institute of Technology

In everyday life, we sometimes evaluate images, such as deciding how appealing they are, while at other times we simply scan them for specific details. These differences in the way we process visual information can influence how well we remember it later. In this study, we explored how differences in the way people process images affect their memorability. One group of participants were asked to rate images based on perceived attractiveness, encouraging global processing, while another group was asked to search for specific targets, engaging local processing. Following the rating or search tasks, both groups were asked to complete a surprise memory task to assess image memorability. Preliminary results indicate that participants who completed the rating task outperformed those who completed search task, suggesting that global processing is associated with superior memory for images compared to local processing. These findings highlights the practical importance of how we attend and process the visual information can influence how well we remember it.

33.349 IS THAT A FACE?: INTERACTION BETWEEN ECCENTRICITY AND CONTEXT IN FACE PAREIDOLIA

Neeranuch Kittikiatkumjom¹ (neeranatk@gmail.com), Waragon Phusuwan^{1,2}, Payachana Victoria Chareunsouk^{1,2}, Chaipat Chunharas^{1,3,4}, Sedthapong Chunamchai^{1,2,3,4}; ¹Cognitive Clinical and Computational Neuroscience Center of Excellence, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand, ²Medical sciences program, Division of Neuroscience, Faculty of Medicine, Chulalongkorn university, Bangkok, Thailand, ³Division of Neurology, Department of Medicine, Faculty of Medicine, King Chulalongkorn Memorial Hospital, Thai Red Cross Society, Bangkok, Thailand, ⁴Chula Neuroscience Center, King Chulalongkorn Memorial Hospital, Bangkok, Thailand

Despite naturally noisy and limited inputs, the human visual system rapidly and robustly detects faces, extending to illusory face percepts in inanimate objects—known as face pareidolia. One hypothesis proposes a filling-in process that integrates visual inputs with prior knowledge. Although previous studies have demonstrated that

misperception at the periphery—with increased sensory uncertainty due to degraded visual acuity—reflects greater reliance on surrounding low-level contextual information (e.g., contrast), it remains unclear whether high-level contextual information (e.g., faceness) similarly biases illusory face perception, and whether such modulation would vary with eccentricity. To address this gap, we used Mooney faces with quantified faceness scores, controlling for low-level properties, in a face detection task. We manipulated high-level contextual information by varying mean display faceness across multiple combinations of high-, low-, and scrambled non-face images. Eighteen participants completed two 300-trial blocks. Images were presented for 150 ms across four spatial locations at 3° or 6° eccentricity. Participants reported whether they perceived a face at each randomly probed position. At non-face locations, false-alarm rates were significantly higher at 6° than at 3° eccentricity, $F(1,17) = 11.63$, $p < .01$. Contextual faceness significantly interacted with eccentricity, $F(6,102) = 2.51$, $p < .05$. The eccentricity effect was strongest under low contextual faceness but diminished as contextual faceness increased, with convergence at higher contextual faceness. In contrast, contextual faceness did not differentially affect miss-rates across eccentricities, suggesting that the pattern was specific to false alarms rather than a uniform response shift from degraded visual acuity. These findings demonstrate that illusory face perception is jointly shaped by sensory uncertainty and high-level contextual information, with contextual effects depending on eccentricity. This pattern is consistent with the Bayesian framework in which the influence of priors increases with higher sensory uncertainty.

33.350 FUNCTIONALLY DISTINCT SUB-REGIONS OF THE PARAHIPPOCAMPAL PLACE AREA REVEALED BY MODEL-BASED NEURAL CONTROL

Ranjani Koushik^{1,2} (ranjanikoushik@gatech.edu), Alex Abate³, Frederik Kamps⁴, N. Apurva Ratan Murty^{1,2}; ¹Cognition and Brain Science, Georgia Institute of Technology, ²Center for Computational Cognition, Georgia Institute of Technology, ³Harvard Medical School, ⁴University of Edinburgh

The parahippocampal place area (PPA) is central to the brain's scene processing system, yet its finer-grained functional organization has remained easier to describe anatomically than functionally. If the PPA contains functionally distinct subregions, those subregions should reflect distinct computations, require distinct computational models, and be independently dissociable with distinct images. Here, we integrated data-driven analyses of neural response structure, subregion computational models, in silico hypothesis testing, and closed-loop experimental validation to define the roles of PPA subregions in scene processing. We first used data-driven clustering to identify PPA voxel groups with similar responses to a large image set (185 images, 20 repetitions per image, $N = 4$). Across participants, more than 95% of response variance was captured by just two anterior-posterior clusters, consistent with prior work (Baldassano et al., 2013; Nasr et al., 2013; Cukur et al., 2016). To uncover what

functionally distinguishes these subregions, we built ANN-based encoding models for each. The models predicted held-out responses well ($R > 0.70$, $P < 0.00001$) and identified image pairs expected to strongly drive one subregion while suppressing the other. We tested these model-derived predictions in a new pre-registered fMRI experiment ($N=15$), revealing a clear double dissociation: the anterior subregion responded stronger to sparse spatial layouts, whereas the posterior preferred images rich in objects and texture. We next asked whether this dissociation reflected a broader organizing principle across the scene-selective network. It did not—the pattern was specific to PPA and did not generalize reliably to other scene-selective regions, including OPA and MPA/RSC ($F(2,22) \geq 14.62$, $p = .001$). Together, these findings reveal reproducible anterior-posterior functional organization within human PPA, with a functional dimensionality of at least two. We further provide localizers for both subregions, computational models for each, and initial insight into their roles in scene recognition.

33.351 BRAIN RESPONSES TO SYMMETRY IN EARLY INFANCY

Aurore Maloh^{1,2} (auroremaloh@gmail.com), Lara Pierce¹, Peter J. Kohler^{1,2}; ¹Department of Psychology, York University, ²Centre for Vision Research, York University

Visual symmetry is a fundamental property of natural environments and plays an important role in object recognition and perceptual organization. In adults, brain imaging studies have demonstrated robust responses to reflection symmetry in several areas of the visual cortex, using dot patterns with one or more axes of reflection. Another approach has used the wallpaper groups, 17 regular textures that comprise the complete set of possible symmetries in the 2D plane. Each wallpaper group contains a unique combination of fundamental symmetries, including reflection and rotation. The visual cortex responds robustly to combinations of symmetries in the wallpaper groups, with reflection symmetry typically producing stronger responses than rotation symmetry. The current study used the wallpaper groups to record brain responses to reflection (PMM) and rotation (P4) symmetry from 18 infants aged 2-8 months using high-density electroencephalography (EEG) with a Steady-State Visual Evoked Potentials (SSVEPs) paradigm. Infants viewed full-field (24°/visual angle) wallpaper textures from the PMM and P4 groups alongside matched phase-scrambled control images that preserved low-level visual properties while removing global structure. This paradigm isolates symmetry-specific responses and low-level responses driven by changes in contrast at each image update. Results revealed robust image update responses for PMM and P4, indicating that infants were generally equally willing to look at the screen for both conditions. We also found adult-like responses to reflection symmetry in infants, but rotation symmetry produced responses that were much less adult-like, if measurable at all. To our knowledge, these are the first measurements of brain responses to symmetry in early infancy. Our results suggest a learned component

of symmetry processing, and hint that infants may learn to see reflection earlier than rotation, possibly due to differential exposure during the initial months of life.

33.352 SHIELDING SACCADDES FROM SOUNDS

Vanessa Michel¹, Mark Eckert², Alex White¹; ¹Barnard College, ²Columbia University Irving Medical Center

Background: Oculomotor freezing (OF) is the rapid inhibition of eye movements caused by sensory stimulation. While well-documented in the visual domain, cross-modal freezing and its modulation by cognitive state remain poorly understood. This study answers two questions: Do visual and auditory stimuli elicit comparable OF, and do concurrent task demands modulate the magnitude of this inhibition? **Methods:** Across four experiments, participants performed two tasks that elicited frequent saccades but differed in top-down control. In the reading task, participants read passages and answered comprehension questions. In the dot-following task, they tracked a randomly moving dot and counted its color changes. During both of these tasks, visual and auditory distractor stimuli were presented at unpredictable intervals. The visual stimulus was a briefly flashed frame around the edge of the display. The auditory stimuli were pure-tone beeps and white noise bursts. **Results:** Visual flashes elicited strong OF during both tasks – the saccade rate dropped to a minimum within ~ 110 ms. In contrast, auditory stimuli showed a task-dependent effect. They elicited minimal freezing during reading but significant freezing during the dot-following task. Notably, when auditory OF did occur, it was roughly 30 ms faster than visual OF. Furthermore, white noise bursts elicited stronger freezing than pure tones, potentially reflecting stimulation of a broader population of frequency-tuned neurons. **Conclusions:** Saccade planning is inevitably disrupted by salient visual onsets, but it can be “shielded” from auditory stimuli during tasks that require strong top-down control, such as reading. In contrast, saccades in the dot-following task are more stimulus-driven and thus more susceptible to inhibition by salient irrelevant sounds. These findings position OF as a sensitive, objective tool for measuring how the brain prioritizes sensory inputs, which may differ across age and in clinical populations.

Special thanks to Martin Rolfs.

33.353 DOES PERSPECTIVE DISTORTION MODULATE THE TEMPORAL TUNING OF SYMMETRY RESPONSES IN THE HUMAN VISUAL CORTEX?

Nikan Movahedi^{1,2} (movahedinikan@gmail.com), Peter J. Kohler^{1,2}; ¹Department of Psychology, York University, Toronto, ON, ²Center for Vision Research, York University, Toronto, ON

Symmetry is a highly salient feature in both natural and man-made environments. Numerous species are sensitive to symmetry, and symmetry is thought to be an important cue for visual tasks, including figure-ground segmentation, shape detection, and viewpoint-invariant representation of objects. However, while symmetries are common in

natural and artificial objects and scenes, in natural vision, they are often distorted by perspective and rarely produce symmetrical patterns on the retina. Here, we build on previous studies showing that perspective-distortion makes symmetry responses weaker and more task-dependent (Makin et al., 2014; Keefe et al., 2018) by investigating its effect on the temporal tuning of symmetry responses. We used novel, naturalistic 3D objects with reflection symmetry about a vertical axis. Symmetric objects and well-matched asymmetric controls were procedurally generated and rendered to produce images in which object symmetries are either present in the image-plane or subject to perspective-distortion. Across 3 studies (N=95), we measured visual system responses to both image-plane and perspective-distorted symmetries using high-density EEG with a Steady-State Visual Evoked Potentials (SSVEPs) paradigm, in which images of symmetrical objects alternate with images of control objects. This allows isolation of symmetry-specific brain activity in the odd harmonics of the stimulation frequency. We investigated the temporal tuning of these responses across four experiments using seven stimulation frequencies between 1 and 10Hz. We found that responses to image-plane and perspective-distorted symmetries generally shared the same temporal tuning, with slower frequencies (1-2 Hz) producing the highest amplitude responses. Perspective-distorted symmetries produces responses that were weaker overall, consistent with prior work, and more right-lateralized, compared to image-plane symmetry. Our results suggest that image-plane and perspective-distorted symmetry is detected by mechanisms that share the same temporal tuning, but may differ in the contribution of the two hemispheres.

This work was supported by the Vision Science to Applications (VISTA) program funded by the Canada First Research Excellence Fund (CFREF, 2016–2023) and by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada awarded to PJK.

33.354 WHO LINGERS LONGER?: TEMPORAL DISCREPANCIES IN MEMORY OF WOMEN VS MEN

Maggie Nichols¹, Xiuyuan Flora Zhang², Kimberly W. Wong¹; ¹Williams College, ²Brown University

How faithfully do we remember those that we encounter? Prior work in social psychology suggests the answer depends on gender and context: In meetings or conferences, female speakers are expected to talk less and male speakers are expected to dominate the conversation—resulting in female speakers being judged as having talked longer than male speakers, even if they actually hadn't. This bias in judgment has been often attributed to gendered cultural stereotypes (e.g. of traditional authority and dominance). But what if this bias goes beyond misattribution of talking time? Here we propose that this could even be captured by more fundamental differences in memory for individuals of different genders—before any words are exchanged at all. In short, even when simply passively viewing men and women, might women be remembered as having lingered longer?

On each trial of the study, observers were shown a full body image of a male or female individual (White, Asian, Latino, or Black), for varying durations (from 500ms to 1500ms). Observers then reproduced the image duration by holding down the spacebar. Despite gender and race being completely irrelevant to the task, observers nevertheless consistently reproduced women for longer durations than men. In particular, this effect was significant for white female images compared to white male images. Effects persisted even with identical bodies, varying only the head. We interpret these results as evidence of representational asymmetry in memory for men vs. women—even in the absence of behavior, words, or context. Because women are culturally backgrounded while men are allocated more time and space, any duration that a woman actually occupies may be encoded as disproportionately long compared to male counterparts. This difference may explain several downstream higher-level social judgments—from a woman's talkativeness to her assertiveness—simply because she is spontaneously remembered as occupying more time.

We would like to thank Williams College Office of the Dean, as well as the Williams College Psychology Department.

33.355 ROLE OF SENSORIMOTOR REPRESENTATIONS IN UNDERSTANDING EMOTIONAL BODY MOVEMENTS BY ONESELF VS. OTHERS

Andrea Marie Pua¹ (apua88@students.claremontmckenna.edu), Alison Harris¹; ¹Claremont McKenna College

Action simulation theory proposes that we understand others' internal states by recreating their actions with our own sensorimotor systems. Supporting this idea, electroencephalography (EEG) has shown that the sensorimotor mu rhythm (10-12 Hz) is reduced both during action execution and observation of others' actions. Recently, we found weaker "mu suppression" for emotional actions, perhaps due to greater variability between individuals in their bodily expression of emotions. If so, we would expect greater mu suppression when observing one's own emotional actions versus those of others. In this study, participants completed a two-session experiment in which point-light displays (PLDs) of affectively-neutral and emotional body movements were recorded with motion capture, followed by a separate session measuring their neural responses during observation of these PLDs by themselves and others. Because the PLDs removed other visual identifying information, discrimination depended solely on the pattern of whole-body movement. We found no significant mu suppression for participants' own versus others' emotional movements, suggesting that decreased action simulation for emotional actions is not simply due to reduced familiarity or ease of simulation. However, separate behavioral experiments confirmed that participants could successfully discriminate whether emotional movements were performed by themselves or another, despite limited visual familiarity with the PLDs. Exploratory EEG analyses implicated the beta band (14-40 Hz) in actor identity perception, consistent with proposed roles in motor inhibition and motor set maintenance.

Although inconsistent with an action simulation account based on mu suppression, these results nonetheless support a role for sensorimotor representation of emotional actions.

33.356 FACIAL ASYMMETRY DRIVES DISTORTIONS IN PROSOPOMETAMORPHOSIA: A CASE STUDY

Ellie Rodgers¹ (eleonor.m.rodgers.26@dartmouth.edu), Antonio Vitor Reis Goncalves Mello, Brad Duchaine; ¹Dartmouth College, ²Dartmouth College, ³Dartmouth College

Prosopometamorphopsia (PMO) is a rare neurological condition in which individuals experience perceptual distortions when they view faces. Although these distortions were first documented more than a century ago, the processes underlying them are poorly understood. Here, we present a detailed case study of a participant with PMO to investigate whether asymmetry between the left and right halves of the face drives distortions. A.S. is a 58-year-old woman who has experienced intense, dynamic distortions affecting facial feature shape and position since 2010. In Experiment 1, A.S. rated distortion intensity (scale: 0-6) of 40 faces. Ratings for symmetric faces (mean) were significantly lower than ratings for natural faces (mean) ($p < 0.001$), indicating that left-right facial asymmetry contributes to distortions. In Experiment 2, we further assessed the effect of asymmetry by asking A.S. to rate distortions in naturally asymmetric faces presented with different parts visible. The stimuli were either full faces (asymmetry present), top and bottom halves (asymmetry present), left and right halves (asymmetry absent), or isolated facial features. Distortions were strongest for the full faces, present but reduced in intensity for the top halves and pairs of eyes, and mostly absent for the bottom halves, left halves, right halves, and individual features ($p < 0.001$). In Experiment 3, we contrasted frontal views of faces (asymmetry present) with full profiles (asymmetry absent); A.S. saw distortions to all of the frontal views and no distortions to faces presented in profile. Together, these findings suggest that the visual system calculates the asymmetry between the left and right halves of the face every time a face is viewed, and therefore, some cases of PMO depend critically on facial symmetry. Disrupting left-right symmetry substantially reduces perceived distortions, providing insight into the perceptual mechanisms underlying this condition.

Dartmouth College Department of Psychological and Brain Sciences

SUNDAY MORNING POSTERS IN PAVILION

Functional Organization of Visual Pathways: Subcortical, clinical

SUNDAY, MAY 17, 8:30 AM – 12:30 PM, PAVILION

33.401 DECODING NEURAL REPRESENTATIONS OF AFFECTIVE SCENES IN THE LOW-ROAD PATHWAY OF EMOTION PROCESSING

Max Lobel¹, Lihan Cui¹, Yun Liang¹, Ke Bo², Xiangren Wang¹, Freydel Espinoza¹, Andreas Keil¹, Mingzhou Ding¹; ¹University of Florida, ²Dartmouth College

In this study, we analyzed two independent fMRI datasets to test the role of the “low-road” pathway, from the superior colliculus, through the pulvinar nucleus of the thalamus, to the amygdala, in visual affective scene processing, with a secondary goal being to evaluate the effectiveness of different types of machine learning tools in decoding fMRI data. In the first dataset, simultaneous EEG-fMRI data were recorded from 20 human participants viewing pleasant, unpleasant, and neutral pictures from the International Affective Picture System (IAPS). In the second dataset, fMRI data were recorded from 30 human participants performing an emotion reappraisal task, in which they were cued to anticipate and then viewed unpleasant and neutral pictures from the IAPS. Single-trial BOLD responses were estimated and subjected to decoding analysis by support vector machine (SVM), convolutional neural network (CNN), and visual transformer (ViT). The results showed that across the two datasets, both CNN and ViT were able to consistently detect the differences in neural representations of different categories of affective scenes in the three structures, whereas SVM’s performance was inconsistent. These results provide support to the hypothesized role of the low-road pathway in emotion processing and demonstrate the importance of incorporating deep-learning based techniques into neuroimaging studies. Additional analyses carried out include (1) assessing of emotion-specific functional connectivity among the three structures and (2) fusing of EEG and fMRI data to examine the temporal dynamics of affective scene processing in the low-road pathway.

NIH grants MH125615 and MH112558 and NSF grant BCS2318984

33.402 NONINVASIVE TEMPORAL INTERFERENCE STIMULATION OF AMYGDALA CIRCUITS MODULATES PUPILLARY RESPONSES DURING THREAT ANTICIPATION

Marnie White¹ (marnewhite@gmail.com), Yoojeong Choo¹, Alena Zheng¹, Joshua MacMillan², Jason Smith¹, Ines Violante³, Alexander Shackman¹, Weizhen Xie¹; ¹Department of Psychology, University of Maryland, College Park, ²A. James Clark School of Engineering, University of Maryland, College Park, ³School of Biomedical Engineering and Imaging Sciences, King’s College London

Deep circuits such as the amygdala-hippocampus complex play a key role in higher-level visual cognition and affective processing, yet they remain difficult to study and modulate noninvasively in humans. Limitations in the spatiotemporal resolution of existing methods constrain efforts to test causal links between subcortical dynamics and visual cognitive functions. Temporal interference (TI) stimulation is an emerging noninvasive technique that leverages the interaction of two

high-frequency electrical currents (e.g., 2000 Hz and 2010 Hz), each too fast to elicit neuronal firing on its own, to generate a lower-frequency envelope (e.g., 10 Hz) in deeper tissue, offering a potential avenue for subcortical modulation without invasive intervention. Here, we examine the feasibility of TI stimulation in modulating amygdala-related pupillary responses during threat anticipation in a double-blinded, sham-controlled study. Participants performed a threat-anticipation task previously shown in fMRI work to engage the amygdala. On each trial, a geometric cue signaled that a threatening or neutral image-sound pair might occur unpredictably, eliciting high versus low levels of anticipatory anxiety. In our pupillometry adaptation, real-time pupil dilation provided a continuous physiological marker of this anticipatory state, devoid of low-level visual confounds (Experiment 1). We next applied 10 Hz TI stimulation targeting the amygdala circuit during the anticipation period in a randomized, blocked design (Experiment 2). Preliminary results indicate that 10 Hz TI stimulation attenuates pupil dilation to threat cues relative to neutral cues, whereas sham stimulation preserves the original threat-related pupil dilation. A high-frequency control study using a 0 Hz envelope (Experiment 3) provided further evidence for the specificity of the 10 Hz effect. Together, these findings highlight TI stimulation as a viable noninvasive approach for modulating deep subcortical circuits central to visual-affective processing.

33.403 OPTOGENETIC ACTIVATION OF PMLS CORTICOGENICULATE FEEDBACK SHIFTS LGN NEURONAL TUNING

Adam Kunz¹, Matthew Adusei², Farran Briggs¹; ¹National Eye Institute, National Institutes of Health, ²National Institute of Mental Health, National Institutes of Health

Layer-6 corticothalamic feedback in the visual system has classically been cast as a modulator of first-order thalamic relay cells rather than a source that drives their activity. Yet the role of corticothalamic feedback arising outside primary visual cortex remains uncertain. Recent discoveries of distinct corticogeniculate cell types in V2 and mid-level extrastriate areas in both ferrets and macaques open new avenues for investigating the function of corticogeniculate pathways. Here we asked whether L6 feedback from the extrastriate posteromedial lateral suprasylvian (PMLS) area, implicated in visual motion processing, reshapes responses in the dorsal lateral geniculate nucleus (LGN) in ways that challenge the modulator-only view. Using retrograde viral gene delivery and optogenetics, we selectively and reversibly activated PMLS L6 corticogeniculate neurons in vivo while recording LGN neuronal responses to drifting gratings, bars, flashed spots, and m-sequence white noise stimuli. Activating PMLS feedback systematically shifted LGN neuronal tuning for temporal frequency toward higher values with little change in overall gain, baseline firing, or spike-timing precision. These changes align LGN output with PMLS neuronal preferences for high temporal frequencies, consistent with PMLS’s role in visual motion perception. Together, the results suggest that PMLS corticogeniculate feedback plays a specialized role in

configuring LGN sensitivity during early visual processing to behaviorally relevant stimuli, an effect more “driver-like” than purely modulatory. They also provide the first visual system evidence that LGN tuning can be modified through “egocentric selection”, a mechanism previously described for corticothalamic circuits in the auditory and somatosensory systems, whereby thalamic responses align with the tuning preferences of their active cortical inputs.

33.404 NEURAL DECODING OF VISUAL REPRESENTATIONS FROM MOUSE SUPERIOR COLLICULUS AND PRIMARY VISUAL CORTEX

Reo Tsukasa^{1,2}, Masatoshi Kasai^{1,3}, Shuntaro C. Aoki^{1,2}, Tadashi Isa^{1,4}, Yukiyasu Kamitani^{1,2}; ¹Kyoto University, ²ATR, ³Kitasato University School of Medicine, ⁴Institute for the Advanced Study of Human Biology (WPI-ASHBi)

The brain processes visual information through two major pathways: the cortical pathway, which includes the primary visual cortex (V1), and the subcortical pathway, which includes the superior colliculus (SC). Although the cortical pathway has been extensively characterized, the functional role of the SC in visual representation remains comparatively understudied. Recent experimental and analytical advances have begun to address this gap. In particular, progress in optical imaging has enabled single-cell-resolution recordings of SC activity (Kasai and Isa, 2016; 2021), providing new opportunities for direct comparisons between V1 and SC. In parallel, studies in humans have shown that both simple visual features and hierarchical deep neural network (DNN) features can be decoded from activity in the visual cortex (Kamitani and Tong, 2005; Horikawa and Kamitani, 2017). Building on these developments, we used neural decoding approaches to investigate how visual information is represented in the mouse SC relative to V1. We examined visual representations in mouse SC and V1 by decoding features from population activity recorded with two-photon calcium imaging in anesthetized animals. The stimulus set included natural scenes and simple artificial stimuli, allowing comparisons across different types of visual features. Using linear regression decoders, we estimated direction, spatial position, and hierarchical DNN features from neural responses. Both V1 and SC contained sufficient information to predict multiple visual attributes; for example, the direction of bar motion and the spatial position of dot stimuli were reliably decoded. Moreover, DNN-derived features—particularly those from self-supervised models trained on low-resolution images—were decoded with high accuracy. While SC and V1 exhibited broadly similar decoding performance, future work employing more fine-grained analyses—such as unit-level comparisons of decoding accuracy—will be necessary to determine whether more subtle differences in visual representations exist between these two regions.

This research was supported by AMED JP24wm0625409 and KAKENHI JP25H00450.

33.405 USING SUPPORT VECTOR MACHINE (SVM) TO CLASSIFY NEURAL PATHWAY CONTRIBUTIONS IN THE VISUAL EVOKED POTENTIAL (VEP) AND TO DETECT VISUAL PATHOLOGY FROM THE VEP

Jawshan Ara¹ (jawshan@gmail.com), Alireza Tavakkoli¹, Michael A. Crognale¹; ¹Institute of Neuroscience, University of Nevada, Reno, USA

Pattern VEPs can noninvasively quantify activity from specific achromatic and chromatic visual pathways (e.g., Murray et al., 1987; Kulikowski et al., 1989; Rabin et al., 1994). Achromatic pattern reversal (PR), which includes high spatial frequencies, is typically employed to stimulate the transient magnocellular pathway. Conversely, chromaticity modulation along equiluminant L-M cone opponent axis and the S-cone axis, using lower spatial frequencies and pattern-onset (PO) modes, is best suited to activate the more sustained chromatic pathways. VEP waveforms recorded with these parameters usually show stereotypical shapes. However, when stimulus parameters deviate from these optimal conditions, waveforms can vary significantly between individuals. For instance, achromatic PO-VEPs often exhibit greater inter-individual variability (e.g., Crognale et al., 1998; Crognale, 2002; Odom et al., 2016), since they reflect contributions from both magnocellular and chromatic pathways. Due to this variability, assessing the integrity and contributions of different neural pathways becomes challenging when stimulus parameters are not optimized as described. We classified VEPs using a Support Vector Machine (SVM) on VEP waveforms to classify the contribution of specific visual pathways. We also classified VEPs elicited by a broader, more general set of stimulus parameters. Next, we constructed a dataset to model VEP responses by varying the amplitude and latency of VEPs from visually normal subjects to simulate VEPs with visual pathologies. We then classified raw VEPs from visually normal subjects versus those from simulated visual pathologies, thereby establishing the clinical significance of simple machine learning algorithms for visual deficit detection. We plan to use SHAP (Shapley Additive exPlanations) analysis to rank VEP signal features, thereby enhancing the interpretability of the classification and extending these methods to multifocal VEPs to aid in the classification of localized visual deficits.

33.406 ELEVATED OCULAR TNF-ALPHA LEVELS AND ALTERED VISUAL PROCESSING: IMPLICATIONS FOR SCHIZOPHRENIA

Tanique McDonald^{1,2,3}, Steven Silverstein^{1,2,3,4}, Farran Briggs^{1,5}; ¹Department of Neuroscience, University of Rochester Medical Center Rochester, New York 14642, ²Center for Visual Science, University of Rochester, Rochester, New York 14627, ³Department of Psychiatry, University of Rochester Medical Center, Rochester, New York 14642, ⁴Department of Ophthalmology, University of Rochester Medical Center, Rochester, New York 14542, ⁵National Eye Institute, National Institutes of Health, Bethesda, Maryland, 20892

Schizophrenia is a severe neuropsychiatric disorder that is characterized by multiple visual system changes. The causes of these changes are unknown, however elevated systemic levels of TNF α , an inflammatory cytokine, have been reported in drug-naïve, first-episode schizophrenia patients. When released by microglia in the retina, TNF α is known to promote the death of retinal ganglion cells which in turn has been linked to degenerative effects along the feedforward visual system pathway, some of which are consistent with altered visual response properties frequently observed in people with schizophrenia. In this study, we report on a new animal (ferret) model that tests the hypothesis that retinal neuroinflammation can generate many of the visual system changes observed in people with schizophrenia. Using optical coherence tomography and single-neuron recordings, we are assessing the structural and functional alterations to early visual structures (retina, visual thalamus, primary visual cortex) following the elevation of TNF α levels in the eye after intravitreal injection. Preliminary results suggest that loss of retinal cells depends on the concentration of TNF α injected as well as the time between injection and neuronal recordings. Moreover, following intravitreal injection of TNF α , significant retinal hypertrophy was observed as well as altered visual response properties including reduced firing rates, elevated contrast sensitivity and altered spatial and temporal frequency preferences for contralesional cells of the lateral geniculate nucleus. With this study, we aim to use a ferret model of elevated ocular inflammation to explore the neural mechanisms underlying perceptual deficits in schizophrenia at different stages of disease progression.

33.407 EVENT-BASED NETWORK DYNAMICS AFTER PEDIATRIC CORTICAL RESECTION: CASE-SERIES EVIDENCE FROM MOVIE-WATCHING FMRI

Sophia Robert¹, Christina Patterson², Marlene Behrmann^{2,1}; ¹Carnegie Mellon University, ²University of Pittsburgh

After unilateral cortical resection in childhood, the preserved hemisphere supports visual perception and broader cognition with markedly altered structural connectivity. Naturalistic movie watching provides a rich, continuous visual stimulus for probing how post-surgical visual and associative networks coordinate over time. Time-varying functional connectivity reveals how cognitive networks coordinate activity moment-to-moment, reflecting dynamic changes in integration and segregation. We used movie-watching fMRI and an edge-centric framework to test how whole-hemisphere dynamics and the properties of transient high-amplitude (HA) network co-fluctuations are altered in patients with unilateral cortical resections for drug-resistant epilepsy during childhood. Eight patients (15–20 years; 3 male; 5 with a preserved left and 3 with a preserved right hemisphere) and seven typically developing controls (9–20 years; 2 male) watched a 10.5-minute film segment twice in a 3T MRI scanner. Cortical time series from the preserved hemisphere were parcellated (Schaefer-400) and converted to edge time series. The root-sum-of-squares (RSS) across edges indexed global co-fluctuation amplitude at each

frame. HA events were defined as contiguous low-motion frames where RSS exceeded a permutation-derived threshold. We quantified intersubject similarity of global and network RSS trajectories, including visual and attention networks, and compared HA-event rate and amplitude between groups using mixed-effects models, complemented by single-case tests with control-as-patient validation. Group models indicated reduced HA-event rate (but not amplitude), and case-series analyses revealed a subset of patients with an intact left hemisphere whose event rates were markedly below the hemisphere-matched control distribution, classified as highly robust abnormalities. Almost all patients showed normative event amplitudes, even when rates were reduced. These results indicate lower rates of high-amplitude co-fluctuation events in the contralesional hemisphere after pediatric cortical resection and underscore the value of case-level analyses for characterizing post-surgical large-scale dynamics during naturalistic visual processing.

This research was conducted at the CMU-Pitt BRIDGE Center (RRID: SCR_023356) and supported by the NIH, National Eye Institute (NEI, R01EY027018), the NSF GRFP (DGE2140739), an NEI P30 CORE Award (EY08098), The Research to Prevent Blindness Inc, NY, and the Eye & Ear Foundation of Pittsburgh.

33.408 MAPPING CORTICAL REORGANIZATION IN AMBLYOPIA BEYOND V1 USING ULTRA-HIGH FIELD FMRI

Martin Szinte¹, Uriel Lascombes¹, Adrien Chopin², Paul V. McGraw³, Denis Schluppeck^{3,4}; ¹Institut de Neurosciences de la Timone, CNRS, Aix-Marseille Université, Marseille, France, ²The Smith-Kettlewell Eye Research Institute, San Francisco, CA, United States, ³Visual Neuroscience Group, School of Psychology, University of Nottingham, Nottingham, United Kingdom, ⁴Sir Peter Mansfield Imaging Centre, University of Nottingham, Nottingham, United Kingdom

Amblyopia affects 3-5% of adults and is characterized by reduced visual acuity in one eye despite best optical correction and absence of any other overt ocular pathology. While visual deficits are well-documented in amblyopia, the cortical organization underlying these deficits remains debated. Previous measures of population receptive field (pRF) size and cortical magnification in primary visual cortex (V1) have been inconclusive, potentially due to differences in methodology and viewing conditions (monocular vs. binocular). Here, we substantially extend the analysis of a high-resolution 7T fMRI retinotopy dataset previously acquired (monocular stimulation) in participants with strabismic amblyopia (n=12) or no amblyopia (n=9). Our new analysis extends beyond V1 to multiple early and intermediate visual areas (V2, V3, V3AB, hMT+). We examined pRF parameters as a function of eccentricity and cortical area to probe potential reorganization at different levels of visual processing. We next applied surface-based geodesic distance methods to estimate cortical magnification across the visual hierarchy. While the original analysis of this dataset (Schluppeck et al., 2025, Vision Research) found largely preserved V1 organization, it did not explicitly test for alterations in higher visual areas or for changes that may be better

captured by a more local cortical magnification measure. This work aims to clarify if and where such reorganization occurs, thereby shedding new light on the quality and extent of potential markers for cortical reorganization in amblyopia, informing ongoing debates about the neural mechanisms underlying amblyopic vision loss.

33.409 ISOLATING NONGENICULOSTRIATE INPUTS TO CATEGORY-SPECIFIC RESPONSES IN THE VENTRAL STREAM IN BLINDSIGHT

Jessica M. Smith¹, Bradford Z. Mahon; ¹Carnegie Mellon University

Category-specific responses in the ventral stream have traditionally been understood as due to processing of visual inputs through the geniculostriate pathway. Recent findings, however, suggest that visual inputs processed by nongeniculostriate pathways modulate patterns of category-specific neural responses in the ventral stream (Garcea et al., 2018; Vuilleumier et al. 2004; Bar et al., 2006). To isolate nongeniculostriate contributions to category-specific processing in the ventral stream, we studied cortically blind participants (n=3) with phenomenological blindness in part of their visual field due to a geniculostriate lesion. Despite that disruption to 'seeing', intact nongeniculostriate pathways can still process visual information in the cortically blind field, a phenomenon called blindsight. While previous research with this population has demonstrated neural responses to unseen stimuli in cortical and subcortical regions, it is unknown if the ventral stream receives input from nongeniculostriate pathways in the absence of seeing. We present here fMRI and connectivity analyses to investigate a) the presence of ventral stream responses to objects presented in the blind field and b) the patterns of structural and functional connectivity that underlie those effects. Each participant completed a task in which neutral faces, fearful faces, tools, or gabor patches were presented to the seen or blind field. Functional ROIs were defined with an independent category localizer in which stimuli were presented at central fixation. We found continued category-specific responses in the ventral stream for 'unseen' stimuli, including right lateral fusiform gyrus activity for unseen faces, and bilateral inferior temporal gyrus activity for unseen tools. Nongeniculostriate pathways generate fast orienting responses to salient stimuli, which may shape detailed perceptual analysis supported through the geniculostriate pathway. These results demonstrate that nongeniculostriate pathways contribute to category-specific neural responses in the ventral stream, challenging models of ventral stream organization that focus entirely on inputs contributed through the geniculostriate pathway.

33.410 DEVELOPING AOSLO IMAGING IN THE MARMOSET TO STUDY TRANSSYNAPTIC RETROGRADE DEGENERATION

Amy Bucklaew¹, David Fernandez¹, Sara Patterson^{1,2}; ¹Center for Visual Science, University of Rochester, ²Flaum Eye Institute and Department of Neuroscience, University of Rochester Medical Center

Stroke induced vision loss is a prevalent source of visual impairment, resulting in nearly 100,000 new cases of hemianopia in the US and Europe each year (Sahraie et al 2007). While there is potential for training-induced vision restoration (Saionz et al 2020), the limited efficacy means that cortically induced blindness (CB) remains a permanent condition for most patients (Zhang et al 2006b). A key factor likely to affect visual recovery is the extent of trans-synaptic retrograde degeneration (TRD) in upstream areas such as the LGN and the retinal ganglion cells (RGCs). Prior investigations of the TRD in RGCs have been limited by the resolution achievable through the eye's optics. These challenges have been overcome by combining calcium imaging with adaptive optics scanning light ophthalmoscopy (AOSLO) to allow for visualization of foveal RGC function in the living primate eye. To investigate TRD at the level of individual RGCs, we created a marmoset model of stroke by unilaterally lesioning area V1 with ibotenic acid. The marmoset is advantageous as a model due to its smooth cortex, which facilitates cortical targeting, and its comparable visual circuitry to humans. We constructed an extension of a macaque AOSLO system to target the marmoset eye by custom designing a marmoset-specific stereotax, optically scaling down the system's pupil size, and optimizing anesthesia/paralysis techniques to facilitate imaging in the marmoset. We can now image the marmoset cone mosaic in vivo for the first time. Next, we performed intravitreal injections of a calcium indicator, GCaMP8s, into each eye to label RGCs and enable functional imaging. Unlike other retinal physiology methods, our approach is both non-invasive and allows for the opportunity to re-image the same population of foveal RGCs for months or years during the progression of TRD.

Research was supported by the National Eye Institute of the National Institutes of Health under R00-EY035323, P30 EY001319. Research was also supported by an Unrestricted Grant to the University of Rochester Department of Ophthalmology from Research to Prevent Blindness, New York, New York.

33.411 MICROSTRUCTURAL CHANGES PREDICT HIGHER-ORDER VISUAL DEFICITS AFTER POSTERIOR CEREBRAL ARTERY STROKE

Sara Ajina¹ (s.ajina@ucl.ac.uk), Yuheng Zeng¹, Alexander Leff², Matthew Lambon Ralph³, Randi Starfelt⁴; ¹Wellcome Centre for Human Neuroimaging, UCL Queen Square Institute of Neurology, University College London, ²Institute of Cognitive Neuroscience, UCL Queen Square Institute of Neurology, University College London, ³MRC Cognition and Brain Sciences Unit, University of Cambridge, ⁴Department of Psychology, University of Copenhagen

Posterior cerebral artery (PCA) stroke can cause diverse higher-order visual deficits affecting word, object, and face perception, yet the supporting white-matter changes remain poorly understood. Here, we examined microstructural changes across visual white-matter pathways and their relationship to domain-specific perceptual deficits in chronic PCA stroke. Forty-four chronic PCA stroke survivors and 21 healthy controls completed a comprehensive visual battery including

face, object and word processing ("Back of the Brain" project). Domain-specific composite scores were derived using principal component analysis. All participants underwent 3T structural and diffusion MRI. Lesions were corrected using virtual brain grafting prior to multimodal-parcellation of 55 visual regions, 11 subregions. Whole-brain tractography between visual subregions used constrained spherical deconvolution. Tracts were weighted with SIFT2, normalized by intracranial volume. Ten major tracts were defined from control data. Mean Fractional Anisotropy was extracted and related to perceptual scores for left- and right-sided stroke using correlation and regression analyses; age and lesion volume as covariates. Word processing showed a modest overall association with left-hemisphere tract integrity ($F=8.1$, $FDR_p=0.05$), alongside a strong 'stroke-hemisphere' x 'tract-hemisphere' interaction: in left-sided stroke, word scores related primarily to left-hemisphere pathway integrity, whereas in right-sided stroke, performance related more strongly to right-hemisphere pathways ($F=13.0$, $FDR_p=0.006$). Face processing showed clear right-hemisphere dominance, associated with integrity in right ventral and dorsal stream tracts ($F=23.6$, $FDR_p=0.0001$). Object processing was selectively associated with integrity of right parahippocampal-ventral stream tract ($F=10.4$, $FDR_p=0.02$). Regression analyses showed that in right-sided stroke, integrity of right and left lateral temporal-STS pathways predicted face ($R^2=0.9$, $FDR_p=0.003$) and word ($R^2=0.8$, $FDR_p=0.007$) performance respectively, independent of lesion volume and age. Word deficits depend largely on tract integrity in the damaged hemisphere with a slight left-hemisphere bias, whereas face and object deficits show strong right-hemisphere dominance. Associations with contralesional tract integrity suggest interhemispheric plasticity and/or transneuronal degeneration may also influence visual outcomes.

This project was funded by the Independent Research Fund Denmark (Sapere Aude to R.S.; DFF - 4180-00201) and supported by a programme grant and intramural funding to M.A.L.R. from the Medical Research Council [MR/R023883/1; MC_UU_00005/18], and a Clinical Research Career Development Fellowship to S.A. from Wellcome [224655/Z/21/Z].

Functional Organization of Visual Pathways: Neuroimaging

**SUNDAY, MAY 17, 8:30 AM – 12:30 PM,
PAVILION**

33.412 FACE NETWORK IS ENGAGED IN SIGN LANGUAGE PROCESSING

Akshi .¹, Elizabeth Saccone¹, Miriam Hauptman¹, Matthew Sampson¹, Marina Bedny¹; ¹Johns Hopkins University

The role of Fusiform Face Area (FFA) in supporting linguistic processing remains an open question. Signed languages such as

American Sign Language (ASL) use faces and body-parts (hands, body orientation etc.) as articulators, with facial expressions being an integral part of ASL grammar. We investigate functional profile of FFA in Deaf signers and hearing non-signers using individual subject functional localization. Congenitally and profoundly Deaf adults ($n=8$) completed an ASL localizer (conditions: signed ASL sentences, meaning and perceptually matched Action videos, and a blurred low-level Control condition). The same participants also completed a dynamic face localizer (conditions: faces, objects, scenes, body parts, scrambled objects) and a written English localizer (conditions: sentences, nonword sequences; $n=7$). A hearing non-signer group ($n=14$) completed the same face localizer as the Deaf group and an analogous spoken English localizer with English and foreign language sentences. We tested whether the lateral fusiform, i.e., the typical location of the fusiform face area (FFA), develops specialization for perceptual processing of faces and bodies in sign language. In the Deaf group, face-preferring vertices inside an FFA anatomical mask (from a separate group of hearing participants) responded comparably to ASL and Action conditions (significantly more than low-level control). However, ASL-selective vertices in the same FFA mask (ASL > Action; leave-one-run-out) showed robust preference for ASL over Action and low-level Control, and also responded significantly more to face and body condition compared to other categories. In contrast, written English-selective vertices (sentences > nonwords) in Deaf participants did not show such preference for faces and bodies. Similarly, spoken English-selective vertices in FFA in hearing non-signers did not exhibit face/body preference. The results suggest that experience with a visuo-manual language, such as ASL, in Deaf signers leads to specialization of a subset of FFA vertices for sign-language-specific effectors (faces, body parts).

33.413 IMPROVED PRF MODELING OF POSITION-IN-DEPTH: TOWARDS MAPPING OF 3D VISUAL SPACE

Stephanie Kristensen¹, Julie Golomb¹; ¹Department of Psychology, The Ohio State University

Accurate perception of positions in depth is essential for human behavior. During natural vision, binocular disparity (the slight image difference between left and right eyes) is an important cue to perceive depth. Neural selectivity for binocular disparity is distributed across the visual cortex in regions as early as V1, though some prior fMRI studies suggest a transition from 2D to 3D processing along the visual hierarchy with more specialized processing of 3D depth perception in later visual areas. Our goal is to describe how neural populations across human cortex represent perceived depth and whether there exists a systematic organization of depth-related neuronal tuning, similar to 2D retinotopic maps. We reanalyzed data from an fMRI population receptive field (pRF) modeling experiment (Golomb, 2018) during which participants ($N=8$) passively viewed a random dot motion stereogram while wearing red/green anaglyph glasses. The stimulus depicts a full-field plane of dots located at varying positions-in-depth relative to fixation. Within a session, the plane moved forward or

backward through 13 depth positions ranging from -18 arcmin (perceptually near, in front of fixation) to 18 arcmin (perceptually far, behind fixation). We adapted standard pRF modeling with a 1D Gaussian to identify the preferred depth and tuning width (range of preferred disparities) for each voxel. To improve model fits, we also incorporated a parameter estimating each voxel's relative sensitivity to depth vs depth-invariant visual stimulation. The results revealed that voxels in early visual cortex exhibit primarily depth-invariant visual preferences, while tuning to position-in-depth increases in later visual areas. In particular, a cluster along the transverse occipital sulcus (overlapping with V3a/b and IPS0) shows a high degree of voxel-specific depth selectivity, with spatially separated voxels preferring negative or positive depths. Overall, this study builds on recent studies indicating that nuanced position-in-depth representations may reside in later visual areas.

NIH R01-EY025648 (JG)

33.414 IN VIVO EVIDENCE FOR PARADOXICAL SUPPRESSION OF INHIBITORY CELLS PREDICTED BY INHIBITION-STABILIZED NETWORKS

Brock M Carlson¹, Joseph Emerson^{1,2}, Vera Lindh¹, Ryan Holland², Yifan Hu², Audrey Sederberg³, Cheryl Olman^{1,2}, Gordon Smith²; ¹Department of Psychology, University of Minnesota College of Liberal Arts, ²Department of Neuroscience, University of Minnesota School of Medicine, ³School of Psychology, Georgia Institute of Technology

Cross-orientation suppression (COS) is a robust form of neuronal response suppression in the early visual cortex, typically attributed to divisive normalization and recurrent excitatory-inhibitory interactions. Theoretical work on inhibition-stabilized networks predicts a "paradoxical" response in which increases in intracortical drive lead to decreases in inhibitory activity (Tsodyks et al., 1997; Rubin & Miller, 2015). Indirect physiological support comes from membrane conductance estimations of intracellular recordings, which show reduced inhibitory conductance during surround suppression in cat area 17 (Ozeki et al., 2009). A direct demonstration of paradoxical suppression in inhibitory neurons in vivo has been lacking. Using two-photon calcium imaging and post-hoc histology in ferret V1, we examined COS responses in functionally characterized excitatory and inhibitory neurons. We observe strong COS in both excitatory and inhibitory neurons, with most inhibitory neurons exhibiting response reductions approaching the theoretical limit (COS index ≈ -1). These results provide preliminary in vivo evidence that inhibitory neurons in ferret V1 exhibit paradoxical suppression during COS, consistent with predictions of inhibition-stabilized networks.

This work was supported by NIH R01 NS123482

33.415 MESOSCALE FUNCTIONAL NETWORKS REVEALED BY INFRARED NEURAL STIMULATION OF MACAQUE PULVINAR IN ULTRA-HIGH-FIELD 7T MRI

Yuqi Feng^{1,2,3} (yuqifeng@zju.edu.cn), An Ping², Songping Yao^{1,2}, Sunhang Shi^{1,2}, Meilan Liu^{1,2}, Henry Evrard⁴, Jianbao Wang⁵, Anna Roe^{6,7}; ¹Key Laboratory for Biomedical Engineering of Ministry of Education, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China, ²Department of Neurosurgery of the Second Affiliated Hospital, Interdisciplinary Institute of Neuroscience and Technology, School of Medicine, Zhejiang University, Hangzhou, China, ³Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY, USA, ⁴International Center for Primate Brain Research, Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai, China, ⁵Affiliated Mental Health Center & Hangzhou Seventh People's Hospital, School of Medicine, Zhejiang University, Hangzhou, China, ⁶Translational Neuroscience Laboratory Division, Center for Biomedical Imaging and Neuromodulation (C-BIN), Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY, USA, ⁷New York University Grossman School of Medicine, New York, NY, USA

Introduction: In human and nonhuman primates, the pulvinar is a key hub for visual pathways, and supports higher-order functions such as attention and emotion. We previously showed, using a novel functional connectivity mapping method INS-fMRI (Infrared Neural Stimulation in ultra-high-field fMRI, 1 site: 1 brainwide network), that focal stimulation (1) in PL/PI (lateral/inferior pulvinar) led to mesoscale (mm-scale), topographic activations in visual cortical areas (in ventral and dorsal pathways) and (2) surprisingly, INS in PM (medial pulvinar) activated limbic areas (cingulate, insula, amygdala) in highly organized patchy fashion. Here, we extend this investigation to activated areas in prefrontal cortex. **Methods:** INS stimulation (1875nm, activating $<500\mu\text{m}$ -sized clusters of neurons, pulse trains 200 Hz, duration 0.5 sec; pulse width 0.25 msec, 200 μm optical fiber) was delivered to 28 loci in pulvinar (PI, PL and PM) of 2 anesthetized macaque monkeys in 7T fMRI (BOLD GRE-EPI, 1.5mm isotropic). Statistical tests and controls revealed reproducibility of activation sites. **Results:** (1) Each INS site in pulvinar PI, PL, and PM subnuclei activated brain-wide functional networks consisting of arrays of mesoscale activations. (a) Consistent with anatomical studies, the mesoscale functional connections of PL/PI and PM are distinct. Whereas INS in PI and PL revealed connections with visual areas (V1-V4, TPO, TEO, LIP, VIP), INS in PM revealed patterns of connections with limbic, medial and orbitofrontal (13,14) areas. Interestingly, stimulation sites in the foveal region of PI/PL revealed activations in high-order regions such as area 7a and dorsolateral prefrontal (8, 46) areas. All activations were consistently mesoscale in size and distributed in patchy, non-overlapping fashion. **Conclusion:** We extend and strengthen the view that pulvinar provides environmental context for both visual (PL/PI) and emotion-related (PM) function. Functional connections exhibit mesoscale, non-overlapping organization. Connectivity of foveal PI/PL with prefrontal areas suggest a potential specialization in foveal engagement.

33.416 SHARED VISUAL CATEGORIES, INDIVIDUAL NEURAL REPRESENTATIONS

Ivette Colón¹, Timothy Rogers¹; ¹University of Wisconsin - Madison

Neural representations of visual categories are reliably distributed across cortex, yet it remains unclear how much of their internal structure is shared across individuals versus individually specific, and how stable such structure is over time (Haxby et al., 2014). Specifically, do neural representations preserve a shared categorical organization while exhibiting stable, individual-specific representations across time? Here, participants viewed controlled sets of familiar face, place, and object images during two fMRI sessions separated by several weeks. Neural responses were analyzed using sparse multiclass decoding to assess category information across cortical activity patterns. Mean within-session decoding accuracy across participants was 70.2% in Session 1 and 66.9% in Session 2 (chance = 33%). To quantify the stability and individuality of representational structure, we trained decoding models within individuals and evaluated their performance across sessions and across participants. To enable cross-session and cross-participant comparisons despite differences in voxel sampling, we aligned voxel sets using bipartite graph matching based on spatial proximity in standardized brain space. Within individuals, cross-session decoding was highly reliable: models trained on Session 1 generalized to Session 2 with a mean accuracy of 65.5%. In contrast, decoding models trained on anatomically-matched data from other participants performed substantially worse when applied to a given individual (42.2%). The within-individual advantage was large when expressed relative to the between-participant distribution (mean z-score = 3.18), demonstrating pronounced individual specificity in representational organization. Importantly, above-chance decoding accuracy persisted when analyses were restricted to voxels outside canonical category-selective regions, indicating that individual-specific semantic structure is distributed broadly across cortex rather than confined to classical functional localizers (Julian et al., 2012). Together, these results show that joint representations of faces, places, and objects are stable over time within individuals, reliably distinct across people, and supported by distributed neural patterns.

33.417 SPIKINESS AND INANIMACY MEDIATE TOOL REPRESENTATION IN THE HUMAN OCCIPITOTEMPORAL CORTEX

Boyang Hu¹, Marvin Chun¹, Yaoda Xu¹; ¹Yale University

A central goal in human vision is to identify the key features supporting visual object representation in the human brain. Our recent fMRI work has identified several cortical regions preferring spiky over stubby objects, regardless of animacy, and regions preferring either animate or inanimate objects, independent of shape spikiness. These regions span ventral and lateral occipitotemporal cortex (OTC) and posterior parietal cortex (PPC), raising the question of how spikiness- and animacy-preferring areas relate to category-selective areas. Tool-selective areas are especially pertinent here, as they are typically

localized by contrasting responses to tools vs animals, and many tools have elongated or pointed shapes. To investigate the relationship between tool-selective areas and areas modulated by object spikiness and animacy, we conducted an fMRI study in 16 right-handed participants. We identified regions selective for spikiness and animacy while participants viewed images from four stimulus categories, including spiky inanimate objects, stubby inanimate objects, spiky animate objects, and stubby animate objects. We separately localized the tool areas in these participants when they viewed images of tools and animals. We found that the tool-selective area lies exclusively within the intersection area responding to inanimate spiky objects in both OTC and PPC, showing that both animacy and spikiness mediate tool selectivity. This area overlap persisted even when tools were excluded from the inanimate objects, showing that shape, rather than the function of the inanimate object, was relevant here. The same results were obtained when object silhouettes were used to define brain areas, indicating that the effect reflects sensitivity to shape contours rather than surface details. Lastly, the same results were found when elongated spiky objects were excluded, showing that spikiness, rather than elongation, played a more critical role here. Together, these results show that both spikiness and inanimacy mediate tool representation in the human brain.

Supported by NIH Grant R01EY030854 to YX.

33.418 TASK-DERIVED FUNCTIONAL CONNECTIVITY IS SUFFICIENT TO CAPTURE INDIVIDUAL DIFFERENCES IN VENTRAL VISUAL SELECTIVITY

Isaac Liao¹, Zeynep M. Saygin¹, David E. Osher¹; ¹The Ohio State University

Most studies of ventral visual networks use resting-state functional connectivity (FC) to characterize the brain's intrinsic functional organization. However, recent work has questioned whether we should rely exclusively on resting-state data to define this organization (Finn, 2021). In this study, we test whether functional connectivity estimated from residualized task fMRI data can predict task responses in ventral visual regions at the individual subject level. While previous studies have demonstrated that responses of visual regions can be predicted from resting-state FC at the level of individual subjects (Saygin et al., 2012; Osher et al., 2016; Molloy et al., 2024; Tavor et al., 2016; Osher et al., 2019), it is unclear whether residual connectivity from task fMRI of a completely different domain (e.g. language) can a) predict responses to visual processing; and b) if it can do so in a selective manner (i.e. for functional regions of interest within each network). We find that connectivity derived from task data predicts responses just as well as connectivity derived from resting-state data. Task data, regardless of the specific task performed, are sufficient to capture fine-grained individual differences in ventral visual activations. These findings suggest that dedicated resting-state scans may not be necessary and broaden the utility of task fMRI for investigating the functional organization of brain networks.

33.419 THE VWFA AS A NEUROCHEMICAL AND FUNCTIONAL BRIDGE BETWEEN LANGUAGE AND VISUAL CORTEX

Laura Bradley¹, Zeynep Saygin¹; ¹The Ohio State University

The visual word form area (VWFA) sits at a unique intersection of high-level visual and language systems, yet the biological factors that anchor it to these systems remain poorly understood. A central question is whether the VWFA is more similar in its underlying organization to its visual neighborhood or by its connectivity to language network. Functional and structural brain asymmetries support uniquely human abilities like language, and connectivity patterns in children can predict later VWFA location, suggesting that both local structure and long-range connections may constrain its emergence. Here, we combine neurotransmitter density profiles and functional connectivity to characterize the architecture of the VWFA. To address this, we examined category-selectivity regions in the ventral temporal cortex (VTC), including VWFA, face-selectivity regions (FFA, OFA), and object-selective regions (LO, PFS), alongside frontal and temporal language regions. Neurochemical architecture was quantified from mean PET-derived receptor density maps (1,238 adults, 20 neurotransmitters/transporters). We applied hierarchical clustering and representational similarity analysis to characterize relationships among neurotransmitter profiles across ROIs. Hierarchical clustering and RSA revealed that the frontal language regions are most distinct from temporal language and visual regions; face ROIs cluster with each other, object ROIs with each other. The VWFA is neurochemically most similar to temporal language regions. The distinct neurochemical profile of the VWFA from other visual regions and its similarity to temporal language regions suggest that neurochemical affinity may be the physical mechanism by which privileged connectivity between VWFA and temporal language cortex is instantiated in development.

33.420 V1 FOVEA-CENTRIC FUNCTIONAL ARCHITECTURE IN HUMANS WITH NEUROTYPICAL AND AMBLYOPIC VISION

Zexi Chen¹ (zexic7@uci.edu), Marianna Elisa Schmidt^{2,3}, Shahin Nasr^{4,5}, Cherlyn Joan Ng^{1,6}, Daniel Ts'o⁷; ¹University of California, Irvine, Irvine, CA, USA, ²Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, ³Max Planck School of Cognition, Leipzig, Germany, ⁴Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA, ⁵Harvard Medical School, Boston, MA, USA, ⁶Brunson Center for Translational Vision Research, University of California, Irvine, CA, USA, ⁷SUNY Upstate Medical University, Syracuse, NY, USA

Introduction: Intra-areal functional connections (FCs) within V1 are thought to support coherent visual perception across the visual field. These connections, however, are not uniform; rather, they exhibit systematic heterogeneity that may help explain known perceptual asymmetries. In this study, we tested whether intra-V1 FC varies

between foveal and peripheral regions as suggested based on foveal confluence—the dense clustering of foveal representations in V1. We then asked whether this organizational principle is altered in amblyopia, a developmental disorder that disproportionately affects foveal vision. Methods: In Experiment 1, 33 adults (aged 22-43) with neurotypical vision were scanned during resting state (eyes closed) in a 3T MRI (3mm isotropic); 20 of these participants also completed a separate 7T session (1mm isotropic). Experiment 2 included 4 neurotypical controls (aged 24-47) and 7 individuals with amblyopia (aged 27-60; 4 with strabismus and 3 with anisometropia) scanned at 7T. For all participants, the V1 surface representing the central 10° was segmented into 16 equal-area subregions. Intra-V1 FC was quantified using partial correlations between vertices across these subregions (Schmidt et al., 2025). Results: In Experiment 1, regions representing more foveal visual field locations showed significantly stronger FC with one another than with regions representing equivalent eccentricities in the periphery ($p < 10^{-4}$). This pattern was robust across both field strengths and detectable across deep, middle, and superficial layers at 7T. Our preliminary results in Experiment 2 showed that, compared to controls, strabismic amblyopia showed both reduced foveal bias and a global reduction in intra-V1 FC, whereas anisometropic amblyopia preserved overall FC magnitude but showed reduced foveal specificity. In both amblyopia subtypes, the weakened foveal bias was more apparent in superficial layers. Conclusion: These findings reveal a pronounced foveal-centric organization of V1 FC and demonstrate its vulnerability to disruptions in balanced binocular vision early in life.

This work was supported by NIH NEI (R01 EY030434), and by the MGH/HST Athinoula A. Martinos Center for Biomedical Imaging. Crucial resources were made available by a NIH Shared Instrumentation Grant S10-RR019371.

33.421 WHAT IS THE RELATIONSHIP BETWEEN FUNCTIONALLY LOCALIZED BRAIN REGIONS AND THE HUMAN CONNECTOME PROJECT'S MULTI-MODAL CORTICAL AREAS IN INDIVIDUAL HUMANS?

David Rahabi¹, Sam Hutchinson¹, Ammar Marvi², Matthew Glasser³, Nagehan Demirci³, Nancy Kanwisher¹; ¹MIT, ²Harvard, ³WashU

The last few decades of fMRI research have identified many cortical regions with distinctive and highly selective response profiles, including the fusiform face area (FFA), parahippocampal place area (PPA), extrastriate body area (EBA), and visual word form area (VWFA). Because the anatomical locations of these regions vary substantially across individuals, group-based analyses often fail to identify them precisely, motivating the widespread use of individual functional localizers. Although recent work has made these localizers more efficient (e.g., Marvi et al., 2025), they still require costly scan time. Meanwhile, multimodally-defined cortical areas based on the Human Connectome Project (HCP, Glasser et al., 2016) have been widely used to summarize multi-modal MRI data, but it is unknown the extent to which they are capable of providing fine-grained

characterizations of local cortical functional selectivity. We therefore quantified the selectivity of the response of the FFA, PPA, EBA, and VWFA when these regions are defined with individual functional localizers versus when they are defined as the HCP cortical areas in the same region. Individual functional localization revealed significant and strong category selectivity in held-out data in all four regions, whereas category selectivity was much weaker (PPA and EBA) or even reversed (FFA and VWFA) in both a) the closest HCP-defined area/s (see also Molloy et al., 2024), and b) the top 10% of vertices within those closest HCP-defined areas with the highest tSNR in resting functional data. Thus, HCP-defined cortical areas do not reproduce the functional specificity obtained from regions defined with individual localizer scans.

NIH Project Number: UM1 MH130981-01, "Functionally guided adult whole 459 brain cell atlas in human and NHP", PI: Ed Lein

33.422 EFFECTIVELY ROBUST MODELS BETTER EXPLAIN BRAIN DATA ON OUT OF DISTRIBUTION STIMULI

Bhavin Choksi^{1,2} (choksi@uni-frankfurt.de), David Trung Nguyen¹, Sari Saba-Sadiya^{1,2}, Gemma Roig^{1,2}; ¹Goethe University Frankfurt, Department of Computer Science, ²The Hessian Center for Artificial Intelligence (hessian.AI), Darmstadt, Germany

Deep neural networks are increasingly used as computational models of the human visual system, yet the extent to which brain-model alignment reflects meaningful generalization beyond natural images remains unclear. While the generalization of the alignment has been investigated on the model side, how it holds when the human brain data, instead of the model, is acquired on out-of-distribution (OOD) settings remains relatively unexplored. In this work, we systematically investigate whether alignment between model representations and human visual cortex responses predicts robustness to OOD stimuli. We quantify robustness using Effective Robustness, which isolates a model's OOD performance beyond what is expected from its in-distribution accuracy. Across 70 ImageNet-trained models, we relate Effective Robustness on multiple OOD benchmarks to representational similarity with fMRI recordings from early visual (V1, V2) and ventral stream (V4, IT) brain regions, measured using both natural and synthetic stimuli. We find that robustness under texture-based distribution shifts consistently correlates with brain alignment in early visual areas, regardless of stimulus type. In contrast, brain-model similarity in higher ventral stream areas correlates with robustness only when derived from fMRI responses to synthetic, rather than natural, stimuli. Together, these results demonstrate that the link between model-brain alignment and model generalization is currently highly dependent on both stimulus characteristics and the cortical regions considered. Our findings highlight the gaps in the current state of the brain-model alignment and the importance of considering the stimulus used when collecting neural data and using ANNs as cognitive models.

This work was funded by the Deutsche Forschungsgemeinschaft: DFG project 5368 and DFG project 539642788, RO 6458/5-1.

3D Shape and Space Perception: Miscellaneous

SUNDAY, MAY 17, 8:30 AM – 12:30 PM, PAVILION

33.423 INTERNAL NOISE AND EFFICIENCY DIFFERENCES UNDERLIE THE STEREOSCOPIC ANISOTROPY

Ichasus Llamas-Cornejo¹ (illamas@ucm.es), Ignacio Serrano-Pedraza¹; ¹Universidad Complutense de Madrid

There is a well-established anisotropy in stereovision: sensitivity for detecting sinusoidal disparity corrugations of low spatial frequency is higher for horizontal than for vertical corrugations. Despite the strong evidence for this stereoscopic anisotropy, its cortical locus, underlying mechanisms, and functional role of this phenomenon remain unclear. Several explanations have been proposed, including the number of underlying channels or the size of the summation fields in stereovision, but none provide a clear explanation of the effect. In the present study, we used a masking paradigm to estimate the internal noise and efficiency for detecting vertical and horizontal sinusoidal corrugations. Using a Bayesian staircase procedure, we measured disparity thresholds at three spatial frequencies (0.2, 0.4, and 0.8 cpd) masked by disparity white noises of different power. As expected, our results show that thresholds increased with noise power. We then fit an equivalent noise model to the threshold vs. noise power data. The model has two parameters: internal noise and efficiency. The fitting results indicate that the mechanism responsible for detecting horizontal corrugations exhibits lower internal noise and higher efficiency than the mechanism detecting vertical corrugations. Interestingly, while internal noise increased with spatial frequency, efficiency remained approximately constant across spatial frequencies for both orientations. These findings support the existence of different orientation mechanisms underlying disparity processing and suggest that differences in internal noise and efficiency contribute to the stereoscopic anisotropy.

Supported by grant PID2024-155243NB-I00 from Ministerio de Ciencia, Innovación y Universidades (Spain) to ISP.

33.424 STEREO VISION WITHOUT 3D GEOMETRY: KEEPING STEREO VISION IN RETINAL (VS WORLD) COORDINATES

Paul Linton¹ (paul.linton@columbia.edu); ¹Columbia University

How we infer 3D geometry from multiple viewpoints has received renewed interest (Tsao & Tsao, 2022; O'Connell et al., 2025; Lee &

Watrous et al., 2025). We formalize an alternative ('minimal') model of stereo vision, where stereo vision isn't about inferring 3D geometry, but eradicating rivalry relative to fixation (Linton, 2023). On this model, the three common stages of stereo analysis are a mistake: 1. RELATIVE DISPARITIES: Stereo depth being nearer vs. further than fixation is a fundamentally different visual experience. It is therefore a mistake to eradicate fixation (move from 'absolute' to 'relative' disparities), as much of the literature does. Depth from diplopia (Ziegler & Hess, 1997) is also hard to explain otherwise. 2. CORRESPONDENCE PROBLEM: The world doesn't change with fixation, but our stereo percept does, leading to manifestly illogical outcomes. Cross-fusing two coins side-by-side leads to a single fused central coin and two monocular flankers. This is hard to justify as a 'solution' to the correspondence problem. Instead, stereo vision appears to fuse whatever is on the fovea, and stereo matching (and depth) radiates outwards from that fixed starting point. 3. DISPARITY SCALING: Linton Stereo Illusion (VSS Demo Night 2024; ECVP 2024) demonstrates that perceived stereo depth is not inferred from the underlying 3D geometry. Instead, stereo depth appears to be a linear function of disparity on the retina. We aim at a model that is independent of physical size and distance. Pilot data suggests $z = 5.4x$, where x is disparity, and z is the x-axis angular size match of the perceived depth. IMPLICATIONS: First, this approach provides a more plausible evolutionary account for stereo vision. Second, it suggests a mechanism consistent with early (V1) processing. Third, it transforms our understanding of visual space, from a fully articulated 3D space, to fixation plus nearer/further than fixation.

This research project and related results were made possible by the support of the NOMIS Foundation. NOMIS Foundation 'New Theory of Visual Experience' grant to PL at the Italian Academy for Advanced Studies, Columbia University. Research conducted in Visual Inference Lab, Zuckerman Mind Brain Behavior Institute, Columbia University.

33.425 LOOKING AT DENSITY AND NUMBER IN DEPTH

Nichole Suero Gonzalez¹, Frank Durgin¹, Alex Huk²; ¹Swarthmore College, ²University of California, Los Angeles

Density is a visual property that seems relevant to the estimation of visual number. Some argue that number is encoded directly; others that density and number are entangled: Both density and number are strongly biased by patch size. One reason patch size may affect perceived density is that size and density, rather than number, are distance cues, so a larger patch might be a nearer patch that is higher in density than it appears. To study effects of patch size on density and number in a more naturalistic situation in which depth matters, we used virtual environments, rendered in a high-resolution head-mounted display (HMD). The basic task consisted of a brief presentation of a circular dot texture (~ 30 dots) at 1m viewing distance, followed by a second dot texture at another viewing distance (0.5-2m). The second texture varied in size, number, and density. Participants (14) made either number comparisons (was the second patch more or fewer dots than the first?) or density comparisons (was the second patch denser

or less dense than the first?) in separate sessions. An additional between-session manipulation involved using either retinally matched textures (the simulated dots and patches of dots were enlarged proportionally with distance to project the same retinal sizes), or real-world image variations in another, such that the dots and patches were retinally smaller at far distances and retinally bigger at near distances. We consistently found that both number and density judgments were strongly and equally affected by the relative sizes of the two patches. This shows that, even in a simulated real-world context, participants' judgments of relative number and of relative density remain entangled. Because number is (in principle) invariant with distance, effects of patch size on number contradict direct encoding, and support the idea that number is entangled with density perception.

Swarthmore College

33.426 SENSORIMOTOR INTERACTION ACTIVELY RESHAPES INTRINSIC BIASES IN THE PERCEPTION OF 3D SPACE.

Chaeun Lim¹, Dhanraj Vishwanath², Fulvio Domini¹; ¹Brown University, ²University of St. Andrews

Systematic biases in distance and depth perception distort the perceptual map of 3D space. Can such biases be actively updated by physical interaction? To address this, we leveraged a specific bias identified in our previous research (Lim, Vishwanath, & Domini, 2025). While standard theories posit that egocentric distance is required to scale exocentric depth, our findings suggest a bidirectional relationship: systematic biases in exocentric depth estimation also distort the perception of egocentric distance. We demonstrated this using stereoscopic paraboloids where the back of a near object (near-back) physically aligned with the front of a far object (far-front). Due to typical depth overestimation in near space, the near-back appears significantly farther away than the far-front. This creates a "spatial discrepancy" where physically equidistant points are perceived at different distances. Here, we investigated whether sensorimotor interaction could correct this discrepancy. In a pre-test where participants estimated the near-back and far-front surface locations relative to a reference stimulus, the spatial discrepancy was replicated: the near-back was judged as significantly farther than the far-front, despite being physically collocated. Participants then performed a learning task, reaching to touch both surfaces with haptic and visual feedback. This signaled that the perceptually distinct near-back and far-front were actually at the same distance. A post-test—identical to the pre-test—revealed that the spatial discrepancy was significantly reduced following learning, unlike in a control group. It is striking that this learning from a short-term (72 trials) visuomotor task transferred to a pure perceptual task involving no motor response. This reduction in bias could not be due to simple visuomotor adaptation but a result of a fundamental rescaling of the perceptual 3D map. These findings demonstrate that the visual system actively utilizes sensory feedback to correct intrinsic biases, updating the perception of space itself.

This material is based upon work supported by the National Science Foundation under Grant No.2120610.

33.427 MODELING SHAPE-FROM-MOTION PROCESSING IN THE PRIMATE DORSOVENTRAL VISUAL PATHWAYS WITH VIDEO-COMPUTABLE NEURAL NETWORKS

Yoon Bai¹ (yhb@mit.edu), Thomas O'Connell², Ani Ayvazian-Hancock⁴, Hannah Maver¹, Yoni Friedman³, Josh Tenenbaum^{2,3}, James DiCarlo^{1,2}; ¹MIT McGovern Institute for Brain Research, ²MIT Department of Brain and Cognitive Sciences, ³MIT Computer Science & Artificial Intelligence Laboratory (CSAIL), ⁴MIT Division of Comparative Medicine

Perceiving shape from dynamic sensory data is a fundamental challenge for the visual system, enabling robust interaction with a dynamic world. How does the primate visual system compute object shape from dynamic motion cues? We investigated this question using textured objects, human behavioral experiments, and dorsoventral neural activity recorded from macaque monkeys. Stimuli (2520 videos, 400ms each) were rendered from abstract 3D shapes with a 50° rotation. Both the object and background were overlaid with one of five high-contrast textures, camouflaging the object shape in the absence of motion. Behavioral experiments show that humans are near-ceiling at matching motion-induced shape across textures. We used multi-electrode chronic arrays to record from three downstream regions in the ventral pathway (pIT, ciT, alT) and two downstream regions in the dorsal pathway (7op, Tpt). Neural decoding revealed that motion-induced shape is represented in both ventral and dorsal visual streams, but shape readouts from dorsal regions show higher correlations to human behavior than shape readouts from ventral regions. To identify candidate computations that may underlie these capabilities, we evaluated a large set (N=1280) of video-computable neural network models as computational analogs of primate visual processing. These include ImageNet CNNs and ViTs, unsupervised image models, foundation image models, video classification models, and autoregressive video models. Autoregressive video models trained to predict missing spatiotemporal features in natural videos exhibit superior alignment with both ventral and dorsal neural responses and human behavior. Finally, we find that the models best predicting neural responses are also best at behaviorally matching shapes across textures. Taken together, our results show that neuronal activity in downstream dorsal regions contain motion-induced shape representations that are closely correlated to human behavior and identifies predictive processing of spatiotemporal regularities as a plausible computational motif for how the brain constructs geometric world models from motion cues.

Office of Naval Research (MURI N00014-21-1-2801)

33.428 DEVELOPMENTAL DIFFERENCES IN A NATURALISTIC EXPLORATION TASK

Qianqian Wan¹, Nathalie Moriarty¹, Shea Duarte¹, Simona Ghetti¹, Joy Geng¹; ¹UC Davis

Exploration is fundamental to learning, yet most developmental research has used simplified paradigms with discrete choices. Everyday exploration, however, requires coordinating movement, perception, and attention simultaneously in continuous environments. Processes that may influence exploration, including spatial navigation, visual encoding, and adaptation to salient events, have typically been studied in isolation. Measuring them together within a single exploration context allows us to assess their relative contributions to developmental differences. We asked children (ages 7–14, N=41) and adults (N=42) to freely explore a large-scale virtual furniture store (4,216 m²) in immersive VR while tracking eye, head, and body movements. The environment contained 10 multisensory objects (MS-objects) that produced sounds and animations when approached, providing controlled probes into how participants adapt to salient events during ongoing exploration. Adults explored significantly more than children, visiting more locations (92.8% vs. 72.5%, $p < .001$) and viewing more objects (42.7% vs. 29.7%, $p < .001$) despite traveling similar distances ($p = .235$). This difference was related to three factors. First, children backtracked more than adults ($p = .008$). Second, children spent more time looking at each object (0.61s vs. 0.47s, $p < .001$), reflecting slower encoding. Third, children adapted more slowly to MS-objects, indicated by less deep processing initially ($\beta = -.15$, $p < .001$), smaller changes across encounters ($p < .001$), and more fixated repetitions to encode MS-objects ($\beta = -0.66$, $p < .001$). Together, these findings suggest that children's exploration may be largely constrained by their developing memory systems. Regression models revealed that backtracking and encoding efficiency together explained 57.5% of the total explained variance (51.7%) in exploration within children, with age contributing only 0.1% unique variance. These results suggest that developmental differences in exploration are best understood through behaviors that improve with age rather than chronological age itself.

James S. McDonnell Foundation, NSF BCS 2446371

33.429 DIFFERENCES IN PERCEIVED SEPARATION IN THE MÜLLER-LYER ILLUSION ARE NON-FUNCTIONAL: PERCEIVED SPEED IS UNAFFECTED BY MIS-PERCEIVED SEPARATION

Fernando Castillo Rodriguez¹, Joseph DeNavas², Arthur G Shapiro², Cathleen M Moore¹; ¹University of Iowa, ²American University

In the Müller-Lyer (ML) illusion, the separation between two wings-out arrow heads is perceived as greater than the identical separation between two wings-in arrow heads. We asked whether this difference in perceived separation was functional in that it would transfer to differences in perceived speed. Specifically, we asked whether the speed of an object moving between two wings-out arrow heads would be perceived as greater than the speed of an identically moving object between two wings-in arrow heads. Using both linear (Brentano) and circular (Todorović) ML displays, large differences in perceived separation were found between wings-out and wings-in arrowheads. The perceived speed of moving discs within the same displays,

however, was unaffected by the direction of the arrow heads, and was generally accurate. Thus, the motion system was blind to perceived differences in separation due to the ML illusion. This indicates that the ML illusion does not give rise to functional representations of different separations, as mechanisms of size constancy do. These results, therefore, challenge explanations of the ML illusion that appeal to size constancy.

33.430 EVALUATING THE INTERACTION OF PRESENCE AND SET SIZE USING THE CONFIGURATION ERROR PARADIGM IN VIRTUAL REALITY

Sean Hinkle¹, Mark Neider¹; ¹University of Central Florida

Though long a hallmark of virtual reality (VR) research, spatial presence has been challenging to operationally define and even harder to objectively measure. The process model of presence claims that a more accurate and rich mental representation of virtual space leads to presence. The configuration error (CE) paradigm is a well-established test of egocentric and allocentric theories of spatial representation and has provided evidence that representation degrades for individual objects after disorientation, supporting an egocentric updating account. We recreated the CE paradigm in VR using simple stimuli and added a set size manipulation to test the process model's claim that a richer representation increases presence. We developed a virtual environment where participants were tasked with learning the location of 4 or 8 colored blocks and pointing to them in several phases (eyes-open, eyes-closed, post-disorientation). In a 2 (set size) x 3 (phase) within-participants design, we evaluated presence, accuracy, and changes in CE (standard deviation of mean error to blocks) after disorientation to test predictions of the process model, and to replicate and extend the CE paradigm in VR. We expected increased self-reported presence, CE, and average error with increased set size. We replicated key findings of the CE paradigm; increased CE following disorientation in both 4 and 8 block conditions ($p < .001$). However, there were no observed differences in self-reported presence ($p = .336$), CE ($p = .709$), or accuracy ($p = .275$) with changes in set size, though RT increased as expected ($p < .001$). The lack of change in accuracy may be explained by a constrained error space (360-degrees). Presence did not correlate with accuracy ($ps > .074$) as the process model predicts. Our results support the classical findings of the CE paradigm, and provide evidence against several claims of the process model of presence.

33.431 WHY DOES THE MOON APPEAR LARGER ON THE HORIZON? THE CURIOUS CASE OF PERCEPTUAL MAGNIFICATION

Keithan Ducre¹ (kmducre1@stanford.edu), Juliet Horenziak¹, Peter Xu¹, Kalanit Grill-Spector¹; ¹Stanford University

The full moon appears larger at the horizon than when elevated. Known as the moon illusion, its magnitude and mechanisms are debated. In experiment 1 (perceptual matching, $n=175$), participants were given circular probes and indicated which one matched the

moon's size for a lower and elevated moon. On the horizon, observers perceptually magnify the visual angle of the moon by 4.1x and magnification decreases with elevation (power law, exponent=-0.33). In experiment 2 (adjusted matching), the same observers were allowed to hold the probes against the moon and adjust until they found a match. Surprisingly, the illusion was abolished but perceptual magnification wasn't: observers perceptually magnify the moon by 1.96x on the horizon, and magnification increases with elevation (power law, exponent=0.07). To determine the underlying factors, we conducted perceptual and adjusted matching experiments ($n=140$) in the Stanford Quad with objects that varied in distance (15.2[m]–151.9[m]), visual angle (0.35[deg]–7.34[deg]), and elevation (0.64[deg]–13.7[deg]). Observers magnify all objects, more in perceptual (1.71x) than adjusted (1.07x) matching and are generally unaware of their magnification. Perceptual magnification is predicted by a power law: $\text{PerceptualMagnification} = C \cdot \text{VisualAngle}^a \cdot \text{Distance}^b \cdot (1 + \text{Elevation})^c$. It is larger for smaller ($a=-0.08$) and farther ($b_{\text{perceptual}}=0.37$; $b_{\text{adjusted}}=0.14$) objects, and the elevation exponent varies across conditions. Since observers close one eye during adjusted matching, we tested if binocular disparity varies with moon ($n=58$) and disk ($n=20$) elevation. We find that when visual angle and distance are fixed, disparity increases with elevation and predicts the moon illusion. These data suggest that perceptual magnification of the moon is particularly large because of its small visual angle (~0.5 [deg]) and large distance (~384,400 [km]) and that binocular vision contributes to the illusion. Together, we not only elucidate the moon illusion, which was enigmatic for over two millennia, but also discover a new phenomenon of perceptual magnification and its underlying mechanisms.

Funding: NIH grant R01EY023915; Stanford VPUE grants to Keithan Ducre and Juliet Horenziak.

Eye Movements: Individual differences, visual preference

**SUNDAY, MAY 17, 8:30 AM – 12:30 PM,
PAVILION**

33.432 SEE WHAT MATTERS: CLUSTERING VISUAL ATTENTION PATTERNS USING MULTIMODAL EMBEDDINGS

Shan Zhang¹ (zhangshan@ufl.edu), Christine Wusylko², Do Hyong Koh¹, Pavlo D. Antonenko¹, Anthony F. Botelho¹; ¹University of Florida, ²Kennesaw State University

Social media serves as a major channel through which adolescents learn about current events. Yet, evaluating this information is difficult and requires integrating visual, textual, and cognitive processes. To understand how learners process such multimodal content, traditional

visual attention studies often rely on Areas of Interest (AOIs), defined by fixed spatial regions. However, these spatially bounded AOIs can oversimplify the semantic relationships between images and text, limiting inferences of understanding. To address this gap, this study introduces a novel, context-driven method that uses shared high-dimensional text and image embeddings, combined with fixations to capture what adolescents attend to, not just where they look, when evaluating information on social media. Visual attention and cognitive data were collected from 29 middle and high school students evaluating 16 social media posts about climate change. Fixations were transformed into 116×116-pixel subimages ($n = 22,143$) from which multimodal embeddings were extracted using OpenCLIP and clustered with UMAP and K-means. Six visual-semantic clusters emerged, and clearly distinguished text-dominant ($n = 4$) and image-dominant ($n = 2$) attention patterns. Regression analyses revealed that participants with higher visuospatial working memory capacity (WMC) focused more on scientifically-relevant textual regions instead of images ($\beta = 0.29$, $p = .01$). Those with lower WMC devoted greater attention to temporally comparative or heuristic text (e.g., “years” “likes”) ($\beta = -0.3$, $p = .01$). Importantly, embedding-based context AOIs explained additional variance ($R^2 = 0.37$) in visuospatial WMC compared to task-specific visual-based AOIs, thus providing richer insights into the interplay between perceptual attention and cognitive capacity. By connecting visual attention patterns to individual cognitive differences, our study advances methodological and theoretical understanding of adolescent credibility evaluation on social media, contributes to vision science research, and lays the groundwork for attention-aware educational tools that foster more critical and informed evaluation practices with (mis)information.

33.433 STIMULUS CENTER BIAS PERSISTS IRRESPECTIVE OF ITS POSITION ON THE DISPLAY

Ohad Ben-Shahar^{1,2}, Rotem Mairon¹; ¹Faculty of Computer and Information Science, Ben-Gurion University of the Negev, ²School of Brain Sciences and Cognition, Ben-Gurion University of the Negev

Since the earliest studies on human eye movements, it has been repeatedly demonstrated that observers fixate the center of visual stimuli more than their periphery, regardless of visual content. Subsequent research suggested only little effect of typical biases in experimental setups, such as the observer’s position relative to the screen or the relative location of the cue marker. While comparative studies of the screen center vs. stimulus center revealed that both conspire in the process, much of the prior art is still confounded by experimental details that leave the origins of the center-bias debatable. We thus propose methodological novelties to rigorously test the effect of the stimulus center, isolated from other factors. In particular, eye movements were tracked in a free-viewing experiment after stimuli were presented at a wide range of horizontal displacements from a counterbalanced cue marker in a wide visual field. Stimuli spanned diverse natural scene images to allow inherent biases to surface in the pooled data. Various analyses of the first few

fixations show robust bias toward the center of the stimulus, independent of its position on the display, but affected by its distance to the cue marker. Center bias is thus a tangible phenomenon related to the stimulus.

33.434 THE ROLE OF VISUAL SENSITIVITY AND TASK DEMANDS IN GAZE BEHAVIOR TOWARDS VISUALLY UNCOMFORTABLE STIMULI

Michaela Klimova¹, Jorge Morales; ¹Northeastern University

Visual discomfort is an aversive experience associated with viewing certain stimuli, inducing perceptual illusions and somatic effects such as eye strain or headache. Repetitive patterns such as gratings with mid-range spatial frequencies are considered the most aversive, particularly for highly sensitive individuals who report stronger discomfort and more illusions when viewing such patterns. While the physical stimulus properties causing visual discomfort are fairly well characterized, its behavioral and cognitive consequences and associated eye movement patterns are less well studied. Specifically, do participants actively avoid such stimuli or alter their gaze behavior? We recruited participants with low and high visual sensitivity, as assessed by the Visual Discomfort Scale (VDS). While tracking their gaze, participants performed a visual search task in which they were presented with multiple high-contrast grating patterns of low (0.5 cpd), mid-range/aversive (1-6 cpd) and high (>7 cpd) spatial frequency and required to discriminate the color of a low-opacity target present within one of the gratings in 50% of trials. Additionally, we obtained free-viewing eye movement data and a discomfort rating for each stimulus, and presented a computerized version of the Pattern Glare Test. Higher spatial frequencies were associated with higher reported discomfort and lower accuracy. However, eye movement data showed that higher spatial frequency gratings were also fixated upon earlier, had longer fixation durations, and higher number of repeated fixations, regardless of VDS status. High-VDS participants showed greater pupil constriction in response to the grating patterns compared to low-VDS, suggesting some impact of high sensitivity. In contrast to the search task, participants directed fewer fixations to the high spatial frequency stimuli in a free-viewing setting, compatible with discomfort reduction behavior. Together, our findings suggest that, in a task setting, participants do not avoid aversive patterns but direct more effort to them despite discomfort, regardless of VDS group.

33.435 WHAT YOU LOOK AT IS WHAT YOU USE: IDIOSYNCRATIC FIXATIONS ENCODE IDIOSYNCRATIC IDENTITY-DIAGNOSTIC INFORMATION

Anita Paparelli¹ (anita.paparelli@unifr.ch), Lisa Stacchi¹, Roberto Caldara¹; ¹Eye and Brain Mapping Lab, Department of Psychology, University of Fribourg, Fribourg, Switzerland

Over the past decade, individual differences in visual sampling strategies during facial identity processing have gained considerable interest. Studies have shown that these idiosyncrasies are reliable over time, generalize to real-world behavior, and are directly related to

neural responses. However, it is still unclear whether observers simply fixate different face regions while extracting the same diagnostic information, or whether these differences instead reflect genuine individual variation in the facial information used to process identity. To address this question, we first asked a large sample of young healthy adult Western observers to learn eight facial identities presented with a neutral expression. These identities were then presented with different emotional expressions to ensure genuine identity-level processing instead of image matching, in two complementary paradigms. In the first paradigm, observers recognized these identities under free-viewing conditions while their eye-movements were recorded, mapping each observer idiosyncratic visual sampling pattern. Subsequently, they completed the same task using a Bubbles reverse-correlation technique. By computing classification images from correct trials, we identified the diagnostic facial information each observer relied on. We then quantified the pixelwise correspondence between each observer's fixation pattern and their diagnostic information map to assess whether the information they foveated aligned with the information they used. Our results revealed individual differences in both information sampling and information usage. Crucially, within-observer comparisons showed a strong correspondence between these two components: observers who fixated the eyes, performed correctly when Bubbles masks revealed eye information, whereas those who fixated the mouth required mouth information to succeed. Our findings demonstrate that idiosyncratic sampling strategies reflect a genuine functional difference in the visual information used for effective identification. Our data provide the first direct mechanistic explanation of individual differences in facial identity recognition, further questioning the universality of this critical biological ability in humans.

This study was supported by the grant no 10001C_201145 from the Swiss National Science Foundation awarded to RC and ARR.

33.436 COMPARISON OF EYE POSITION STABILITY BETWEEN HEAD STABILIZATION APPROACHES IN MACAQUES PERFORMING COVERT ATTENTION TASKS

Sabrina Mai^{1,2,3}, Farran Briggs^{1,4}; ¹Thalamocortical Visual Processing Section, National Eye Institute, National Institutes of Health, Bethesda, MD, USA, ²Graduate Partnership Program, National Institutes of Health, Bethesda, MD, USA, ³Neuroscience Graduate Program, University of Rochester, Rochester, NY, USA, ⁴Department of Neuroscience, University of Rochester School of Medicine, Rochester, NY, USA

Electrophysiology studies in vision science require precise measurements of subjects' eye position over time. To ensure measurement accuracy, the standard approach with behaving nonhuman primates (NHPs) has been head stabilization via surgically implanted head posts. In recent years, there has been an increase in efforts to utilize less invasive methods, and develop paradigms that encourage animals to voluntarily engage in behavioral tasks. Here, we quantify eye position stability during fixation across two different covert

attention discrimination tasks, comparing between NHPs whose heads were stabilized via a head post versus a guide for voluntary head positioning. The guide was a 3D-printed mask of each individual NHP's face with a large central cutout to allow for eye-tracking and reward delivery. The head-fixed system was the Precision Positioning System by Thomas RECORDING GmbH. Importantly, identical eye-tracking and reward delivery systems were utilized with both head-fixed and mask-stabilized animals. The two NHPs whose data were analyzed here, were trained to perform the same covert attention tasks, requiring discrimination of changes in attended stimulus contrast, or orientation. We analyzed average eye position before stimulus changes to assess fixation precision for each head-stabilization method. We also assessed whether NHPs displayed any systematic biases in eye position as a function of behavioral requirements (covertly attended location) or outcomes (answer by saccading up or down). We found that neither NHP displayed systematic biases in eye position dependent upon behavioral requirements or outcomes ($p > 0.7$ for all comparisons across both NHPs). Interestingly, the head-fixed NHP displayed slightly larger positional variance compared to the voluntarily head stabilizing NHP, which we attribute to reduced motivation and overall task performance in this animal. These results contribute to growing evidence that less invasive head stabilization approaches are as effective as conventional head-fixing in enabling accurate measurements of eye position.

33.437 FREE-VIEWING BIASES FOR COMPLEX SCENES IN MACAQUES AND HUMANS

Saloni Sharma¹ (saloni_sharma@hms.harvard.edu), Marcel Linka^{2,3}, Maximilian Davide Broda², Benjamin de Haas², Margaret Livingstone¹; ¹Harvard Medical School, ²Justus-Liebig-Universität Giessen, ³Stanford University

Human free-viewing behavior towards natural scenes can be predicted by semantic object information (Xu et al., 2014) with substantial variability across individuals (de Haas et al., 2019) and age (Linka et al., 2025). An open question is whether such gaze biases reflect conserved primate characteristics or if they are human-specific. Here, we analyzed free-viewing behavior of 103 humans (age = 25.17 ± 5.5 ; 72 females) and 8 monkeys (age = 9.13 ± 5.1 ; 1 female) across 200 natural scenes. Humans exhibited a pronounced center-bias with fixation density concentrated near the center of the scenes. Monkeys also showed a center-bias, but it was weaker and less spatially concentrated. Humans made more fixations per image than monkeys (humans: 9.25 ± 0.91 ; monkeys: 7.73 ± 1.88 , $p = 0.0011$), though the mean fixation duration differed weakly across species (humans: $240.52\text{ms} \pm 24.25$; monkeys: $242.33\text{ms} \pm 55.96$; $p = 0.044$). Next, we quantified group differences in the dwell time and first fixations proportion across six semantic categories (Faces, Text, Emotion, Hands, Touched and Edible objects). Despite a history of face deprivation during earlier development in two of the monkeys, both species showed the strongest attentional bias towards Faces. However, this face bias was more pronounced in humans (dwell time:

humans 26.48%, monkeys 21.21%; first fixations: humans 39.31%, monkeys 28.96%; both $p < .001$). Compared to humans, monkeys looked less at Text (dwell: 4.22% vs. 14.15%; first fixation: 4.38% vs. 7.23%, both $p < .001$) and spent slightly less dwell time on Touched (13.08% vs. 9.48%; $p < .001$) and Edible objects (10.32% vs. 9.06%; $p < .05$). Overall, these findings indicate that free-viewing behavior in humans and monkeys is characterized by both shared and species-specific differences in oculomotor behaviors and in the strength and structure of semantic object-directed biases.

33.438 MATH ANXIETY SHAPES VISUAL INFORMATION PROCESSING IN MATH LEARNING: EFFECTS OF A SELF-REGULATION INTERVENTION

Joanna Lewis¹, Caterina Azzarello¹, Gulden Karakok¹, Molly Jameson², Kristin Kang¹, Sara Rohani¹, Katherine Morrison¹; ¹University of Northern Colorado, ²University of Strathclyde

Math anxiety (MA) is a known factor that increases cognitive load during mathematical operations, disrupts attentional control, and can impair performance (Ashcraft & Krause, 2007). We examined the relationship between MA (assessed through a math disposition survey) and visual attention during an arithmetic task in pre-service elementary teachers enrolled in a math content course. Participants completed a two-digit pair addition task in the SR EyeLink 1000+ at the beginning and end of the semester. Students either received an anxiety-targeted self-regulation intervention or served as the baseline comparison group. Across conditions, participants showed high accuracy and improved arithmetic task efficiency over time. Eye-movement patterns indicated reduced cognitive load over the semester: fixation count, fixation duration, and blink duration all decreased, reflecting more efficient visual information sampling during mental mathematical operations. MA influenced several aspects of eye-movement behavior. Individuals with high MA exhibited greater variability in fixation duration and increased total dwell time, along with reduced accuracy gains relative to low-MA students. These results align with attentional control theories, which suggest that anxiety increases cognitive effort and interferes with efficient deployment of attention despite similar arithmetic performance (Calvo & Avero, 2002; Hunt et al., 2015). Pupillometry revealed reduced variability in pupil dilation for students in the intervention condition, a pattern consistent with evidence linking pupil dynamics to cognitive workload (Hepsomali et al., 2019). This suggests that the intervention may modestly impact effort-related aspects of visual processing. Our findings highlight the effects of MA on visual processing and cognitive effort during arithmetic tasks and indicate that the math course experience enhances eye-movement efficiency. While the self-regulation intervention had limited efficacy in altering cognitive load eye-movement markers, despite changes in reported MA, the results highlight the relationship between MA and oculomotor behaviors during mathematical operations.

NSF Award #2235588

Eye Movements: Mechanisms, perception, fixational

**SUNDAY, MAY 17, 8:30 AM – 12:30 PM,
PAVILION**

33.439 HUMANS ANTICIPATE HIGH-LEVEL VISUAL INFORMATION DURING ACTIVE VISION

Luca Kämmer^{1,2,3}, Carmen Amme⁴, Philip Sulewski⁴, Martin N Hebart^{1,2}, Tim C Kietzmann⁴; ¹Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, ²Department of Medicine, Justus Liebig University, Giessen, ³Department of Psychology, Humboldt-University Berlin, ⁴Institute of Cognitive Science, University of Osnabrück

How we make sense of our visual world has largely been studied using paradigms in which people are presented with images while maintaining stable fixation. Yet natural vision is dominated by frequent saccades that bring relevant information into foveal view. Psychophysical work shows that perception changes when we actively move our eyes, but it remains unclear how eye movements reshape visual information processing in the brain. To uncover how incoming visual information is processed before and after fixation onset, we used the AVS dataset, a large-scale MEG and eye-tracking dataset of 5 participants free-viewing 4,080 natural images. To differentiate between low- and high-level visual features, we used the SPoSE dimensions, which are behaviourally derived dimensions that have been shown to account for visually evoked activation in fMRI, EEG, and MEG data. To isolate the effect that the dimensional loadings of each fixation patch have on neural responses over time, we combined them with the Unfold method, which deconvolves neural responses across fixations. Using this approach, we found that neural responses prior to fixation onset were modulated by high-level, but not low-level dimensions. Using beamformer source reconstruction, we also found that this predictive activation is present primarily in hippocampal and medial temporal regions, but not in early visual cortex. These findings suggest that during natural vision, humans actively predict incoming high-level features of saccade targets to facilitate subsequent processing.

33.440 SEGMENTATION AS PROACTIVE CONTROL: GAZE ENTROPY REVEALS RISING UNCERTAINTY PRIOR TO EVENT BOUNDARIES

Sophie Su¹, Yining Ding, Jeffrey Zacks; ¹Washington University in Saint Louis, ²Washington University in Saint Louis, ³Washington University in Saint Louis

Humans segment continuous experience into discrete events, but the cognitive signals that trigger segmentation remain unclear. We tested whether gaze entropy—a measure of spatial dispersion in eye movements—indexes moment-to-moment predictive uncertainty

during passive viewing of naturalistic activities. Eighty-one participants viewed everyday activity videos while gaze was recorded. A separate group provided fine and coarse event segmentation judgments. We derived computational measures of prediction uncertainty and prediction error using Structured Event Memory (SEM) model variants (Nguyen et al, 2024) . Mixed-effects models revealed that gaze entropy was positively related to computational prediction uncertainty, but not prediction error. This suggests that gaze entropy reflects internal prediction uncertainty, distinct from prediction error. In addition, gaze entropy robustly predicted coarse event boundary probability but not fine-grained boundaries. Time-series FIR analyses further showed that gaze entropy reliably increased prior to both fine and coarse boundaries and decreased afterward, revealing a proactive uncertainty signal that precedes segmentation. Fixation duration and saccade amplitude showed weaker temporal relationship with event boundaries. These findings support a proactive control account of event segmentation in which rising predictive uncertainty triggers boundary detection and situation model updating during naturalistic perception.

33.441 TEMPORAL WEIGHTING OF EXTRAFOVEAL EVIDENCE UNDERLIES THE PREVIEW EFFECT

David Melcher¹, Michele Deodato²; ¹New York University Abu Dhabi

During active vision, the extrafoveal preview of a saccade target facilitates recognition of the postsaccadic stimulus when it is valid, but impairs it when invalid, a phenomenon known as the preview effect. It is unclear whether this effect reflects genuine facilitation from valid previews or whether it is inflated by the perceptual cost of invalid ones. Moreover, it remains unknown whether preview effects for object recognition scale with the amount of time that parafoveal information is available before a saccade, as predicted by continuous evidence-accumulation accounts. Across two experiments (behavioral and EEG), we used a gaze-contingent paradigm in which participants executed rapid saccades toward peripheral stimuli that served as valid, invalid, or absent (placeholder) previews. After the saccade, a target stimulus appeared and participants reported whether it depicted a face or a house. Behaviorally, we found that reaction times reflected clear perceptual benefits from valid previews and costs from invalid ones, demonstrating that preview effects arise from both facilitation and interference, similar to findings in reading research. Benefits and costs increased as a function of preview duration, indexed by saccadic reaction times. Interestingly, even absent previews produced time-dependent shifts in performance, suggesting the presence of a temporal weighting of extrafoveal evidence. Overall, this indicates that preview mechanisms are not “one-shot” but instead reflect continuous temporal integration of peripheral information. EEG responses supported this view: early postsaccadic neural responses were modulated by preview validity and preview time, revealing distinct neural signatures for facilitation and interference. Together, these findings show that preview effects in object recognition arise from a time-dependent weighting of extrafoveal evidence. This supports a

framework in which the visual system continuously accumulates and updates sensory evidence across saccades to guide rapid object recognition.

NYUAD Center for Brain and Health, funded by Tamkeen under NYU Abu Dhabi Research Institute grant CG012

33.442 BINOCULAR PERCEPTION OF DEPTH FROM MOTION IS STRONGLY MODULATED BY EYE MOVEMENTS

Callista Dyer¹ (callista@sas.upenn.edu), Victor Rodriguez-Lopez², Johannes Burge^{1,3,4}; ¹Department of Psychology, University of Pennsylvania, PA, ²Institute of Optics, Spanish National Research Council, IO-CSIC, ³Neuroscience Graduate Group, University of Pennsylvania, PA, ⁴Bioengineering Graduate Group, University of Pennsylvania, PA

Differential image motion underlies motion parallax, an important cue for the perception of depth and 3D structure. Eye movements affect the pattern of image motion. It is well-appreciated that eye movements can alter the perceived speed of moving stimuli (Freeman et al., 2010). Here, we examine how eye movements alter the perception of depth from stimuli without and with differential motion, in the presence of interocular delays. The stimulus consisted of two dichoptically-presented strips of horizontally moving bars—one above and one below fixation—that moved either in the same or in opposite directions. The fixation target was stationary, moved left, or moved right. The task was to report—while maintaining fixation—the depth order of the bars relative to fixation. Interocular delays were induced by reducing the luminance in one eye. Psychometric functions were measured. The aim was to find the onscreen delay that nulled the depth percept. These critical onscreen delays should have the same magnitude and opposite sign as the neural delay. Without differential target motion (i.e. both bars move in the same direction), the results are straightforward: target motion causes illusions with stationary fixation (gain=0.0), smooth pursuit (gain=1.0) eliminates them, and anti-pursuit (gain=-1.0) accentuates them. These results suggest that retinal image motion is required for neural disparities to result from interocular delays. In the presence of differential target motion, the pattern of results is more complicated and unexpected. With differential target motion, leftward eye movements produced depth illusions that required critical onscreen advances of the left-eye image, and vice versa. It is as if leftward eye movements delay left-eye visual information processing, and rightward eye movements delay right-eye visual information processing. We discuss attempts to understand the patterns in the data.

This work was supported by the National Eye Institute and the Office of Behavioral and Social Sciences Research, National Institutes of Health Grant R01-EY028571 to J.B., and by the National Institutes of Health Vision Training Program T32EY007035 to C.D

33.443 A COMPUTATIONAL MODEL OF NEURAL ADAPTATION IN THE GAZE HOLDING SYSTEM

Terence L. Tyson¹, Jorge Otero-Millan^{1,2}; ¹University of California, Berkeley, ²Johns Hopkins University

The gaze holding system can adapt when introduced to different input statistics. For example, gaze holding to an extreme eccentricity over a long duration biases the input to one side of the oculomotor range. This adaptation can manifest as rebound nystagmus, a directional drift that brings the eye back to the previously held gaze location followed by a corrective saccade in the opposite direction. From our experimental work, we showed that adaptation can change two aspects of the gaze holding system – the oculomotor setpoint and the leakiness of the oculomotor integrator. We define the setpoint as the eye position with zero velocity and the integrator leakiness as how quickly the eye drifts back to central gaze. Here, we attempt to model these adaptive mechanisms. The proposed model tracks the expected eye position and its variance, representing the setpoint and integrator leakiness, respectively. For the simulation, we provide artificial inputs to the model that mimic our gaze holding experiment, producing a curve that relates eye drift velocity to eye position. We then estimate the setpoint (x-intercept) and integrator leakiness (inverse slope) and compare these simulated results with our empirical findings. We show that our computational model can accurately reveal both changes in the setpoint and in the integrator leakiness given the biased input statistics introduced from prolonged eccentric gaze holding.

National Institute of Health (NIH) R00EY027846 research project grant

33.444 THE IMPACT OF VISUAL FEATURES ON FIXATIONS WHILE VIEWING NATURAL SCENES ACROSS THE LIFESPAN

Xavier H. Pham¹, Jessica N. Goetz¹, Angelica S. Busciglio¹, Lillie C. del Real¹, Mark B. Neider¹; ¹University of Central Florida

During free-viewing of natural scenes, oculomotor deployment is highly biased toward the center. However, when accounting for central fixation bias, image features such as luminance, luminance contrast, and edge density are highly predictive of fixation location (e.g., Nuthmann & Einhäuser, 2015). Studies have suggested that the ability to detect and differentiate low-level features like luminance degrades with age (e.g., Monge & Madden, 2016), but few studies of age-related change control for central fixation bias. While controlling for central fixation bias, we investigated the influence of image features on oculomotor behavior between age groups. In the current study (N = 64; 32 younger and 32 older adults) participants completed a free-viewing task. Scenes were segmented into 8 x 6 regions, and four feature values (central fixation bias, edge density, luminance, luminance contrast) were calculated for each region. For analysis, we utilized a generalized linear mixed model containing ten fixed effects (one intercept; four main effects of image features (central fixation bias, luminance, luminance contrast, edge density); five interactions (age group x intercept, age group x each image feature)) and two

random effects (participant and image). We only analyzed the first three fixations from stimulus onset. The outcome variable was whether a region was fixated. Across age groups, there were no significant differences in the predictive power of luminance ($b = 0.01$), luminance contrast ($b < -0.001$), and edge density ($b = 0.01$) (all $ps \geq .20$); visual features were similarly diagnostic of fixation location between older and younger adults. Interestingly, central fixation bias was significantly less predictive for older adults ($b = -0.25$, $p < .001$). Our findings suggest that older adults process scenes more globally, and despite evidence of age-related degradation of visual feature discrimination, low-level features continue to predict oculomotor behavior.

33.445 ILLUSORY TEMPORAL SEGMENTATION OF TRANSIENT VISUAL INPUT INDUCED BY VOLUNTARY EYEBLINKS

Yosuke Sakamoto^{1,2,3}, Yuki Murai^{1,2}; ¹National Institute of Information and Communications Technology, ²The University of Osaka, ³Japan Society for the Promotion of Science

Visual input is frequently disrupted by saccades and blinks, yet our perception remains remarkably stable. To achieve such stability, visual sensitivity is transiently reduced around the time of these oculomotor events via predictive efference copy signals (saccadic/blink suppression; e.g., Burr, Morrone, & Ross, 1994; Volkmann, Riggs, & Moore, 1980; Wurtz, 2008). While saccadic suppression has been extensively investigated, much less is known about how eyeblinks shape ongoing visual perception. Here we report a blink-induced illusion in which a single visual flash presented shortly before a voluntary blink is perceived as two flashes. On each trial, a random-noise patch masked by a Gaussian annulus (peak radius = 2 dva, SD = 0.5 dva) was presented either once or twice, with each flash lasting 33 ms. Participants were instructed to voluntarily blink at various time points and then report the number of flashes they perceived. We found that a single flash was often reported as two when it was presented shortly before the blink. This illusion exhibited sharp temporal tuning, peaking at ~100 ms prior to blink onset and confined to a narrow temporal window of roughly 100 ms. To test whether mere visual interruption elicits this effect, we introduced an artificial-blink condition using shutter glasses. This illusion was absent under artificial blinks, highlighting the critical role of blink-related predictive efference copy signals in shaping perception. These findings suggest that a brief visual input gives rise to visual persistence in the underlying representation, and that a blink-induced predictive signal engaged prior to blink onset selectively attenuates part of this persistence, producing a bimodal temporal profile that is read out as two events. This illusion highlights eyeblinks, alongside saccades, as a potential active temporal filtering mechanism that may contribute to maintaining perceptual stability across oculomotor events.

33.446 AUTOCORRELATED KDE FOR FIXATION STABILITY: CAPTURING SPATIAL STRUCTURE WHILE

ACCOUNTING FOR TEMPORAL DEPENDENCE IN GAZE DATA

Jaeseon Song¹ (jsong32@central.uh.edu), Kimberly Meier¹;
¹University of Houston College of Optometry

The stability of fixational eye movements is commonly quantified using the bivariate contour ellipse area (BCEA), which assumes a unimodal Gaussian distribution. While sufficient for simple fixation patterns, this assumption fails when fixation produces multimodal or irregular spatial structures. Isoline-based metrics using kernel density estimation (KDE) have recently been adopted to address these limitations, providing greater flexibility in representing the spatial distribution of gaze. However, KDE models gaze samples as independent, ignoring the temporal autocorrelation inherent in natural gaze behavior. We introduce Autocorrelated KDE (AKDE), adapting a method from ecology used for estimating animal ranges to quantify fixation stability. AKDE incorporates temporal dependencies through effective sample size estimation, which adjusts for information redundancy, and locally-adaptive bandwidths that capture underlying spatiotemporal structure. We compared area estimates from AKDE with BCEA and standard KDE using gaze recordings from the GazeBase dataset (N = 298 of 322; Griffith et al., Scientific Data 2021;8:184). AKDE showed a strong correlation with BCEA ($r = 0.84$), yet Bland-Altman analysis revealed systematic differences, with BCEA consistently estimating larger areas (mean difference = $0.16 \log \text{deg}^2$, $p < .0001$). AKDE correlated moderately with KDE ($r = 0.55$), which estimated smaller areas (difference = $0.42 \log \text{deg}^2$, $p < .0001$). KDE and BCEA showed the weakest correlation ($r = 0.29$) and largest disagreement (difference = $-0.58 \log \text{deg}^2$, $p < .0001$). These findings demonstrate that AKDE captures structure missed by both BCEA's normality assumption and KDE's independence assumption. By addressing multimodality and temporal dependence, AKDE provides a flexible, distribution-free metric that combines spatial precision with temporal realism. This method offers a principled alternative to traditional approaches: while BCEA may suffice for controls with tightly clustered, near-Gaussian fixation patterns, AKDE provides greater flexibility for populations with unstable or multimodal fixation, making it well-suited for diverse clinical applications.

This work was supported by NIH grant EY007551 to the University of Houston College of Optometry.

33.447 SYSTEMATIC BIASES OF FIXATIONAL DRIFT DURING FREE VIEWING AND SUSTAINED FIXATION

Evalie Rehor¹ (evalie_rehor@berkeley.edu), Jorge Otero-Millan¹; ¹UC Berkeley

The eyes are never still, even during attempted fixation. Fixational eye movements (microsaccades, drift, and tremor) are typically studied during sustained fixation at central locations. While microsaccades have been characterized during natural viewing, fixational drift during free viewing remains comparatively unexplored. To investigate how fixational drift differs between sustained fixation and free viewing, we

recorded binocular eye movements from head-fixed participants under both conditions ($n = 5$). Fixation targets were presented across a range of eccentricities matching those reached during free viewing. Eye position was measured using dual Purkinje imaging (DPI) eye tracking within the OpenIris framework. To quantify the systematic components of drift, we fit a linear regression model of drift velocity onto eye position. The slope of the regression measures the position-dependent drift velocity bias (arcmin/s per degree of eccentricity), and the intercept of the regression measures the overall position-independent drift velocity bias. Across subjects, drift velocity exhibited a significant position dependence in the vertical direction in both conditions (fixation: $-0.34 \pm 0.07 \text{ arcmin/s/deg}$; free viewing: $-0.71 \pm 0.06 \text{ arcmin/s/deg}$), and in the horizontal direction during free viewing ($-.69 \pm 0.18 \text{ arcmin/s/deg}$). Negative slope indicates that drift velocity goes opposite the current position. Vertical drift intercepts were significantly negative (fixation: $-16.4 \pm 4.7 \text{ arcmin/s}$; free viewing: $-21.5 \pm 4.6 \text{ arcmin/s}$). When comparing conditions, vertical drift velocity dependence on eye position was stronger during free viewing ($p = .02$), and vertical intercepts were more negative during free viewing ($p = .03$). Horizontal slopes and intercepts showed no difference across conditions. These findings show that drift exhibits systematic velocity biases during both sustained fixation and free viewing, with a stronger vertical bias during free viewing.

Grant Support: NIH Grant T32-EY007043. Thank you to Ravi Srinivasan for assistance in building the DPI eye tracker.

33.448 TEMPORAL-NASAL ASYMMETRIES IN OPTOKINETIC NYSTAGMUS OF COMMON MARMOSETS

Yunshan Cai¹, Jude F. Mitchell^{1,2,3}, Sara S. Patterson^{1,2,3,4};
¹Neuroscience Graduate Program, University of Rochester, ²Center for Visual Science, University of Rochester, ³Brain and Cognitive Sciences, University of Rochester, ⁴Flaum Eye Institute, University of Rochester

Optokinetic nystagmus (OKN) is a reflexive eye movement for gaze stabilization with slow-phase tracking followed by a resetting saccade. OKN is highly conserved across vertebrates, with differing patterns of monocular temporal-nasal symmetry depending on eye placement and cortical involvement. In lateral-eyed species, horizontal OKN (hOKN) is primarily driven by nasally-tuned ON direction-selective retinal ganglion cells (ON-dsRGCs) projecting to the nucleus of the optic tract (NOT). In contrast, forward-facing species show abundant cortical projections to the NOT that support binocular integration, improving tracking gain at higher speeds and providing monocular hOKN symmetry. Cortical inputs are thought to be the sole driver of adult primate hOKN. Failed maturation of corticofugal (cortex-to-NOT) projections leads to pathological nystagmus and permanent hOKN asymmetry, consistent with a direct contribution from retinal direction signals. The existence of ON-dsRGCs in the primate retina was only recently confirmed; however, their contributions to hOKN remain an open question. We systematically quantified hOKN in two marmosets. We measured slow-phase gain and temporal-nasal asymmetries

across stimulus velocities under binocular and monocular viewing using full-field drifting gratings and a digital dual-Purkinje eye tracker (Wu et al., 2023). Under monocular viewing, both animals exhibited stronger nasal than temporal gain. Under binocular viewing, gain increased and greater nasal-temporal symmetry was observed. An unexpected multi-modal distribution of slow-phase velocities occurred in 17% of the trials (374/2243). A Gaussian mixture model captured a clear bimodal distribution within these trials. The most prominent bimodal distributions were found at 40°/s, with two stable gain modes (monocular: 20.8±8°/s, 8.1±5.6°/s; binocular: 26.3±5.7°/s, 11.9±5.3°/s). Such bimodality represents a previously undescribed feature of hOKN behavior that could reflect differential cortical and retinal contributions. Overall, the common marmoset hOKN behavior is highly consistent with previous measurements in other non-human primates and humans.

YC and SSP from NIH R00-EY035323 and Research to Prevent Blindness, JFM from NIH R01-EY030998, and P30-EY001319

33.449 TEMPORAL DYNAMICS OF VISUAL PROCESSING DURING FIXATION AS A FUNCTION OF ECCENTRICITY

Cristina de la Malla¹, Martina Poletti^{2,3,4}; ¹Vision and Control of Action (VISCA) Group, Department of Cognition, Development and Psychology of Education, Institut de Neurociències, Universitat de Barcelona, Barcelona, Catalonia, Spain, ²Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, USA, ³Department of Neuroscience, University of Rochester, Rochester, NY, USA, ⁴Center for Visual Science, University of Rochester, Rochester, NY, USA

Although fixation is normally considered as a temporal unit, humans' ability to localize changes in a scene varies over the course of brief fixation periods, and it is differently modulated in time for foveal and peripheral stimuli. While change localization accuracy for peripheral stimuli improves with time after saccade landing, it remains relatively stable for foveal stimuli. Here we examine whether the degree of improvement increases as a function of eccentricity (0.3°, 2°, 6°, or 15°; fovea to periphery). A fixation point was surrounded by four bars presented either all at the same eccentricity, or split between two eccentricities. Stimuli were enlarged as a function of eccentricity to compensate for cortical magnification. Participants (n=10) executed a saccade toward the fixation point, and at random time delays after saccade landing (up to 450 ms), one of the bars briefly (50 ms) changed its orientation. Participants were required to indicate which of the bars moved. The degree of orientation change was kept fixed for each stimulus location throughout the experiment to yield 71% of correct localizations when the change occurred ~200 ms after saccade landing independently of stimulus eccentricity. Results show that localization performance improved over time during the course of fixation, with larger improvements at greater eccentricities (9.6%, 8.6%, 14.0% and 21.8% increases for smaller to larger eccentricities). When stimuli were presented at two different eccentricities, performance was higher for the stimuli closer to the center of gaze

early on during fixation. As fixation time progressed, performance for both near and far stimuli improved following a similar pattern to that observed when each eccentricity was tested in isolation. These results indicate that monitoring stimuli across eccentricities leads to an initial cost for the more eccentric stimuli, in part compensated by a quick recovery over the course of fixation.

This work is funded by grant PID2023-150883NB-I00 from the MCIN/AEI/10.13039/501100011033 to CM and NIH Grant EY029788 to MP.

Eye Movements: Saccades

**SUNDAY, MAY 17, 8:30 AM – 12:30 PM,
PAVILION**

33.450 SACCADIC EXECUTION AMPLIFIES ATTENTIONAL INTERFERENCE IN VWM

Golnaz Forouzandehfar¹ (gforouz@vols.utk.edu), Ilke Kavusturan¹, A. Caglar Tas¹; ¹The University of Tennessee-Knoxville

Previous work suggests competing accounts of how attention interacts with the contents of visual working memory (VWM). Tas et al. (2016) showed that executing a saccade to an irrelevant secondary object (SO) during the retention interval impaired VWM, whereas covertly attending to the SO produced no interference, indicating a dissociation between overt and covert shifts of attention. In contrast, Hamblin-Frohman and Becker (2019) argued that this interference only arises when attention selects an object's features, suggesting feature-based attention (FBA), rather than saccades themselves, drives VWM costs. To distinguish these possibilities, we similarly asked participants to complete either an FBA or a spatial attention (SA) task during the retention interval of a change-detection task. In separate blocks, participants either executed a saccade to the SO (overt attention) or remained fixated throughout the trial (covert attention). Using the same FBA (reporting an arrow direction) and SA (reporting a dot location) tasks as in Hamblin-Frohman and Becker (2019), Experiment 1 replicated their results for covert attention: FBA task significantly decreased VWM performance compared to baseline (where no SO was presented). However, in overt attention blocks, both tasks similarly impaired performance compared to baseline, suggesting that the apparent selectivity of FBA interference can be explained by differences in task demands under fixation, but saccade execution introduces an additional decrement beyond these costs. Experiment 2 modified FBA task (reporting SO identity: triangle or square) to ensure equivalent difficulty as SA task. Both tasks impaired VWM performance relative to baseline. Crucially, this impairment was larger for overt than covert attention. Together, these findings indicate the disruptions created by the covert attention tasks are most likely due to increased dual-task demands. However, overt shifts of attention

produce a separate source of disruption, most likely due to saccade targets being automatically encoded into VWM to aid visual stability.

33.451 WHICH GEOMETRIC CENTER DRIVES CENTER BIAS IN FREE VIEWING? EVIDENCE FROM GEOMETRICALLY OPTIMIZED POLYGON-MASKED NATURAL SCENES

Braham Wassen-Aklilu¹, Ohad Ben-Shahar²; ¹Ben-Gurion University of the Negev

Observers often fixate near the center of images ("center bias"), but it is unclear which geometric "center" attracts early gaze. We address this with two eye-tracking experiments (EyeLink 1000, fixed head) using masked scenes from CAT2000. In the pilot Experiment 1 (N=20, 540 trials), each image was masked with a randomly generated star-shaped polygon having several spatially distinct centers. Trials began with a peripheral cue followed by 4's of free-viewing. Measuring mean distance to each candidate center and their average, we confirmed that the second-fixation (first post-cue) showed the strongest centrality bias that followed the polygon geometry rather than the monitor center or image content. Out of the different possible centers tested, four exerted the strongest and nearly identical mean distances: center of mass (CoM), bounding-box center (BBC), convex-hull center (CHC), and inscribed-circle center (ICC), which Experiment 2 examined more closely. Experiment 2 (N=25, 429 trials) was designed to resolve this ambiguity by jointly controlling image content and polygon geometry within a decision tree architecture. We selected 390 images from six CAT2000 categories spanning strong versus weak center bias (Object, Sketch, Fractal vs. Jumbled, Satellite, Indoor). Each image was masked with one of 13 evolutionarily optimized polygons, carefully structured to isolate and evaluate the predictive power of each of the four candidate geometric centers. Stimuli polygon set comprised a baseline of coincident centers, maximally separated-centers concave-shape, three pairwise-cluster configurations, and eight isolated-center configurations. The same polygons were also shown apictorially to measure purely geometric bias. Pooled results indicate that post-cue-fixations are closest to CoM, suggesting that center bias is dictated by the center of mass of the visible stimulus region. While being modulated but factors such as pictorial structure and semantics, these results provide a precise target for center priors in computational salience models.

33.452 TEMPORAL RECALIBRATION OF SACCADIC-CONTINGENT VISUAL CONSEQUENCES SHIFTS SUBJECTIVE SIMULTANEITY WITHOUT REDUCING CONFIDENCE

Wiebke Nörenberg^{1,2} (wiebke.noerenberg@hu-berlin.de), Pascal Mamassian³, Martin Rolfs^{1,2,4}; ¹Humboldt-Universität zu Berlin, ²Berlin School of Mind and Brain, ³École Normale Supérieure, PSL University, CNRS, Paris, ⁴Cluster of Excellence 'Science of Intelligence', TU Berlin

Temporal recalibration allows the perceptual system to adjust perceived timing when systematic delays arise between actions and their sensory consequences. This process extends to saccadic eye movements, but it is unclear whether the resulting adaptation reflects a simple decision-criterion shift or a genuine perceptual change. If the former is true, observers should doubt their timing judgements. We developed a paradigm to dissociate these two possibilities by jointly measuring how temporal recalibration reshapes the perceived window of saccade execution and observers' metacognitive access to it. Observers executed horizontal saccades across a dynamic noise background while brief, high-contrast, horizontally-extended Gaussian ellipses (4 ms "flash") were presented at various times relative to movement onset. In a matched replay session, observers maintained fixation while horizontal background movements and flashes reproduced the kinematics and timing of each observer's own saccades, reproducing the same stimulus-movement relations. In 60% of trials per block, we introduced systematic temporal delays by presenting the flash predominantly either immediately (20 ms) or with a delay (80 ms) after movement onset, while the remaining trials sampled a wide range of timings (± 250 ms around movement onset). Reports were only collected in this latter subset. Observers first judged simultaneity of flash and movement. They then rated whether their confidence was 'higher' or 'lower' than the average across the experiment so far. In the saccade sessions, reported simultaneity shifted to later times when flashes were systematically delayed. Participants' confidence in the accuracy of their reports remained stable across delay conditions, thereby denoting a genuine perceptual shift in stimulus timing rather than a shift in decision criterion. Interestingly, metacognitive sensitivity was reduced when flashes were perceived simultaneously with the saccade. In the replay condition, neither simultaneity reports nor metacognitive sensitivity were affected by the systematic delay, indicating that the effect requires a motor act.

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. [865715 – VIS-A-VIS]).

33.453 WIDE-FIELD HIGH-RESOLUTION EYE-TRACKING

Anis Idrizovic¹, Ethan Hundley¹, Ruei-Jr Wu¹, Jannick P. Rolland¹, Michele Rucci¹; ¹University of Rochester

Precise measurement of eye movements is essential for understanding the mechanisms that govern visual perception. Recent work has shown that fine oculomotor behavior – the small fixational movements occurring between saccades – actively structures the spatiotemporal input to the retina, profoundly influencing spatial sensitivity, visual acuity, fine pattern vision, and even stereopsis. These findings emphasize the need for eye-trackers capable of resolving eye motion with arcminute precision. At the same time, many natural tasks require accurate tracking over tens of degrees of visual angle. The Dual Purkinje Image (DPI) method introduced by

Cornsweet and Crane (1973) achieves arcminute-level precision by measuring the relative motion of reflections from the cornea (P1) and posterior lens surface (P4). Building on this principle and recent advances in digital imaging, we recently developed a digital DPI (dDPI) eye-tracker (Wu et al., J Vis, 2023) that detects sub-arcminute eye rotations within the central 15° of the visual field. To accommodate diverse experimental demands, we have now generalized this approach into a modular architecture that can be tuned to the desired tracking span. Here we present an implementation with two laterally offset illuminators and optimized magnification, extending the high precision afforded by the dDPI across a 40° range. The design supports two operating regimes: a central region where two P1 and two P4 reflections are visible and both Purkinje pairs contribute to tracking, and peripheral regions where two P1s and a single P4 support robust single-pair tracking following the classical DPI principle. We validated the system with an artificial eye mounted on a high-precision rotational stage and with human observers, confirming both the underlying optical principles and operational performance. The expanded tracking span enables oculomotor measurements with arcminute resolution across a broad eccentricity range, supporting reliable tracking of eye movements over large amplitudes and speeds.

Supported by NIH EY18363 and P30 EY001319.

33.454 IMPAIRED SACCADIC KINEMATICS DURING BINOCULAR VIEWING IN CHILDREN WITH ANISOMETROPIC AND STRABISMIC AMBLYOPIA

Krista Kelly¹, Norah Nyangau², Mina Nouredanesh³; ¹School of Optometry and Vision Science, University of Waterloo, ²Retina Foundation of the Southwest, ³Max Rady College of Medicine, University of Manitoba

Introduction: Amblyopia (lazy eye) is characterized by reduced vision in one eye and impaired binocular vision (suppression, reduced stereoacuity), which can impact ocular motor control. Previous studies report abnormal saccades in amblyopia, but primarily under monocular viewing or in adults. Since amblyopia typically develops by age 7 years, understanding saccadic function in affected children during binocular viewing is critical. **Methods:** We examined saccadic kinematics in 35 children (7-14 years) with amblyopia and 31 age-similar controls. Amblyopic children were classified by etiology (19 anisometropia, 16 strabismus) and stereoacuity status (18 present, 17 absent). Using an EyeLink 1000 binocular eye tracker, saccades were recorded during binocular viewing at 40 cm. Children fixated a central cross, then shifted gaze to dot appearing left or right at 5° or 10° (10 trials per location, 40 total). **Outcomes** included onset latency, peak velocity, primary and corrective saccade gain, gain error, and % trials with corrective saccades. **Results:** No between-eye differences were observed; fellow eye data only are reported. Compared to controls, amblyopic children exhibited delayed onset latency (control, 201±25 msec vs amblyopia, 219±33, $p=0.02$), reduced gain (0.95 ± 0.06 vs 0.89 ± 0.07 , $p<0.001$), increased gain error (0.10 ± 0.03 vs 0.14 ± 0.04 , $p<0.001$), and more corrective saccades ($52\pm16\%$ vs $63\pm16\%$,

$p=0.01$). Strabismic amblyopia showed delayed latency ($p=0.03$), while anisometropic amblyopia showed increased corrective saccades ($p<0.001$). Amblyopic children with absent stereoacuity exhibited delayed latency, reduced gain, increased gain error, and increased corrective saccades ($ps<0.04$) compared to controls, whereas those with stereoacuity present were similar to controls ($ps>0.10$). No associations were found with amblyopic eye visual acuity ($ps>0.17$). **Conclusions:** During binocular viewing, amblyopic children demonstrate impaired saccade kinematics. These deficits are linked to binocular vision status, particularly absent stereoacuity, rather than visual acuity. These findings may have implications for understanding functional deficits, such as slow reading, in children with amblyopia.

NIH EY028224; NSERC RGPIN-2024-03992

33.455 A COVERED EYE REVEALS INDEPENDENT EYE-CONTROL SIGNALS DURING SACCADIC EYE MOVEMENTS

Scott Watamaniuk^{1,2}, Devashish Singh², Stephen Heinen²; ¹Wright State University, ²The Smith-Kettlewell Eye Research Institute

During fixation, a covered eye can deviate either inward or outward from its viewing position, an eye movement known as a phoria. Phorias occur in most people and are usually present in both eyes, but occasionally become pathological in which case the deviation progresses in magnitude to eventually manifest as strabismus. Because covering an eye in normals reveals the phoria, we hypothesize that when uncovered, the oculomotor system must continuously generate an independent signal to align that eye with a target. If so, models that drive binocular eye movements using unitary neural signals (conjugate or vergence) could be incorrect. To reveal if independent signals persist during normal saccade generation, we measured both eyes as observers repeatedly made horizontal or vertical saccades during binocular and monocular viewing. The target was a small (.2 deg) white dot presented on a uniform grey background. Each trial began with a randomized (1-1.5 sec) fixation period during which the static phoria was assessed. For monocular viewing, an eye was occluded with an infrared pass filter to allow binocular eye movement recording (EyeLink 1000+ eye tracker). Under binocular viewing, both eyes had the same trajectory as expected. Under monocular viewing for horizontal saccades, the fixation-phoria in the covered eye was maintained during the saccade. However, saccade magnitudes in the eyes were the same. For vertical saccades the covered eye was also displaced horizontally while vertical alignment was maintained. The results suggest that a unitary signal drives both eyes during saccades, but independent control signals driven by vision correct each eye's phoria during binocular viewing.

NIH 1R01EY034626-01

33.456 PREFERENCE ELICITATION REVEALS STABLE INDIVIDUAL DIFFERENCES IN COSTS OF SACCADIC EYE MOVEMENTS

Tobias Thomas^{1,2} (tobias.thomas@tu-darmstadt.de), Constantin A. Rothkopf^{1,2}; ¹Technische Universität Darmstadt, ²Hessian Center for Artificial Intelligence (hessian.ai)

The human visual system relies on saccadic eye movements to bring information from the environment to the high acuity region of the retina. The decision-making process underlying each shift has predominantly been studied to investigate the effect of stimulus or task properties. Yet, in recent years, the idea that gaze shifts themselves might vary in cost has gained prominence, with research both qualitatively and quantitatively describing these effects, as well as demonstrating that they influence behavior in different contexts. In this work, we propose two preference elicitation experiments that enable us to quantify the effects of most previously discussed spatial saccadic features. These include the amplitude, direction, change in direction, and orbital eye position. In both experiments, subjects repeatedly revealed their preferences by making a forced choice gaze shift to one of two targets on the screen. While the first experiment included forced return saccades to the center of the field of view to ensure that all decisions started there, the second experiment generated a carefully designed sequence of gaze targets across all parameters, thereby additionally investigating the influence of the orbital eye position. All subjects showed a linear preference for smaller saccades; however, for most other features, we found idiosyncratic behavior that was highly consistent across experiments and across days. To quantify the individual differences in preferences and facilitate gaze prediction, we inferred parametric cost functions using a random utility model. The inferred costs of the model align with the empirically observed preferences for all saccadic features, and the model is able to predict approximately 80% of the decisions across participants. In summary, using two preference elicitation experiments designed to investigate factors contributing to saccadic costs, along with a random utility choice model, allows for the characterization of individual differences in gaze selection.

The authors acknowledge the support of the cluster project 'The Third Wave of AI' as part of the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art

33.457 SACCADIC INHIBITION IS MODULATED BY SACCADIC DIRECTION AND DISTRACOR LOCATION

Jiacheng Hao^{1,2}, David Melcher^{2,4}, Marisa Carrasco^{1,3}; ¹Department of Psychology, New York University, New York, USA, ²Department of Psychology, New York University Abu Dhabi, Abu Dhabi, UAE, ³Center for Neural Science, New York University, New York, NY, USA, ⁴Center for Brain and Health, New York University Abu Dhabi, Abu Dhabi, UAE

[BACKGROUND] Saccadic Inhibition (SI), the transient suppression of saccade generation that follows an abrupt visual transient (distractor), is a fundamental mechanism of oculomotor control. This inhibition typically peaks at ~90ms post-distractor-onset and is followed by a brief saccade rebound above baseline. Although SI is known to

depend on the distractor's physical properties and its spatial relation to the planned saccade, it remains unknown whether SI exhibits typical polar-angle asymmetries—better performance along the horizontal than the vertical meridian and along the lower than the upper vertical meridian. We investigated whether these perceptual asymmetries influence the magnitude and temporal profile of SI. [METHODS] We conducted two experiments in which participants made saccades to peripheral Gabor targets. Shortly after target onset, a larger Gabor distractor was flashed at a different location to induce SI. To isolate asymmetries related to saccade direction from those related to distractor location, we manipulated their spatial configuration. Participants made a saccades to one of four cardinal locations while a distractor appeared at one of four intercardinal locations. We quantified the magnitude and temporal profile of SI as a function of target and distractor locations. [RESULTS] SI emerged across all target and distractor locations, with the maximum inhibition occurring ~100 ms after distractor onset. For saccade direction, upward saccades exhibited weaker inhibition than the other cardinal directions, regardless of distractor location. Moreover, overall there was a stronger SI with upper-field than lower-field distractors. [CONCLUSION] The results showed that both target and distractor location modulate the timing and magnitude of SI. These findings suggest that to comprehensively understand the SI phenomenon, and its neural correlates, we should consider saccade direction. To further investigate this issue, we will also present results of a reversed condition-intercardinal target and cardinal distractor locations.

Funded by the Global PhD Fellowship, New York University Abu Dhabi to author J.H.

Face and Body Perception: Wholes, parts, configurations, features

SUNDAY, MAY 17, 8:30 AM – 12:30 PM,
PAVILION

33.458 THE ROLE OF EYEBROWS IN IDENTIFYING FACES WITH SUNGLASSES

Charles C.-F. Or¹, Wei Xuan Ang², Yi Ning Tan², Alan L.F. Lee³; ¹University of Plymouth, ²Nanyang Technological University, ³Lingnan University

The eyebrows have been argued to provide important diagnostic information for identifying faces, especially when the eyes are occluded (e.g., Sadr et al., 2003, Perception). In real-life settings, the unoccluded eyebrows might be useful when sunglasses only occlude the eyes. Here, we tested this hypothesis with multiple face memory tests that involved studying and testing with different combinations of face occlusions, including (1) full, unoccluded faces, (2) faces with small sunglasses occluding the eyes only but not the eyebrows, and (3) faces with big sunglasses occluding both the eyes and eyebrows.

In each condition, observers first studied 21 faces in a randomized sequence (each face presented twice; stimulus duration: 2000 ms/presentation). Afterwards, their task was to determine whether each of the 42 test faces (21 studied and 21 unseen; stimulus duration: 350 ms) was seen before or novel. The occlusion type of the test faces (e.g., with small sunglasses) either matched with the studied faces (congruent conditions; e.g., study small, test small) or not (incongruent conditions; e.g., study full, test small; study big, test small). Results showed impaired performance (d') when a memory test involved any type(s) of sunglasses during study and/or test, compared to studying and testing with full faces. This confirms that the eye region is crucial for face identification. Critically, unoccluded eyebrows (small sunglasses) did not lead to any significant increases in d' compared to occluded eyebrows (big sunglasses). Interestingly, studying faces with eyebrows unoccluded and then testing them occluded (study small, test big) did not significantly lower d' further, while studying faces with eyebrows occluded and then testing them unoccluded (study big, test small) impaired performance. These results suggest that the eyebrows may not necessarily provide useful information for face identification, especially when the eyes are occluded.

Supported by UoP School of Psychology travelling funds to C.O., SG-MOE AcRF Tier 1 Grant (RG59/19) to C.O. & A.L., and NTU-URECA funds to C.O. & Y.N.T.

33.459 A VISUAL BEHAVIORAL STUDY OF CATEGORICAL FACE PATTERN RECOGNITION IN MICE

Yanhong Chen¹ (chen7@uab.edu), Lingna Guo², Ri Jiang², Ke Zhou³, Xudong Zhao^{2,4,5}, Ming Meng¹; ¹The University of Alabama at Birmingham, Birmingham, USA, ²State Key Laboratory of Cognitive Science and Mental Health, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China, ³Department of Psychology, Beijing Normal University, Beijing, China, ⁴CAS Center for Excellence in Brain Science and Intelligence Technology, Beijing, China, ⁵University of Chinese Academy of Sciences, Beijing, China

Can rodents form a categorical face concept, or is their ability to discriminate faces and objects driven solely by visual feature attributions? To answer this question, we trained sixteen C57BL/6J mice with a touchscreen face-object discrimination task and examined their behavior in two experiments. The image stimulus set comprises faces and control objects, which were matched on low-level visual properties. Mice entered unrewarded test sessions once they achieved >70% accuracy on two consecutive days in the training phase. In experiment 1, task accuracy across stimulus conditions was modeled with generalized linear mixed-effects models (GLMMs). We found that, after training, mice exhibited above-chance performance on unencountered new face stimuli. Moreover, this performance decreased when stimuli were inverted or contrast-negated, consistent with the classic effects of these manipulations on categorically “face-specific” tasks in humans. In Experiment 2, sigmoid, linear, and quadratic functions were fitted to assess categorical face-recognition ability, the data were augmented using a simulation method. We found

that along a continuum from low-similarity pareidolia to real faces, mice's performance followed a sigmoidal rather than a linear or quadratic function, which suggests a categorical face concept recognition boundary. Together, these results suggest that mice can form an abstract, configuration-dependent face category and apply it in a human-like manner. The work provides evidence that a face concept can emerge even in a species whose social communication is not primarily visual, paving the way for future investigations on neural and genetic substrates of categorical pattern recognition.

33.460 HOLISTIC FACE PROCESSING EMERGES THROUGH CONTEXTUAL PART ENRICHMENT IN FACE-TRAINED VISION TRANSFORMERS

Srijani Saha^{1,2}, Talia Konkle^{1,2}, George Alvarez^{1,2}; ¹Harvard University, ²Kempner Institute for the Study of Natural and Artificial Intelligence at Harvard University

Seminal work in psychology has demonstrated that faces are recognized via holistic processing, where identities and face parts appear changed in different facial contexts, yet we lack formal image-computable models of these effects. Here we leverage face-trained and ImageNet-trained deep neural network models to provide a computational account of at least one form of holistic processing — updating local part representations based on global context (contextual enrichment of parts). First, we examined whether any of these networks showed classic signatures of human face processing, specifically the inversion effect. To do so, we computed identity prototypes (averaged upright exemplar embeddings) and compared distances of upright versus inverted test images. Confirming prior studies, face-trained models ($n=4$), irrespective of architecture, showed robust evidence of the Inversion Effect, while non-face-trained models ($n=3$) did not. We next leveraged the transformer models, whose patch-based representation and self-attention mechanisms allow us to understand more deeply if and how holistic processing is operating. We locally perturbed one face part (nose) and measured the identity shifts. Critically, by recording activations to each local patch in the context of the original vs. modified nose, we could ask if the identity shift was due to the change in the nose (new information), or due to the impact of the new nose on neighboring parts (contextual influence). We found that contextual effects alone created identity shifts in face-trained, but not object-trained models. This contextual effect dramatically decreased for inverted faces, tying the Inversion Effect observed to the likely disruption of contextual part integration. These findings unveiled how face-identification fine-tuning supports a mechanism where local face parts in upright configurations contextually update each other to build identity representations. This provides a computational formalization of holistic processing, offering a testing ground for other signatures such as the Thatcher and Composite Face Effects.

This work was supported by NSF PAC COMP-COG 1946308 to GAA and funding from the Kempner Institute to TK

33.461 HOLISTIC PROCESSING REQUIRES EXPERTISE, BUT EXPERTISE MAY DEVELOP RAPIDLY

Yuxuan Zeng¹, Ren E. Hentz¹, David E. Osher¹; ¹The Ohio State University

The debate regarding the origins of holistic processing remains unresolved. Some theories posit that this mechanism is domain-specific to faces, whereas others argue it is a domain-general consequence of expertise. Our previous work challenged the strict face-specific view by demonstrating that learning configural information allows observers to process novel non-face stimuli holistically. However, a critical question remains regarding the role of proficiency. Does holistic processing emerge immediately upon learning a categorization rule, or does it depend on the higher level of proficiency associated with expertise? To investigate this, we trained participants to categorize computer-generated abstract line patterns. In the criterion-based condition, training ceased immediately after participants achieved 90% accuracy in the final 30 trials. The training duration for this group varied from 30 to 180 trials and averaged approximately 100 trials. In the fixed-training condition, participants completed a regimen of 200 trials. We ensured that participants in this group also maintained over 90% accuracy in their final block to match the performance level of the first group. Preliminary results reveal a functional dissociation between the two groups. Participants who stopped immediately upon reaching the accuracy criterion showed limited evidence of holistic processing. We observed a composite effect in this group, which suggests this measure might be particularly sensitive, but we found no significant inversion, misalignment, or part-whole effects. In contrast, the group trained for 200 trials exhibited robust holistic processing across all four measures. These findings indicate that simply learning the categorization rule is insufficient. Instead, the visual system requires a degree of proficiency characteristic of expertise to engage holistic mechanisms. Crucially, our data suggest that for certain tasks, the functional transition from learner to expert does not require years of experience but can occur within a short window of 200 trials.

33.462 SOLID SIGHT SOCIAL VISION DATABASE: HIGHLY REALISTIC HUMAN AVATARS FOR VISION RESEARCH

Carl Bunce¹ (c.bunce@reading.ac.uk), Katie L. H. Gray¹, Peter Scarfe¹; ¹University of Reading

Research on face and body perception has benefited enormously from large, well-curated stimulus databases. However, nearly all publicly available sets consist of 2D photographs or videos, which is becoming increasingly at odds with the growing use of immersive virtual reality (VR) in vision science. VR offers many advantages for social vision research over traditional screen-based methods, including naturalistic depth cues, full-field visual presentation, a strong sense of presence and immersion, and access to rich behavioural measures—such as precise tracking of eye gaze and head and body position—that can be used to reliably infer implicit social and perceptual processes. Yet

progress in this area has been slowed by a lack of realistic-looking 3D human stimuli suitable to present in such tasks. Here we introduce the Solid Sight Social Vision Database, a collection of highly realistic full-body human avatars designed for real-time rendering in VR. The database was created by combining state-of-the-art modelling techniques to achieve high geometric accuracy and photorealistic textures of real-life individuals. Participant likenesses were captured using i) a structured-light 3D scanner (Einscan Pro HD), providing sub-millimeter scanning accuracy, and ii) photogrammetric reconstruction from multi-angle photographs. Raw scan data underwent a post-processing workflow involving alignment, geometry and texture cleaning, and scale and position standardization. We additionally captured multiple facial expressions per participant, producing separate meshes for different emotions. The resulting meshes from this pipeline form a high-quality avatar set intended to support VR-based social vision research, as well as any other research applications that require realistic human models for 2D or 3D rendering outside of VR. The database will be made freely accessible to researchers alongside sample MATLAB and Psychtoolbox code to simplify integration into experimental tasks, including immersive VR.

This project was supported by an award from the Leverhulme Trust.

33.463 DOES GAZE DETECTION RELY ON HOLISTIC PROCESSING?

Hyerin Shin¹ (hshin1@hallym.ac.kr), Kyungmin Yang¹, Hoon Choi¹; ¹Department of Psychology, Hallym University

Gaze carries socially meaningful information, and observers readily orient their attention in response to changes in where others are looking. In this study, we examined whether gaze-shift detection depends more heavily on holistic processing or on feature-based processing by assessing how detection efficiency varies with face orientation (upright vs. inverted), stimulus type (graphic vs. real-face), and the gaze-shift angle (small vs. large deviation) in a visual search task. Participants judged whether a target face was present within arrays of upright or inverted faces; all distractor faces displayed a straight gaze, whereas only the target's gaze was shifted a small or large angle to the left or right from straight ahead. Detection efficiency was quantified using the Inverse Efficiency Score (IES), which integrates response time and accuracy. Larger gaze-shift angles consistently yielded higher detection efficiency than smaller angles across all stimulus types. Real-face stimuli further showed higher efficiency in the upright than in the inverted condition, demonstrating a robust face-orientation effect, whereas graphic stimuli exhibited minimal orientation differences. Overall, these findings indicate that gaze-shift detection is not guided solely by local eye features but draws more substantially on holistic facial information, particularly when observers process real faces.

33.464 HOW DOES FACE RECOGNITION IMPROVE? INSIGHTS FROM 12,000 TRIALS WITH THE SAME 10 FACES

Jérémy Lamontagne¹, Vicki Ledrou-Paquet, Anthony Proulx, Laurianne Côté, Isabelle Charbonneau, Caroline Blais, Justin Duncan, Daniel Fiset; ¹Université du Québec en Outaouais, ²Université d'Ottawa

Although studies using the Bubbles technique (Gosselin & Schyns, 2001) typically require large numbers of trials, they have not directly examined how information-use strategies change after very extensive exposure to the same faces. To quantify such changes, we asked four participants to complete 12,000 trials of a face-recognition task over 8–11 sessions. Previous work from our laboratory has shown that better face recognizers rely more heavily on the eye region than poorer recognizers (Royer et al., 2018; Tardif et al., 2019), suggesting that information use may become increasingly optimized during learning and with increasing contact with a given identity. Here, 3D Bubbles were used to investigate how the facial features and spatial frequencies supporting recognition evolved as participants were repeatedly exposed to the same initially unfamiliar identities. Participants were also required to study the faces every day to maintain a high level of contact with those identities and to mimic the kind of repeated exposure that occurs in everyday life. Across all sessions, the eyes remained the most useful information for recognition in all spatial frequencies, but further analyses revealed significant quantitative changes in the use of facial features over the course of the experiment. In particular, participants' accuracy correlated significantly more with the use of both eyes, in the last session compared to the first, specifically in higher spatial scales, suggesting that the eyes become more important as a face becomes more familiar and that learning sharpens an already eye-centered strategy rather than shifting it to other features. Overall, this experiment shows that changes in the information used for face recognition during learning are quantitative and reflect a progressively more efficient use of the eyes. Extending these results to more naturally varying stimuli would help demonstrate the robustness and ecological validity of this finding.

33.465 ADULTS' AND CHILDREN'S FACE PAREIDOLIA IN SYMMETRIC NOISE IS DIFFERENTLY TUNED TO ORIENTATION

Benjamin Balas¹, Emily Westrick¹; ¹North Dakota State University

The detection of face-like structure in noise patterns (face pareidolia) is sensitive to the statistics of those images. In prior work, we reported that adults' rates of reporting face pareidolia in fractal noise images depends on spectral power (steeper power spectrum coefficients reduce pareidolia rates) and orientation content (horizontal or vertical orientation filtering reduces pareidolia rates). The tuning of face pareidolia to these image statistics may reflect sensitivity to statistical regularities in natural images that develops slowly during the school-age years. For example, children only acquire an adult-like bias favoring horizontal orientation energy in face recognition tasks gradually during middle childhood. Presently, we investigated this possibility by measuring face pareidolia as a function of orientation

statistics in adults and a developmental sample. We hypothesized that compared to adults, children would exhibit more consistent pareidolia rates across isotropic, horizontally-filtered, and vertically-filtered noise images. We asked children (aged 5-12) and adults to complete an unconstrained face pareidolia task in which they were shown 4 exemplars each of isotropic fractal noise textures (1/f power spectrum) and orientation-filtered versions of those images with horizontal and vertical passbands applied. Unlike our prior study examining orientation energy and pareidolia, we additionally imposed symmetry across the vertical meridian (Palas & Webster, 2013) to increase overall face pareidolia rates. Participants were given 2 minutes to view each image and circle the pareidolic faces that they saw in this interval. Contrary to our previous results, we found that while orientation filtering of both types did significantly reduce pareidolia rates in our adult sample, horizontally-filtered images yielded more pareidolic faces than vertically-filtered images. Also, consistent with our initial hypothesis, children exhibited stable rates of pareidolia across all conditions. We discuss these results in terms of the emerging template for face detection during childhood and cues that support perceptual grouping.

This work was supported by NSF grant BCS-2338600

33.466 CAN GUIDED HEURISTICS IMPROVE DEEPPFAKE DETECTION?

Lillie C. del Real¹, Jessica N. Goetz¹, Mark B. Neider¹; ¹University of Central Florida

Humans typically classify deepfake face images at near chance levels (50%). However, they also appear to experience synthetic and real faces differently. For example, humans attribute higher ratings of certain perceptual judgments (e.g., attractiveness) to synthetic faces compared to real faces. Could these percepts serve as guided heuristics to aid in distinguishing between real and synthetic faces? Across two experiments, our goal was to identify percepts that were indicative of face type and prompt participants to use those percepts to inform classification of real and synthetic faces. In Experiment 1, 124 participants were randomly assigned to one of four conditions in which they rated real and synthetic StyleGAN2 faces. There were three heuristic conditions where participants rated each face on one facial percept (attractiveness, symmetry, and unusualness) and an unguided condition where participants rated only on the face ground truth (real/synthetic). Participants in the heuristic conditions rated synthetic faces more attractive, symmetrical, and usual than real faces (all p s < .001) without knowing the face ground truth, suggesting these percepts could serve as indicators for deepfake classification. In Experiment 2, 96 naïve participants were assigned to one of the four conditions where they conducted the same rating task as Experiment 1. Participants then used their rating of the face to inform their real/synthetic classification. Results suggested that larger rating differences between real and synthetic faces, specifically in the symmetry, unusualness, and unguided conditions, were associated with higher classification accuracy (all p s < .001). Thus, when making

perceptual judgments of faces, individuals who utilized a wider range of the ratings scale tended toward higher deepfake detection accuracy. While guided heuristics emphasizing face attributes may serve as useful cues for classification, the effectiveness of those heuristics may depend on one's ability to differentiate between synthetic and real faces in perceptual judgment space.

33.467 RED FACILITATES ANGER PERCEPTION ACROSS CULTURES

Lijing Guo^{1,2} (lijing.l.guo@jyu.fi), Hongjin Sun³, Taosheng Liu⁴, Dan Nie², Penglan Liu², Weizhen Xie⁵, Chaoxiong Ye^{1,2}; ¹School of Education, Anyang Normal University, Anyang, China, ²Department of Psychology, University of Jyväskylä, Jyväskylä, Finland, ³Department of Psychology, Neuroscience, and Behaviour, McMaster University, Hamilton, Ontario, Canada, ⁴Department of Psychology, Michigan State University, East Lansing, United States, ⁵Department of Psychology, University of Maryland, College Park, MD 20742, USA

Colour is not only a perceptual attribute but can also carry emotional implications. Previous research has shown that people tend to associate specific colours with particular emotions, with red being associated with anger. However, the implication of red could possibly also be culturally specific as in Chinese culture, red is often associated with positive emotions such as joy and celebration. The present study aimed to examine whether, within a bivalent cultural context, red activates both associations simultaneously (coexistence hypothesis) or engages one over the other (competition hypothesis). Experiment 1 examined both Chinese and Canadian participants in a facial emotion identification task using either White and East Asian facial images. Participants were asked to rapidly identify angry and happy emotions presented against a red, green, or gray background. Using the reaction time measure, we found that a red background significantly facilitated the identification of angry emotion and had little effect on the identification of happy emotions among participants from both countries. These findings support the competition hypothesis rather than the coexistence hypothesis. Experiment 2 replicated this pattern in a new sample of Chinese participants under reduced perceptual salience of the display. Experiment 3 further replicated the effect despite variation of stimulus factors, such as background size (full-screen vs. local). Overall, we consistently demonstrated red-anger association across three experiments, which was not moderated by the racial profile of the emotional faces, participants' cultural background, or sensory-level factors. These consistent findings highlight the robustness of the red-anger association in visual-emotion processing. These consistent findings highlight the robust impact of conceptual association between color and affect (e.g., red and anger) in visual-emotion processing - a potential indication of cognitive penetrability of visual perception.

Acknowledgements This work was supported by the Research Council of Finland (former Academy of Finland) Academy Research Fellow project (#355369 to Chaoxiong Ye) and Finnish Cultural Foundation (#231373 to Chaoxiong Ye).

Decision Making: Perception 2

**SUNDAY, MAY 17, 8:30 AM – 12:30 PM,
PAVILION**

33.468 LATENT PERCEPTUAL STATES MODULATE SEQUENTIAL BIASES IN DERMATOLOGY JUDGMENTS

Katrina A. Wolters¹, Veith Veilnhammer², Jefferson Ortega¹, Zhihang Ren¹, Haley Frey¹, David Whitney¹; ¹University of California, Berkeley, ²Max Planck UCL Centre for Computational Psychiatry and Ageing Research

Clinician accuracy in dermatological malignancy judgments is essential for optimal patient outcomes. Previous work has identified serial dependence in skin lesion judgments (Ren et al., 2023), and recent findings suggest that human perception alternates between two distinct modes: an internal mode, characterized by decreased attentional engagement and attractive perceptual biases (e.g. serial dependence), and an external mode, where task engagement is heightened and perception is aligned more closely with ground truth (Veilnhammer et al., 2023). Here, we investigated if sequential biases were present in a 2AFC lesion malignancy classification task, and whether machine learning tools like Hidden Markov Models (HMMs) could identify distinct perceptual states. We analyzed over 750,000 lesion judgments made by medical students and residents on an online app, and found significant serial dependence in malignancy judgments ($p < 0.01$), replicating Ren et al. (2023). We extended these results by fitting an HMM to the behavioral data, finding that a two-state model was strongly favored over a one-state model (log Bayes Factor = 6681.36). The two inferred states were consistent with internal and external perceptual modes: the internal mode was associated with significant serial dependence, whereas the external mode showed a repulsive sequential bias, and exhibited 14% higher classification accuracy than the internal mode. Further, given that serial dependence is feature-similarity tuned (Manassi et al., 2023), we explored the impact of sequential lesion similarity (computed using Google's Derm Foundation machine learning model) on malignancy judgments, and found that sequential lesion similarity significantly predicts perceptual bias, with net serial dependence increasing as image similarity increases. Taken together, these results suggest that there may not be one singular perceptual bias or decision-making state in effect when observers make sequential malignancy judgments; instead, sequential lesion similarity and fluctuations in perceptual mode both play a substantial role in bias and diagnostic accuracy.

Supported in part by NIH R01CA236793.

33.469 INCONSISTENT ORIENTATION DISCRIMINATION INDICATES ELEVATED INTERNAL NOISE IN SCHIZOPHRENIA: A DOUBLE PASS STUDY

Jose Reynoso^{1,2,3}, Diana Freed^{1,2,3}, Maya Glasman^{1,3}, Brian P. Keane^{1,2,3}, Dujie Tadin^{1,2,3}; ¹University of Rochester, ²University of Rochester Medical Center, ³Center for Visual Science

Vision is inherently noisy. This noise principally arises from external properties of stimuli (external noise) and brain activity unique to each individual (internal noise). There is growing evidence that increases in internal noise are linked to natural changes like aging (Bogfjellmo, 2013) as well as to psychiatric and neurological conditions (Wolff, 2022). This includes schizophrenia, where increased internal noise might help explain perceptual and cognitive deficits. Here, we assessed internal noise using a double pass technique and hypothesized that people with schizophrenia would show higher levels of internal noise compared to healthy controls. To test this, we recruited 21 healthy controls and 21 schizophrenia patients to complete an orientation discrimination task using virtual reality, with data collection ongoing. In the double pass technique, subjects saw repeated instances of noisy and/or ambiguous stimuli. A noise-free observer will provide the same response to repeated instances of identical stimuli, regardless of a correct response. As internal noise increases, so does the likelihood that responses to repeated instances of identical stimuli will differ. In our study, subjects were shown 201 unique, noisy Gabor patches twice and tasked to verbally report the orientation of the signal grating (left or right). After 67 trials, subjects were shown a second pass of the previous 67 stimuli (402 trials total). Internal noise was estimated based on consistency of responses to repeated stimuli with a correction for baseline accuracy. Our preliminary results revealed marginally elevated internal noise in patients compared to controls ($p=0.07$, Cohen's $d=0.57$), reflecting less consistent responses across repeated stimuli ($p=0.02$, Cohen's $d=0.74$). Patients also had lower task accuracy ($p=0.02$, Cohen's $d=0.75$). We are currently investigating whether elevated internal noise can explain deficits in visual working memory as well as proprioception- and vision-guided reaching, potentially offering insight to how internal noise may affect perceptual and cognitive function.

33.470 SINGLE-TRIAL LATENT NEURAL POPULATION DYNAMICS IN FRONTAL EYE FIELD DURING DECISIONS ABOUT BRIEF STATIC STIMULI

Hiroichi Yoshida^{1,2,3,5}, Zhongxuan Wu^{1,3,4,5}, Xue-Xin Wei^{1,2,3,4,5}, Eyal Seidemann^{1,2,3,5}; ¹Department of Neuroscience, The University of Texas at Austin, ²Department of Psychology, The University of Texas at Austin, ³Center for Perceptual Systems, The University of Texas at Austin, ⁴Center for Learning and Memory, The University of Texas at Austin, ⁵Center for Theoretical and Computational Neuroscience, The University of Texas at Austin

Introduction: Perceptual decisions are thought to arise from neural dynamics in decision-related areas such as the frontal eye field (FEF) that transform sensory evidence into choices. These dynamics have been studied using stochastic stimuli that promote evidence accumulation. However, it remains unclear whether static stimuli presented on the timescale of a single natural fixation also engage an

integrative, continuous process (e.g., drift-diffusion), which may benefit performance in the presence of internal noise, or a near instantaneous, discrete state transition (stepping). Methods: Using Neuropixels probes, we recorded spiking activity from 244-392 single- and multi-units in the FEF of two monkeys performing a coarse orientation discrimination task, in which they reported a binary saccadic choice based on the orientation of a briefly presented Gabor patch at varying contrasts. We analyzed the response properties of these neurons at the level of individual units and neural populations. Results: Our analysis of individual units revealed three subpopulations: choice-selective, stimulus-selective, and units selective for both stimulus and choice. We next trained linear classifiers on the population activity to predict choice. Trial-averaged decision variable (DV) traces for the two choices gradually diverged over time. While this result suggests that choice-predictive information increases as trials progress towards saccade onset, it can be consistent with both drift-diffusion and stepping dynamics. To further adjudicate between the two, we developed a computational approach based on Hidden Markov Models. From single trial DV traces, we inferred the number of latent states and the transition probabilities between them. Our results show that the data are best explained by models with a substantial number of states and smooth state-to-state transitions. Conclusions: Our current results suggest that neural population responses in FEF during decisions about brief static stimuli are more consistent with drift-diffusion dynamics than stepping dynamics.

33.471 RECURRENT NEURAL NETWORKS WITH IMPERFECT MEMORY SHOW CONFIRMATION BIAS

Marvin Maechler¹ (marvin.maechler@gmail.com), Ansh Soni¹, Alan Stocker¹; ¹University of Pennsylvania

Decision-making often involves categorical choices that can bias subsequent judgments (e.g., confirmation bias). We have previously shown that when observers categorize a stimulus before estimating it, their estimates are systematically biased. This bias is not due to poor calibration or insufficient motivation because it persists in the presence of detailed feedback (enabling learning) and monetary incentives (enhancing motivation). This suggests that confirmation bias reflects deliberate behavior, although its underlying motivation remains unclear. We used recurrent neural networks to test whether confirmation bias emerges naturally from resource-rational behavior. We artificially imposed memory limitations on the networks by adding varying amounts of noise to every neuron's activation function at each time step. We trained the networks to perform the same psychophysical task used in our previous experiments on human observers. The task in these experiments was to infer the mean of a stimulus generating distribution based on a sequence of stimulus samples. The observer was asked to categorize the stimuli relative to a decision boundary before estimating the generative mean. Importantly, the networks were trained solely to minimize estimation error without explicit penalty for inconsistency or categorical errors. In

the presence of noise, the network model with access to its own past choice learned to rely on it when making an estimate, thereby introducing confirmation bias. Despite the bias, the model achieved lower root mean squared error than the unbiased comparison model without access to its own choice. The performance advantage demonstrates that incorporating the categorical choice as a form of compressed memory preserves some of the information that is otherwise lost due to memory noise. Our results show that confirmation bias can emerge from computational constraints that are ubiquitous to neural systems (i.e., noise). In conclusion, choice-induced bias may reflect efficient neural computation rather than cognitive limitations.

This work was supported by the NSF CRCNS grant IIS-1912232 to A.A.S.

33.472 CONTRAST-RELATED UNCERTAINTY IS REPRESENTED PROBABILISTICALLY IN VISUAL CORTICAL ACTIVITY

Yuxuan Dai¹, James Cooke¹, Ilona Bloem^{1,2}, Sam Ling³, Janneke Jehee¹; ¹Donders Institute for Brain, Cognition and Behavior, Radboud University, ²Netherlands Institute for Neuroscience, ³Boston University

Percepts of our visual environment are uncertain, due to noise and ambiguity in our sensory signals. While previous work has shown that this uncertainty can be extracted from cortical activity as the width of a probability distribution, these studies mainly focused on spontaneous fluctuations in uncertainty. Here, we ask whether experimentally manipulated sources of noise, such as changes in stimulus contrast, are similarly reflected in visual cortical activity. Participants viewed gratings of random orientation (0-179 degrees), while their brain activity (areas V1-V3) was measured with fMRI. Observers reported the orientation of each grating on a trial-by-trial basis. Stimulus contrast was varied to manipulate uncertainty (contrast levels: 1%, 5%, 10%). Using a generative model-based analysis (van Bergen et al., 2015; van Bergen & Jehee, 2021), we extracted from each trial of cortical activity the posterior distribution of stimulus orientation. The width of the distribution was taken as a measure of uncertainty on that trial. We found that the degree of uncertainty about orientation in visual cortical activity was reliably greater for lower levels of stimulus contrast. Moreover, these contrast-related changes in decoded uncertainty predicted the observer's behavior: behavioral orientation judgments were less precise when stimulus contrast was lower, and the width of the decoded distribution was broader. Interestingly, the decoded distribution also tracked spontaneous trial-by-trial fluctuations in the precision with which the viewed orientation was encoded in cortical activity. Further analyses revealed that these flexible across-noise adjustments in uncertainty appear to rely on mechanisms of neural gain. Taken together, our results suggest that the neural code for uncertainty is remarkably robust across diverse sources of noise – whether that noise arises from the physical environment or is intrinsic to the observer.

33.473 DYNAMICS OF VISUAL, VESTIBULAR, AND DECISION SIGNALS IN MULTISENSORY CORTICAL AREAS DURING A REACTION-TIME HEADING DISCRIMINATION TASK

Yueh Chen Chiang¹ (j1155665@gmail.com), Yared R. Haile¹, Christopher R. Fetsch¹; ¹Zanvyl Krieger Mind/Brain Institute & Solomon H. Snyder Department of Neuroscience, Johns Hopkins University

In natural environments, animals must integrate information from multiple modalities to guide action, while also balancing decision speed vs. accuracy. Although the neural basis of multisensory integration has been extensively studied, previous work mainly used fixed-duration tasks which limit the ability to isolate the temporal window of decision formation on a trial by trial basis. We trained rhesus monkeys to perform a reaction-time (RT) heading discrimination task based on vestibular cues (translational inertial motion), visual cues (optic flow), or both. Animals reported a binary heading judgment with a saccadic eye movement as soon as they were ready, followed by a binary confidence rating with a second saccade (post-decision wagering). Neural recordings were obtained from multisensory cortical areas MSTd and VPS using 32-channel linear arrays (N=440 units in 2 monkeys) or Neuropixels probes (N=932 units in 1 monkey). At the population level, both areas showed clear condition-dependent structure, with stimulus, choice, and confidence components separable in low dimensional space. However, single-unit stimulus encoding was notably weak in the RT task. Within the time frame of individual decisions, neurometric thresholds averaged between 55 and 90 degrees, 2.5-fold higher than during a passive fixation task (after controlling for stimulus duration) and 18-fold higher than psychophysical thresholds. Population decoding confirmed the relatively poor fidelity of heading signals in the RT context, and revealed that choice and confidence signals emerged only in the final few hundred milliseconds before the first saccade. The results suggest that either (1) momentary evidence for speeded heading decisions resides elsewhere in the brain, or (2) perceptual readout is targeted to a small subset of neurons and/or a brief window before commitment to the choice. We consider these results in the context of alternatives to the standard evidence accumulation framework implemented within a circuit model of the visual-vestibular cortical network.

Supported by NIH-NINDS (RF1NS132910), the E. Matilda Ziegler Foundation for the Blind, the Whitehall Foundation, the France-Merrick Foundation

33.474 SCENE STRUCTURE PREDICTS PERCEPTUAL DECISIONS IN NATURALISTIC DETECTION TASKS

Jun Yang^{1,2} (junyang1728@gmail.com), Tiziana Vercillo², Teresa Emma Cutrona², Simone Azeoglio^{1,3}, Giandomenico Iannetti^{2,4}, Peter Neri^{1,2}; ¹Ecole Normale Supérieure-PSL, France, ²Italian Institute of Technology, Italy, ³Institut de la Vision, France, ⁴University College London, UK

We recognize objects effortlessly, even when embedded within complex natural scenes. It remains unclear how this feat is achieved—in particular, we do not know how scene structure shapes perceptual decision-making. We tackle this issue by integrating four complementary approaches: controlled psychophysics, augmented reality (AR)-based naturalistic paradigms, deep neural network (DNN) modeling, and electroencephalography (EEG) decoding. All paradigms involved the same detection protocol: participants were asked to report the absence/presence of a target probe (e.g. short segment) randomly placed within the scene. We trained neural networks to predict psychophysical responses (correct/incorrect) generated by humans under various manipulations (probe type, image category, task structure, and timing of spatial cues). DNNs achieved reliable prediction using only contextual information (scene background), without any information about target configuration. Network predictability was greater for conditions under which humans were presented with greater spatial uncertainty, highlighting the role of global scene context in determining the operation of local perceptual processes. AR experiments extended these findings to real-world, naturalistic behavioral settings. Under some conditions, elementary image features (edge density and texture entropy) supported reliable classification of correct/incorrect responses via traditional machine learning methods, providing an interpretable account of perceptual variability. EEG recordings revealed neural signatures of perceptual outcome prediction during context-only presentations, and fusing EEG information with image-based information significantly improved decoding accuracy, showing that neural activity captures aspects of scene structure that modulate probe detection outcomes. Together, these converging results establish that natural scene statistics systematically modulate local perceptual processes independently of target/task configuration, and provide a unified framework for linking environmental structure, cortical dynamics, and perceptual decision-making in real-world detection tasks.

33.475 A BAYESIAN NETWORK APPROACH TO EXAMINE HOW TODDLERS WITH AND WITHOUT AUTISM SPECTRUM DISORDER LEARN FROM WATCHING NATURALISTIC VIDEOS OF SAME-AGED PEERS

Caius Gibeily^{1,2} (cgibeil@emory.edu), Warren R. Jones^{1,2}, Sarah R. Shultz^{1,2}; ¹Emory University, ²Marcus Autism Center

Background—Autism spectrum disorder (ASD) is characterized by altered social and communicative functioning. Eye-tracking can predict diagnostic status and developmental outcomes in children with and without autism, yet most approaches quantify what children look at rather than how past viewing behavior shapes future viewing. A method for quantifying this dynamic, moment-to-moment “learning” from visual experience could offer new insight into cascading developmental processes. Methods—We developed a Bayesian network (BN) approach to model how toddlers’ viewing history shapes subsequent viewing states. TD toddlers (n=150) and toddlers with ASD (n=216) were eye-tracked while viewing movies of toddlers

playing in a daycare. Visual targets were annotated using a data-driven procedure and semantically clustered using OpenAI’s CLIP model. Dynamic viewing patterns were then hierarchically clustered to condense moment-by-moment variation in viewing patterns and BNs fit for each movie and group. Results—TD and ASD toddlers differed in the distribution of targets viewed, with TD viewers showing greater attention to social-action/facial targets and ASD viewers to object-action/body-image targets ($\chi^2(df=3, n=133)=31.5, p<0.05$). BN structures, constructed from edges found in $\geq 70\%$ of bootstrapped samples, revealed both immediate and latent dependencies in future viewing states. Models incorporating prior viewing history improved prediction of subsequent states (TD: 12.5 ± 3.7 ; ASD: 15.7 ± 3.5 ; metric: percent accuracy above chance). Conclusion—Toddlers with ASD exhibit distinct patterns of contingent visual engagement when viewing naturalistic social scenes. By quantifying the extent to which future viewing behavior is changed by past viewing behavior, this framework provides a tractable metric of learning from visual experience. Such information may identify what is and is not being learned versus what is not, informing intervention targets such as appropriate levels of attentional or semantic complexity. More broadly, this approach offers a principled way to examine how moment-to-moment attentional biases exert cascading influence not only within a viewing session but potentially across developmental time.

33.476 STUDYING PERCEPTUAL DECISIONS IN ISOLATION FROM BEHAVIORAL DECISIONS USING REFLEXIVE EYE MOVEMENTS

Jan Brascamp¹, Morgan Gaston¹; ¹Michigan State University

Perceptual decisions are central to decision-making research: by asking participants to select a simple response to a controlled stimulus, the decision process can be distilled to a tractable core. However, even classic two-alternative-forced-choice tasks—e.g. judging the direction of random-dot motion—include processing stages downstream from perception. A participant may perceive leftward motion, yet still must select and execute the corresponding motor response. These perceptual and post-perceptual stages may be serial or intertwined, but either way the decision is not purely perceptual. We leverage perceptual bistability and reflexive eye movements to study a perceptual decision process in isolation. Participants viewed an ambiguous structure-from-motion stimulus that can be perceived as rotating leftward or rightward. Perception at stimulus onset therefore reflects the visual system’s choice between two interpretations. To assess this decision without engaging response selection, we recorded gaze direction, which reflexively tracks perceived motion for such stimuli. Across a series of experiments, we first confirmed that gaze direction at stimulus onset reliably distinguishes the perceived rotation. We then asked whether gaze could reveal more fine-grained aspects of the decision process. Using manual reports, we replicated the prior finding that decisions are slower when they are more difficult: when the two percepts are a priori equally likely, participants take longer to report their percept than when

one percept is strongly favored. This indicates that balanced percepts yield a more protracted decision process. Critically, gaze-following movements showed the same pattern: they were delayed on trials with relatively balanced percepts. This demonstrates, first, that the slowing arises within the perceptual decision itself rather than solely in response selection, and second, that gaze provides a suitable index of fine-grained decision dynamics, including decision duration. These results lay the groundwork for using gaze as a tool to study perceptual decision making independently from overt behavioral responses.

33.477 SIGNAL DETECTION UNDER MULTIPLE OBSERVATIONS

Zahra Hussain¹ (zahra.hussain@plymouth.ac.uk), Patrick Bennett²,
¹University of Plymouth, ²McMaster University

How do observers integrate information across observations? In detection tasks where the signal is known exactly, or the stimulus is selected randomly from N signals on every observation, d' for an ideal observer improves with the square-root of the observation number. However, if the stimulus is selected randomly on observation 1 and remains constant on subsequent observations, Swets & Birdsall (1978) showed that the initial increase in d' produced by optimal integration exceeds square-root law. We evaluated this prediction by measuring d' in 2IFC detection tasks comprising four observations per trial. The stimuli were band-limited noise textures or 1 cy/deg gratings that varied in orientation. In the no-uncertainty condition, the stimulus was the same in all observations and trials. In the composite condition, the stimulus varied randomly across trials but was the same in all observations within a trial. In the simple uncertainty condition, the stimulus varied randomly across all observations and trials. The effect of observations varied across conditions. In the no-uncertainty condition, d' followed square-root law for gratings but was sub-optimal for textures. In the composite condition, the increase in d' between observations 1 and 2 exceeded square-root law for gratings but not textures. Hence, Swets & Birdsall's prediction was supported for gratings but not textures, and is consistent with the absence of stimulus uncertainty effects for textures shown by Hussain et al (VSS 2025). In the simple uncertainty condition, integration was suboptimal for the first two observations and followed square-root law thereafter. These results are consistent with the idea that the effects of uncertainty depend strongly on the visual stimuli and the way that uncertainty is manipulated across observations.

SUNDAY AFTERNOON POSTERS IN BANYAN BREEZEWAY

Visual Memory: Long-term memory

SUNDAY, MAY 17, 2:45 – 6:45 PM, BANYAN BREEZEWAY

36.301 WHAT MAKES A MATERIAL MEMORABLE?

Yijin Wang¹ (yw1695a@american.edu), Chenxi Liao¹, Bei Xiao¹,
¹American University

From moment to moment, we encounter countless materials. Materials exhibit a wide range of appearances varying within and across categories. The same material can take many forms. What makes an image of material memorable? Are some materials easier to memorize than others? Memorability has been conceptualized as a stimulus-driven, stable statistical property that is consistent across observers. Understanding the visual memorability of materials is a crucial step toward revealing how the brain efficiently processes and represents visual information about them. Here, we investigate the visual memorability of materials using a traditional memorability recognition task on the STUFF dataset, which comprises 3,514 images from 200 material categories. Individual participants ($N = 11$) viewed a continuous stream of images and pressed a key whenever they recognized an image they had encountered; over the course of the experiment, each participant viewed the entire dataset. We computed material memorability for each image as the difference between the hit rate (proportion of participants who made a correct recognition) and the overall false alarm rate. Despite the challenging dataset, participants showed moderate accuracy (target: 56.25%; overall: 87.76%; false alarms: 8.09%; mean memorability: 48.17%) with consistent patterns—Spearman correlations between randomly split groups were high, confirming stable stimulus-driven memorability ($r=0.33$). Results from GLMM showed that material category strongly predicted memorability ($p<0.001$). Snow, tooth, and pearl were highly memorable; garnet and chlorine were least memorable. Conversely, low- to mid-level visual features extracted from pretrained neural networks (LPIPS, DreamSim) correlated only weakly with memorability. Our results suggest that memorability, through reflecting the statistical regularities of the environment, can serve as an intrinsic dimension in constructing the representation of materials. In addition to low- to mid-level visual features, an efficient store of category-driven information is fundamental for future recognition.

36.302 MODULATION OF VISUAL LONG-TERM MEMORY BY FEATURE-BASED ATTENTION

Shruti Nishith^{1,2}, Nicholas B. Turk-Browne^{1,2}; ¹Department of Psychology, Yale University, ²Wu-Tsai Institute, Yale University

Selective attention has well-known effects on perception and working memory, but its influence on long-term memory remains underexplored. Here, we characterize how feature-based attention modulates the encoding of competing visual features of an object into long-term memory. Objects were created by sampling shape and color values from circular, perceptually uniform feature spaces. Each color-shape object was paired with a unique face image to form an associative memory. Participants completed two sessions in which they were instructed to attend to the color or shape of the objects, respectively. Each session started with six rounds of learning, with each round consisting of three phases: (1) study phase to memorize

associations between the attended feature and the face, (2) feature test phase to probe memory for the attended feature using a shape/color wheel, and (3) face test phase to probe memory for the associate face. Across learning rounds, participants significantly improved their memory performance in the feature test and the face test. The session ended with a final test phase in which participants were cued with a face and had to recreate both the color and the shape of the associated object from long-term memory (regardless of which feature they had attended to during learning). Participants showed lower mean error for color memory when color had been attended versus unattended during learning and lower mean error for shape memory when shape was attended versus unattended during learning, resulting in an interaction between attended feature and the feature memory. There was also a main effect of feature type, with greater memory accuracy for color than shape. Objects varied in the similarity of their features to those of other objects, allowing for additional analyses of how feature-based attention modulates memory competition and interference. These findings suggest that task-relevant features are prioritized in visual long-term memory.

36.303 THE ROLE OF STIMULUS PROPERTIES IN DIRECTED FORGETTING OF NATURAL SCENES

Jia Gu¹ (jiag6@illinois.edu), Laurel B. Walther, Huiyu Ding, Curtis Rogers, Lili Sahakyan; ¹University of Illinois

People regularly make intentional efforts to forget irrelevant or unwanted information. While researchers have consistently observed robust forgetting of verbal information (e.g., words), there has been mixed success in capturing intentional forgetting of visual information (e.g., natural scenes), often observing minimal or unreliable results. To understand intentional forgetting of visual items, this study investigated whether stimulus properties like list structure or image representativeness influence intentional forgetting. Firstly, we implemented mixed and pure study lists using natural scenes to explore the role of stimulus structure. At learning, participants viewed 40 natural scene images, each followed by a remember or a forget cue. In the pure-list condition, all studied images came from the same category, whereas in the mixed-list condition, images were equally selected from five categories. Participants then completed an old-new recognition test with 40 old and 40 new images. Successful directed forgetting would result in better recognition of remember-cue items compared to forget-cue items. Results showed that directed forgetting was obtained in the pure list condition but not in the mixed-list condition, suggesting that intentional forgetting of visual information is more likely to occur when the studied images belong to the same category. Additionally, within the pure-list condition, we observed that different image categories exhibit different directed forgetting effects. We then followed up on whether representativeness of images explain these category-specific directed forgetting. In a second experiment, we used naturalistic scenes with goodness-of-exemplar ratings and manipulated the type of stimuli (good vs. bad exemplars). At learning, participants viewed 30 good and 30 bad exemplars of natural scenes,

each followed by a forget or a remember cue. Then, they took an old-new recognition test with 60 old and 60 new images. This design clarifies the role of scene representativeness in directed forgetting, offering new perspectives in mechanisms of forgetting visual information.

36.304 SENSORY BASED MODULATIONS ON VISUAL MEMORY ARE PRESERVED IN OLDER ADULTS: CONVERGING EVIDENCE FROM ONLINE AND IN-PERSON STUDIES

Hofit Azrad^{1,2}, Eden A. Eliyahu³, Olga Kreichman^{1,2}, Shaimaa Masarwa^{1,2}, Boaz M. Ben-David^{3,4,5}, Sharon Gilaie-Dotan^{1,2}; ¹School of Optometry and Vision Science, Bar-Ilan University, ²Gonda Multidisciplinary Brain Research Center, Bar-Ilan University, ³Baruch Ivcher School of Psychology, Reichman University, ⁴University of Toronto, ⁵KITE Research Institute, University Health Network

Sensory visual degradation is one of the most well-known implications of aging. Theories on compensation mechanisms in aging suggest that with sensory degradations, processing shifts to prioritize more preserved non-sensory sources of information such as semantic and contextual information. Therefore, it is unclear whether image size changes (evident in profound sensory changes to the retinal image) that significantly affect memory in young adults, would also affect memory in older ages. Here we first tested this in a large cohort (n=294, 18–92 years). Participants freely viewed 160 3°–15.5° sized pictures without being notified about the memory aspect of the study. Afterwards they were given a surprise old-new recognition memory test on these and 160 additional new images. Image memory was not affected by age (as evident in accuracy, dprime, decision criterion and in AUC). Furthermore, larger images were better remembered across all age groups, and the extent of the image size-related memory benefit was also not affected by age. We then tested this in-person with 30 younger (19–31 years) and 30 older (60–77 years, all cognitively healthy, MoCA score of 24 or higher) adults that underwent an assessment of basic visual functions (near and far visual acuity (VA) and contrast sensitivity). Despite significant age-related differences in near and far VA and in contrast sensitivity, we replicated the results of the online study and the only age-related effect that was found here was that older adults were significantly slower to respond. Although older adults undergo multiple vision-related sensory changes, our results indicate that the benefit from visual-sensory enhancement supports memory to a similar extent in older adults as in younger adults. These findings may have important implications for understanding memory processes and for developing methods to improve information accessibility and interface design for older adults.

This study was funded by ISF Individual Research Grants 1485/18 and 1462/23 to SGD, by CHE Lev-Zion Scholarship to OK, and by ISF 1726/22 to BBD

36.305 LARGER IMAGES ARE BETTER REMEMBERED EVEN WHEN PERCEPTUAL MEASURES ARE NOT AFFECTED BY IMAGE SIZE

shaimaa masarwa^{1,2} (shaimaa2.mas@gmail.com), Olga Kreichman^{1,2}, Sharon Gilaie-Dotan^{1,2}; ¹The Gonda Multidisciplinary Brain Research Center, Bar Ilan University, Ramat Gan, Israel, ²School of Optometry and Vision Science, Faculty of Life Sciences, Bar Ilan University, Ramat Gan, Israel

We have recently found that during naturalistic encoding larger images (in the range of 3-12 degrees) are better remembered. However, it is unclear whether this reflects more detailed perception afforded by larger images or genuine memory processes that critically rely on bigger image size. To test this here we measured perception and memory for these image sizes hypothesizing that if better memory for larger image sizes reflects perception then we shall be able to predict memory from perceptual measures. 45 adult participants were presented with 160 images each presented at 3, 6, 12, or 15.5 degrees and perceptually encoded them (reporting for each image whether (i) there was a person in the image, (ii) it was taken indoors or outdoors) without knowing a memory test would follow. Afterwards they were given a surprise old-new recognition-memory test for 320 images (half "old"). While perceptual performance was at ceiling and not influenced by image size, image memory was significantly affected by image size such that larger images were better remembered. Our results may suggest that when the visual system is confronted with challenging tasks (as in the small images condition) immediate processes (as perception) may be prioritized and less resources may be available for longer-term processes (as long term memory).

This study was funded by ISF Individual Research Grants 1485/18 and 1462/23 to SGD, and by CHE Lev-Zion Scholarship to OK

36.306 HONEY, I SHRUNK THE OBJECTS (IN MEMORY): BIASES IN REMEMBERING PHYSICAL SIZE AND THEIR MODULATION BY CONCEPTUAL SIZE

Nurit Gronau¹, Rotem Avital-Cohen¹, Izhak Eliach¹, Keren Aloni-Carmeli¹; ¹The Open University of Israel

An object's physical size is represented logarithmically relative to its real-world magnitude, a conceptual representation known as canonical size (Konkle & Oliva, 2011). We asked how this representation affects long-term memory (LTM) for specific episodic instances in which an object is viewed from a particular distance and in particular retinal dimensions. Specifically, does viewing an object in dimensions consistent with its familiar (conceptual) size enhance memory relative to viewing it in an atypical size? Prior work shows that visual LTM favors retinally large over small images (Masarwa et al., 2022). We tested whether retinal and conceptual size interact in LTM, hypothesizing that large presentations of conceptually large objects (e.g., a tree) would enhance memory more than large presentations of conceptually small ones (e.g., an apple), due to congruency between physical and conceptual size representations. Across a series of

preregistered behavioral experiments, we replicated the LTM advantage for physically large items but found no interaction with conceptual size when probing memory for item identity ("Did you see an apple or a pear?"). However, when assessing memory for the actual physical size of stimuli ("Did you see a large or small apple?"), we observed a robust bias: participants tended to misremember objects as smaller than presented, for both targets and lures, and across different size scales. Notably, this bias was reduced for conceptually small objects shown in large dimensions, likely due to their increased salience and oddity. Thus, a conceptual-by-retinal size interaction emerged, but in an unexpected direction. These findings suggest distinct mechanisms for memorizing object identity versus physical size: identity memory is largely size-invariant, whereas physical-size memory is systematically reconstructed toward smaller (more distant) representations. We discuss how these results relate to object and scene memory reconstruction, offering new insights into the canonical size framework and the Boundary Extension phenomenon.

36.307 ILLUSORY FACES, REAL BIAS: PEOPLE TEND TO FALSELY REPORT REMEMBERING ILLUSORY FACES MORE OFTEN THAN THEY DO FOR OTHER VISUAL CATEGORIES

Olga Kreichman^{1,2} (oplaksin88@gmail.com), Limor Brook^{1,2}, Susan G. Wardle³, Sharon Gilaie-Dotan^{1,2}; ¹School of Optometry and Vision Science, Bar-Ilan University, ²Gonda Multidisciplinary Brain Research Center, Bar-Ilan University, ³National Institute of Mental Health

Perception of illusory faces (aka "face pareidolia" (FP)) happens when the perception of veridical visual information (e.g. objects) is accompanied by an illusory face-like perceptual experience. We assumed that FP images would be remembered better than face or control images (objects images that do not elicit illusory face percept) since they (i) activate both face and non-face related neural mechanisms, (ii) capture both face-related and object-related attention, and (iii) are very atypical in the visual environment. Here we tested this in 4 experiments where participants (N = 231 young adults) were only instructed to freely view the 160 images (of FP, faces, control objects, and outdoor scenes) and were later given a surprise old-new recognition memory test on 320 images (160 old, new images matched for categories). In contrast to our hypothesis, memory sensitivity for FP images was not consistently higher than for faces, for control objects or for outdoor scenes. However, surprisingly, FP images elicited a reliably more liberal response criterion relative to that of the other visual categories such that participants were more likely to report that FP images had been previously seen even when they were not. Moreover, higher faceness FP images led to more liberal response criterion than lower faceness FP images. This pattern may reflect a general decision bias or may indicate that FP images specifically induce false memories.

This study was funded by ISF Individual Research Grants 1485/18 and 1462/23 to SGD, and by CHE Lev-Zion Scholarship to OK

36.308 REPETITION LEARNING VS. WIDE SEMANTIC LEARNING: EFFECTS ON SEMANTIC AND VISUAL MEMORY

Yael Schems Maimon¹ (yael.schems@gmail.com), Roy Luria¹; ¹Tel Aviv University

It is well established that semantic information in long-term memory (LTM) is organized as networks of associations between concepts. Yet many memory studies rely on repetition-based paradigms to investigate learning in LTM. We previously showed that Wide Learning, a rich and varied exposure to texts, images, and videos of objects from an unfamiliar category, creates a semantic network in LTM through three processes: binding category members to one another, linking the new network to relevant pre-existing knowledge, and separating it from irrelevant concepts. The present study directly contrasted Wide Learning with repetition-based learning. Whereas repetition learners repeatedly studied the same facts and the same visual exemplars of the objects, Wide Learning participants were exposed to numerous details, diverse exemplars, and rich contextual information. The structure of the semantic network and memory performance were measured three times in both groups: before learning, about two days post-learning, and about six weeks post-learning. In both groups, learning produced a semantic network with the same three processes of within-category binding, linking to prior knowledge, and separation from unrelated concepts. Critically, despite these similarities in network structure, and despite the advantage repetition learners had in repeatedly studying the exact content and exact visual exemplars that appeared in the test, Wide Learning participants consistently outperformed repetition learners across all tasks. They showed better object naming, higher knowledge scores, and improved visual working-memory performance, both in the immediate test and after six weeks. Furthermore, in two tasks designed to assess the use of semantic networks in natural behavior, a natural language production task and a visuospatial object-organization task, Wide Learning participants demonstrated better category separation at both post-learning time points. Overall, our findings indicate that repetition learning, widely used in exam preparation and in many experimental paradigms, produced weaker memory performance than context-rich, association-based learning.

36.309 RESTRICTED VISUAL EXPLORATION MEDIATES THE IMPACT OF DIVIDED ATTENTION ON MEMORY FORMATION

Chloe Kindell¹, Meghan Garcelon¹, Heather Lucas¹; ¹Louisiana State University

Visual exploration is essential for memory formation, but effective exploration may depend on attentional resources. During free-viewing tasks, divided attention has been shown to restrict visual exploration, producing fewer fixations and longer fixation durations. We examined whether taxing attention during intentional encoding induces similar restricted exploration and whether these viewing constraints are linked

to worse subsequent memory. We also asked whether the effects of attentional load differ when load remains constant versus when it increases or decreases mid-task. Eye movements were recorded while participants viewed 80 displays, each containing six abstract line drawings and followed by a spatial reconstruction test. Attentional load during encoding was manipulated during study via a concurrent auditory task. Participants either completed all trials with an easy concurrent task, all trials with a hard concurrent task, or switched task difficulty halfway (easy to hard or hard to easy). During the first half of the experiment, higher attentional load led to the same restricted exploration observed in free-viewing tasks (fewer fixations and longer fixation durations), as well as poorer subsequent memory. Across groups, memory performance generally improved in the second half, reflecting practice-based improvements. However, viewing behaviors showed little change from the first to the second half, suggesting that practice-related gains arise from mechanisms distinct from those associated with attentional load. A key exception occurred in the easy-to-hard group: the mid-task increase in attentional load appeared to override practice-related improvements and induce a restricted exploration pattern similar to that seen under high load initially. Together, the results point to two interacting influences on visuospatial memory performance. Limited attentional resources restrict exploration and impair memory, whereas practice improves performance without altering viewing behavior.

36.310 TRACKING THE MIND'S EYE: A NOVEL FRAMEWORK FOR STUDYING ATTENTIONAL ORIENTING IN LONG-TERM MEMORY

Melinda Sabo¹, William Narhi-Martinez, Anna C. Nobre; ¹Yale University

A large body of research has established that attentional selection shapes perception and working memory. Yet its consequences for long-term memory remain less clearly defined. Understanding whether and how internal attention impacts long-term memory retrieval requires a framework for manipulating the orienting of attentional selection between competing contents in long-term memory. The present set of three studies addresses this gap by introducing a new experimental framework. The proposed design includes two major phases: a Learning task and a Selective Retrieval task. During Learning, participants study associations between scenes and two embedded objects. After learning, participants complete a five-minute Tetris game. In the Selective Retrieval task, the same scenes reappear. In Experiments 1a and 1b, a cue appears after scene onset (retrocue). The cue either indicates the object to-be-tested (75% validity) or offers no predictive information. In 25% of the trials, informative cues are invalid: they point to the nontarget item. While valid cues enable the study of attentional benefits, invalid cues allow for the examination of attentional costs. Experiment 2 uses 100% informative and neutral cues presented before scene onset (precues). At the end of the Selective Retrieval trials in all experiments, participants are shown a probe array containing two objects and must

indicate which one was associated with the scene. Across all experiments, informative valid cues led to response times benefits during long-term memory retrieval. Additionally, invalid cues were associated with slower response times, suggesting attentional costs. These consistent effects demonstrate that orienting internal spatial attention using retrocues or precues influenced memory performance. Additionally, in Experiment 2, the fast timings of the cues and probe arrays suggested that internal attention may operate directly upon long-term memory representations, unmediated by working memory. Our new design and findings introduce a rigorous approach for studying attentional selection in long-term memory.

This research was supported by a Walter Benjamin Fellowship offered by the German Research Foundation (Deutsche Forschungsgemeinschaft; SA 5174/1-1)

36.311 FALSERESMEM: A NEURAL NETWORK TO PREDICT FALSE ALARMS IN IMAGE MEMORY

Anastasiia Mikhailova¹, Wilma A. Bainbridge¹; ¹Department of Psychology, University of Chicago

Predicting false alarms, i.e., incorrectly identifying novel images as previously seen, is critical for understanding how visual memory fails, quantifying memory bias and the reliability of memory decisions in basic research and applied settings. Prior work has shown that people are similar in their propensity to false alarm for a given image, suggesting some images are consistently more ‘false alarmable’ than others. This finding has received little attention but allows us to develop models that explicitly capture these error patterns, thereby providing a more complete and ecologically valid account of visual memory. It is particularly important because hit rate and false alarm rate capture distinct components of recognition performance and can vary independently. In this work, we introduce a neural network, titled FalseResMem, that predicts image-level false alarm rates. The model uses features extracted from an ImageNet-pretrained ResNet50 backbone, which are concatenated with additional retrained AlexNet-like convolutional layers and subsequently retrained fully connected layers to learn image properties that drive false recognitions. After training with MemCat, a large-scale image memorability dataset of objects, the model demonstrates consistent performance (test Spearman’s rank correlation on 10-fold cross-validation: 0.48 ± 0.13 ; max. 0.58) in anticipating which images are more likely to elicit false recognitions, complementing existing hit-rate-based memorability predictions (ResMem). Furthermore, the present model successfully makes out-of-sample predictions for other types of images, such as scenes, art, symbols, and even those images that elicit the visual Mandela Effect, confirming robust results on a diverse set of images. This contribution highlights the potential of integrating error-based metrics into visual memory modeling and opens new pathways for studying bias, confusion, and reliability in computational measures of image memory. To further facilitate broader adoption and experimental use, we provide an easy-to-use guide that supports researchers in using FalseResMem to test images for predicted false alarm rates.

The present work was supported by the National Science Foundation under Grant No. CAREER 2441710. The authors declare no competing interests.

36.312 SYSTEMATIC BIAS TOWARD AUTHENTIC PAINTING STYLES REVEALED THROUGH STYLE-TRANSFER TECHNIQUES

Xiyan Ji¹ (jixiyan@cv.jinkan.kyoto-u.ac.jp), Hiroyuki Tsuda², Jun Saiki¹; ¹Kyoto University, ²Shizuoka Institute of Science and Technology

Human memory does not simply reproduce the visual world; it shows systematic biases. The present study investigates whether long-term memory for painting styles shows categorical biases toward authentic artistic styles, and whether styles that elicit stronger biases related to the images’ instinct properties—memorability. To address these questions, we conducted Experiment 1 and a replication, Experiment 2. In Experiment 1, 81 style-transferred images were created by combining 75 landscape contents with four authentic styles—Braque, Kandinsky, Monet, and Rembrandt. Participants remembered 25 study styles and reported it in a continuous-report task. Memory responses did not distribute uniformly: they concentrated on the authentic styles and on mixed styles near the center of style space. According to the analysis of standard memory error, the memory precision for authentic styles was better than mixed styles around the center, showing the memory bias for authentic styles and mixed styles had different mechanisms. Moreover, among the four authentic styles, the strongest bias appeared for Kandinsky’s style, which had the highest memorability score. Experiment 2 repeated this paradigm with four different authentic styles—Gogh, Marc, Hokusai, and Rubens—to test whether the pattern from Experiment 1 reflects a general property of style memory rather than an effect of specific styles. The results also showed higher response frequencies for both the authentic styles and the mixed styles near the center of style space. Moreover, Marc’s authentic style produced the strongest bias, and Marc also had the highest memorability score. Our results reveal two biases in memory for painting styles: attraction toward authentic styles and attraction toward the mean style of the stimulus set. These biases may reflect partly distinct mechanisms. In particular, authentic styles may act as reference points, with more memorable authentic styles producing stronger attraction.

36.313 A GENERATIVE MODEL TO MANIPULATE THE MEMORABILITY OF FACE IMAGES

Nakwon Rim¹, Marc G. Berman¹, Stefan Uddenberg², Shunichi Kasahara³, Wilma A. Bainbridge¹; ¹University of Chicago, ²University of Illinois Urbana-Champaign, ³Sony Computer Science Laboratories

Recent advances in generative artificial intelligence now enable the generation of synthetic human faces that observers cannot reliably distinguish from real ones. These models offer a powerful methodological opportunity for research on face perception and cognition: they are able to produce an arbitrary number of

photorealistic faces with rich latent representations that can support systematic sampling and targeted transformation. Here, we evaluated whether such generative models can serve as an effective tool for understanding and augmenting face memorability. We first conducted two analyses on memorability scores for 1,004 StyleGAN2-generated faces to test whether their memorability aligns with that of real faces. In the first analysis, a neural network trained exclusively on the memorability data of real faces successfully predicted memorability in our synthetic set. In addition, we found that facial judgment features previously shown to relate to the memorability of real faces also showed the same relationships with the memorability of synthetic faces. For example, faces judged as more “typical” were less memorable, paralleling previous findings for real faces. Finally, having established this correspondence, we leveraged StyleGAN2’s latent space to test whether memorability could be systematically manipulated. Across two experiments employing distinct manipulation strategies, we were able to consistently increase the memorability of synthetic faces but failed to decrease it reliably. This asymmetry suggests that memorable faces may share a relatively homogeneous set of visual attributes that are readily captured in generative latent spaces, whereas forgettable faces may arise from more heterogeneous, idiosyncratic configurations that are harder to capture. Together, these findings demonstrate that generative models can serve as a valuable tool for studying and influencing face memorability. More broadly, the ability to algorithmically enhance face memorability opens the door to applications such as real-time enhancement of facial recall in communication or educational settings.

The research was supported by the National Science Foundation under CAREER Grant No. 2441710.

Visual Memory: Encoding and retrieval, capacity

SUNDAY, MAY 17, 2:45 – 6:45 PM, BANYAN BREEZEWAY

36.314 MULTI-ITEM ARRAYS ARE RETRIEVED FROM LONG-TERM MEMORY INTO WORKING MEMORY AS UNITIZED CHUNKS

Woohyeuk (Leo) Chang¹, Ed Awh¹; ¹University of Chicago

Classic models propose that information retrieved from long-term memory (LTM) is reinstated into working memory (WM), engaging the same capacity-limited operations that support online storage. For example, alpha oscillations decline as a function of the number of items stored in WM, and a similar pattern is observed when observers retrieve multi-item arrays from LTM. Although this suggests that LTM retrieval mirrors WM encoding, other work indicates that alpha activity may instead reflect the spatial extent of covert attention rather than the number of WM representations per se. Thus, there is strong motivation

to study reinstatement using neural signals more directly linked to WM storage and designs that de-confound spatial attention from the number of stored items. We used a sequential design in which items were separated in time rather than space, holding spatial attention constant while manipulating the number of individuated items. A multivariate decoding approach known to track WM load independently of spatial attention was used. Participants memorized LTM shape arrays containing 1 or 3 targets and then completed a sequential WM/LTM task during EEG recording. Voltage-based signatures of WM load generalized across conditions requiring the maintenance of temporal order and item identity. The key question was whether retrieving multi-item arrays from LTM would elicit a multi-item WM load signature or whether prior learning would result in unitization (i.e., chunking). Consistent with the latter, LTM retrieval elicited the same neural pattern as a WM load-1 trial, regardless of the array’s size. Alpha power was likewise insensitive to the number of retrieved items, likely because each sequence appeared at a single location. Thus, retrieval of multi-item arrays from LTM reveals a unitized representation that is insensitive to set size, providing insight into how associative learning shapes the reinstatement of multi-item LTM structures in WM.

36.315 DOES ADDED SEMANTIC INFORMATION ALTER HOW YOU DRAW AMBIGUOUS ‘MOONEY’ IMAGES FROM MEMORY?

Emma Megla¹ (emegla@uchicago.edu), Ophra A. Atar¹, Wilma A. Bainbridge¹; ¹University of Chicago

Visual experiences are not just composed of unique visual details, but also semantic information. For example, your friend may remark that your potato chip “looks just like Elvis Presley,” providing information beyond what you are just receiving visually. Does this added semantic information change your visual memory of the potato chip? Whereas sensory reinstatement accounts would predict that visual memory is a recapitulation of perceived visual details, semantic transformation accounts would predict that semantics have the power to alter memory. Here, we presented Prolific participants (N=118) with ambiguous, two-toned ‘Mooney’ images of objects from the THINGS database that are difficult to identify without the object label (Linde-Domingo et al., 2025). To capture visual memory when it is just composed of visual details, participants drew these images from memory without the object label (visual drawings) and then—to capture visual memory after gaining semantics—drew these images again while presented with the object label (semantic drawings). By analyzing over a thousand of these drawings, we found that semantics made the objects in the drawings significantly more identifiable to the neural network CLIP, but at the cost of removing visual detail from memory, with semantic drawings containing significantly fewer edge pixels. However, could it be that semantics changed how participants encoded the image itself, rather than how they retrieved it? In a second experiment, Prolific participants (N=120) drew Mooney images while viewing them (i.e., during encoding) both with and without their object

labels. Semantics still pruned visual detail even during encoding, with semantic drawings containing significantly fewer edge pixels, but the objects in these semantic drawings were not more identifiable than in the visual drawings. In other words, we found evidence that semantic information does change visual memory, leading to encoding fewer visual details and retrieving an image that better resembles the object.

The present work was supported by the National Eye Institute (R01-EY034432) to W.A.B.

36.316 RECOGNITION MEMORY IS BIASED TOWARD PERCEIVED NEAR SPACE

Tasfia Ahsan¹ (ahsant@my.yorku.ca), Laurie M. Wilcox¹, Erez Freud¹; ¹Department of Psychology, York University

Depth cues influence how we perceive and act upon objects, with near-space stimuli receiving prioritized processing. This prioritization of near space is reflected in the “close advantage effect” in perceptual and attentional processing, a phenomenon that occurs even when depth is defined only by pictorial cues. Some studies show that spatial memory can differ between near and far space, but it remains unknown whether perceived depth influences memory for visual objects. To assess this, we used the Mnemonic Similarity Task (MST) to examine whether perceived depth modulates memory fidelity. Participants viewed a series of objects, presented either on the near or far surface of the Ponzo illusion. They were not told they would be asked to do a recall task. In our first experiment, we kept object size constant, irrespective of position (near or far). Recognition accuracy, which serves as an indicator of familiarity-based memory, was significantly higher for items encoded on the close surface. However, lure discrimination (LDI), which reflects the ability of the hippocampus to separate similar items into distinct representations, did not vary with perceived depth. Thus, perceived proximity strengthens recognition without altering memory precision. Our second study evaluated whether this version of the close advantage reflected size-distance scaling, where an object perceived as nearer is interpreted as smaller than an identical object placed farther away. A size-matching pretest was performed to equate the perceived size of the stimuli presented on the close and far surfaces; then we repeated Experiment 1 using these equivalent-size stimuli. We found that the close-surface recognition advantage persisted, and again there was no change in LDI. Taken together, these findings show that perceived proximity enhances recognition memory even when size cues are controlled, consistent with privileged encoding of near-space stimuli. Follow-up studies probing the limits of this effect will also be discussed.

Natural Sciences Research Council of Canada (NSERC); CF-REF program Vision Sciences to Applications (VISTA)

36.317 PICTURE SUPERIORITY NOT STRONGLY SENSITIVE TO IMAGE TYPE

Anne T Gilman^{1,2}, Abigail Noyce³; ¹SUNY Albany, ²Bennington College, ³Carnegie Mellon University

The picture superiority effect, a long-documented memory advantage for images over words, yields insights into the impact of aging and degenerative disease on memory, largely due to greater activation from conceptual processing of images (Ally 2012). Do perceptual characteristics play any role in the strength of the picture superiority effect? While this memory advantage has been found for photos (Shepard 1967) and line drawings (Curran & Doyle 2011), those precedents only contrasted image and word stimuli. Rossion & Pourtois (2004) observed faster, more-accurate image naming for color drawings versus monochrome outlines. Heuer (2016) found greater attention paid to photographs than drawings in another image-naming protocol, and a study of verbal associations with images also found an advantage for photographs over color and outline drawings (Gilman 2016). Attempts to clarify whether stimulus characteristics alter picture superiority have shown a consistent trend towards an advantage for drawings over photos (Bran 2025), so long as the photos are nameable. The present study used more trials per block to increase memory load in an attempt to clarify that trend. Since older adults show reduced depth of processing benefits (Craik 2002, 2012), to support future aging memory studies, the present study also omitted liking ratings from study trials. Fifty-five undergraduates completed the study and identified their gender (27 women, 23 men, 1 non-binary). Participants gave consent and completed a tutorial before completing two rounds of 75-item study and 150-item test trials. Study blocks included 25 photos, drawings, and words each, with items counterbalanced across participants; all test items were words, following Curran & Doyle (2011). Study format significantly predicted recognition accuracy, $F(2,159)=25.7$, $p<.001$. Drawings (63%) and photos (62%) both showed a TukeyHSD-significant difference from words (51%) but not each other ($p>.5$), suggesting that image type is not a primary driver of the picture superiority effect.

36.318 TO WHAT EXTENT DOES EXISTING KNOWLEDGE INFLUENCE STIMULUS MEMORABILITY?

Jie Zheng¹, Sami R. Yousif¹, Brynn E. Sherman¹; ¹The Ohio State University

The information we encounter in our environments is seldom novel: Virtually everything that we see is a variation of something that we've seen before. That existing knowledge can powerfully shape the way we process and encode new experiences into memory. For instance, a familiar chess position might be easier to remember than an unfamiliar one. Yet our memories can also be influenced by intrinsic properties of the stimuli themselves. That is, some stimuli are more memorable than others; some images are consistently remembered or forgotten across people. But how does intrinsic memorability interact with an individual's existing, idiosyncratic knowledge of the world? Here, we explored this question using a stimulus set for which people have a wide range of existing knowledge: Pokémon. Specifically, we had two groups of participants perform a simple encoding task with Pokémon images, followed by a recognition memory task. One group of participants had existing knowledge of Pokémon (Familiar group),

and one group had no such knowledge (Unfamiliar group). Across two experiments using different image sets, we found that memorability was shaped by existing knowledge. Despite having worse memory overall, participants in the Unfamiliar group were more reliable in their memories, in the sense that there was greater consistency in which images were memorable (hit rate) and which were familiar (false alarm rate). The relatively lower consistency in the Familiar group may suggest that existing knowledge causes memories to become more idiosyncratic, leading to greater variability within the population. In other words, while increased knowledge may improve memory overall, it may also make some memories more distinctive to individuals, partially counteracting the impact of intrinsic memorability in driving consistent memory across people.

36.319 THE ILLUSION OF KNOWING: WHEN METACOGNITION DIVERGES FROM MEMORABILITY

Dyllan Simpson¹ (dysimpson@ucsd.edu), Joseph Saito¹, Timothy Brady¹; ¹University of California San Diego

In a given context, some visual stimuli are consistently remembered while others are consistently forgotten resulting in the phenomenon referred to as stimulus memorability. However, observers' metacognitive access to stimulus memorability is unclear, with studies providing mixed evidence for the ability to predict which images will be consistently remembered versus forgotten. Here, we hypothesized that inconsistent access to memorability may reflect reliance on a unique distribution of features that does not universally overlap with those that are most distinct. In Experiment 1, we trained separate ridge regression models to predict memorability and judgments of learning (JOLs) within the THINGS image database. Consistent with our prediction, we found that the visual features that predicted memorability and JOLs overlapped, but not perfectly. For example, bathroom-related features were positively correlated with memorability in the database but were negatively correlated with JOLs. In Experiment 2, we manipulated the prevalence of the unique and overlapping features identified by each model within a new set of images to see how this would influence the correspondence between JOLs and memorability. In doing so, we found that selectively overrepresenting the features that originally predicted high memorability eliminated their distinctiveness and inverted rank-order memorability, but did not invert rank-order JOLs. However, when we overrepresented the features that originally predicted both high memorability and JOLs, both rank orders inverted. Two additional experiments with scenes and sentences confirmed JOLs rapidly calibrated to local statistics producing correspondence. The dissociability suggests that metacognitive access to stimulus memorability arises incidentally when the features used to form JOLs happen to align with the features that drive memorability in the database. We speculate that these unique feature distributions reflect visual statistics acquired across different timescales, with memorability primarily reflecting the flexible statistics of the database and JOLs primarily reflecting the stable statistics of the natural environment.

36.320 NEURAL SIGNATURES OF VISUAL MEMORABILITY REVEALED BY EEG DECODING

Igor Utochkin¹, Woohyeuk Chang¹, Edward Vogel¹, Edward Awh¹; ¹University of Chicago

Stimulus memorability is a systematic tendency for some visual items to be remembered better than others despite variations in learning and test context. Prior studies show that memorability can be decoded from neural activity. However, the extant studies have treated memorability as a single dimension. We highlight that memorability is a compound phenomenon based on two parameters: "Hittability" (H-ability), the likelihood that a studied item will be correctly recognized, and "false-alarmability" (FA-ability), the likelihood that the same item will be erroneously recognized when not studied. Strikingly, these parameters are uncorrelated, and whether their neural bases are distinct or shared remains unclear. Using multivariate decoding of EEG, we show that both H-ability and FA-ability can be robustly decoded from neural activity. Importantly, these signatures are evident not only during the test phase, where retrieval success might drive decoding, but also during study. Neural patterns associated with each parameter generalize across study and test, indicating that their representational bases are not contingent on explicit memory judgments. We next asked whether decoding of H-ability and FA-ability reflects movement along a shared representational dimension. Despite their behavioral independence, cross-decoding revealed strong generalization: classifiers trained to decode H-ability reliably decoded FA-ability, and vice versa. Thus, these parameters share a common dimension in neural state space. A natural hypothesis is that this dimension reflects item familiarity. Contrary to this prediction, neural patterns associated with high H-ability aligned with those associated with low FA-ability. This inversion rules out a simple familiarity account and instead suggests that neural signatures of memorability track systematic biases in decision efficacy based on relative familiarity. Presumably, stimuli with high H-ability and low FA-ability are easier to endorse or reject based on mnemonic status, whereas those with low H-ability and high FA-ability make these decisions more difficult.

ONR N00014-22-1-2123

36.321 TRACING THE EMERGENCE OF PREDICTION-BASED FALSE MEMORY

Olya Bulatova^{1,2} (olya.bulatova@mail.utoronto.ca), Mahveen Salman Mubarak², Keisuke Fukuda^{1,2}; ¹University of Toronto, ²University of Toronto Mississauga

Explicit prediction alone, even without statistical regularity, is sufficient to implant visual memories of predicted yet unexperienced events (Bulatova & Fukuda, 2025). This prediction-based false memory is extremely robust: it cannot be eliminated when a contradicting event immediately rebuts the prediction, and when predictions were highly unreliable. However, it remains unclear whether the prediction-based false memory is encoded at the time of prediction or inferred at the time of retrieval. To test this, we recorded EEG as participants 1) made

predictions of upcoming visual events that were either confirmed, violated or not verified, and 2) reported their memories of the actual events. Specifically, during encoding, participants saw a colourless silhouette of an everyday object and they predicted what colour the silhouette would be. A thousand ms afterwards, the silhouette turned into either the predicted colour, the opposite colour, or remained colourless with equal likelihood. Importantly, participants were instructed to remember the actual, rather than predicted, colour of the silhouette. After encoding 360 silhouettes, participants' memory for the actual colour of the silhouettes was tested. First, participants saw a colourless silhouette and indicated whether they remember seeing it 1) definitely in colour, 2) maybe in colour, 3) definitely not in colour, or lastly that 4) they did not see the silhouette. Afterwards, participants reported the remembered colour on the colour wheel along with their confidence in their report. Our results so far ($N = 12/30$) reveal that the midline theta (3-7Hz) power during encoding was significantly stronger when prediction turned into false memories than when they did not. Furthermore, the response-locked Late Posterior Positivity was significantly larger when indicating the presence of colour memory for the actual events than for predicted events. Taken together, our results demonstrate that both encoding and retrieval processes contribute to the prediction-based false memory.

36.322 REVISITING EVIDENCE AGAINST THE DISCRETE-CAPACITY ACCOUNT OF STM LIMITS: OBJECTIVE GUESSING OBSCURED BY REPRESENTATIONAL MODELING

Tianye Ma¹ (tma039@ucr.edu), Inik Kim¹, Daniel Punzalan¹, Weiwei Zhang¹; ¹University of California, Riverside

The discrete-capacity account of short-term memory (STM) posits a fundamental distinction between stored items and "objective guessing" (zero information). In contrast, continuous-resource models, particularly the Target Confusability Competition (TCC) model, reject discrete limits, arguing that "guessing" is merely an artifact of low signal-to-noise ratios scaled by psychophysical similarity. TCC's superior fit in statistical model comparisons has been widely cited as evidence against discrete capacity. Here, we demonstrate that such conclusions are unwarranted. First, for simulations using a generative process model (combining binding pool and population coding mechanisms) with known discrete limits, TCC consistently outperformed Mixture models, failing to recover the ground truth of capacity bottlenecks. Second, in an empirical experiment, we manipulated the probability of target absence (20% vs. 50%) to establish a ground-truth lower bound for guessing. TCC continued to outperform guess-explicit models (Mixture, TCC-g) even when objective guessing was forced. Critically, TCC-g failed to accurately recover the known guess rates, frequently estimating zero guessing despite 20%/50% target absence. Second, TCC's strength parameter (d') conflates representational strength with noise. In line with this parameter conflation, an individual-difference analyses reveal that the variance explained by TCC d' can be partitioned into variance

stemming from the capacity limit and the precision limit of STM across individuals. Thus, TCC d' acts as a composite measure of both quantity (capacity) and quality (precision), rather than a pure index of memory strength. Together these findings indicate that standard statistical model comparisons are insensitive to the structural differences between these theories, rendering the rejection of objective guessing premature.

36.323 DURING LESS ENGAGING ENCODING CONDITIONS VISUAL LONG-TERM MEMORY IS MORE LIMITED THAN PREVIOUSLY ASSUMED

Reout dayan¹, (reoutdayan@hotmail.com), Olga kreichman², Sharon Gilaie-Dotan³; ¹School of Optometry and Vision Science, Faculty of Life Science, Bar Ilan University, Ramat Gan, Israel, ²The Gonda Multidisciplinary Brain Research Center, Bar Ilan University, Ramat Gan, Israel

Studies have shown that people can remember thousands of images with remarkable detail when they are deeply engaged in encoding these images (e.g., Brady et al. 2008). However, it remains unclear whether this reflects everyday memory when people are typically not highly engaged in encoding the images they come across. Across a series of comprehensive experiments ($n=339$), we recently found that under much less engaging conditions, memory for object images (taken from Brady et al. 2008) is affected by the number of encoded images and by the number of interfering images at test. Since when people have rich representation of objects they are able to successfully distinguish between them and other similar items (e.g., finding their luggage on the conveyor belt at the airport), here we further tested memory strength by examining whether it would be affected by the similarity of the distractors at test. Here too, memory ($n=62$) was significantly affected by the number of encoded images and by the number of distractors at test. Importantly, memory performance was affected by the similarity of the distractors (e.g., for 60 encoded images: 2AFC accuracy of 86% (similar distractors) vs 91% (non-similar distractors), for 180 encoded images: 6AFC accuracy of 41% (similar distractors) vs 55% (non-similar distractors)). While image memory capacity can reach its extensive potential during demanding encoding conditions, our results suggest that during less engaging conditions, image memory is not as strong and extensive as previously assumed.

This study was funded by ISF Individual Research Grants 1485/18 and 1462/23 to SGD, and by CHE Lev-Zion Schola hofital@gmail.com rship to OK

36.324 IS REPULSIVE SERIAL BIAS A TOP-DOWN DRIVEN PHENOMENON?

Scott Janetsky¹, Jordan Dolinar¹, Kuo-Wei Chen¹, Gi-Yeul Bae¹; ¹Arizona State University

Visually guided behaviors are systematically influenced by recent perceptual history. In location estimation tasks, for example, reports

are often repelled away from the previous-trial target, even when the previous target is task-irrelevant. Prevailing models attribute this repulsive serial bias to sensory adaptation: the previous target alters the encoding of the current target via bottom-up mechanisms. However, the potential contribution of top-down mechanisms to the bias effect has been under explored. To investigate whether repulsive serial bias could arise when bottom-up sources of bias are minimized, if not completely prevented, we had participants perform a location estimation task with implicit locations: they remembered the center of two reference objects and reported the center location by adjusting the center of two new reference objects. Because there was no explicit stimulus at the center of the two reference objects, participants had to imagine the center location during both encoding and reporting. To avoid the potential adaptation effect from the reference objects, the reference objects for the encoding and reporting were physically deviated from each other across trials. We predicted that, if repulsive serial bias is solely driven by bottom-up sensory adaptation, then the repulsive serial bias in this task would be absent or substantially smaller compared to the repulsive bias for the explicit locations. However, we found that repulsive serial bias for the implicit locations was comparable to the repulsive bias for the explicit locations. This was replicated in follow-up experiments where participants performed a location estimation task either with an implicit target but with an explicit response object or with an explicit target but with an implicit response object. Together, these results demonstrate that repulsive serial bias can occur for higher-level abstract representations and further suggest that bottom-up sensory adaptation is not necessary to explain repulsive serial bias effects.

36.325 PHRASAL MOMENTUM

Chaz Firestone¹, Tal Boger¹; ¹Johns Hopkins University

The world looks different from one moment to the next, and various mental processes anticipate such changes. A foundational example of this anticipation is representational momentum (RM), wherein the mind ‘plays forward’ visual events. For example, falling objects are recalled as closer to the ground, and rotating rectangles are recalled as more rotated. Though early studies hypothesized that RM primarily concerns the anticipation of motion, recent work demonstrates that it arises for other continuous properties, such as the brightness of a stimulus or even how melted an ice cube appears. Here, we extend representational momentum even further than these sophisticated phenomena, to discrete (rather than continuous) properties, as well as to more conceptual (rather than purely visual) properties of the world. Participants watched animations of common phrases being typed; the animations were interrupted partway through (e.g., “The quick brown fox jum”), and participants used a slider to recall the phrase’s state at the moment it disappeared. 5 experiments revealed ‘phrasal momentum’: The mind plays forward the phrases we see, in ways that distort visual memory. This distortion was greater for real phrases than for (a) strings of “X”s (Experiment 1), (b) length-equated strings of random letters (Experiment 2), and (c) scrambled versions of the words composing the phrases (e.g., “heT Kicqu rbnow oxf”; Experiment

3). We also observed a stronger distortion for normal, in-order phrases than for word-shuffled versions of those same phrases (e.g., “The quick brown fox...” vs. “Fox quick the brown...”; Experiment 4). Finally, Experiment 5 found that phrasal momentum follows a conceptual ‘gradient’; in a single experiment combining all the aforementioned conditions, normal phrases produced the strongest RM effects, followed by shuffled phrases, scrambled letters, random letters, and Xs. Thus, representational momentum arises even for discrete and non-visual stimuli.

36.326 DIFFERENTIAL IMPACT OF INCIDENTAL AND INTENTIONAL ENCODING ON PUPILLARY RESPONSE DURING RECOGNITION MEMORY

Wen Jian¹ (wen_jian@brown.edu), Elena K. Festa¹, William C. Heindel¹; ¹Brown University

Our pupils change size not only in response to light, but also to internal cognitive processes. Pupil dilation has been linked to activity within the locus coeruleus noradrenergic arousal system. During tests of recognition memory, greater pupil dilation is elicited by studied than new items, although the mechanism underlying this pupillary old/new effect remains elusive. Recent findings suggest that it may reflect the interaction between the retrieval cue and the stored memory trace, as pupils dilate more in response to greater perceptual cue-trace match. We aimed to investigate how this pupillary match effect is influenced by encoding intentionality and stimulus repetition. Two groups of healthy young adults completed either incidental or intentional encoding of visual stimuli presented under two repetition conditions: Distributed learning (which enhances contextual recollection) and massed learning (which minimizes contextual recollection but supports familiarity). A subsequent test phase consisted of previously studied items under varying cue-trace match plus novel items. As expected, recognition accuracy was greater following intentional than incidental encoding as well as distributed than massed learning. Following distributed learning, incidental and intentional encoding produced comparable pupillary old/new and cue-trace match effects. Following massed learning, incidental encoding produced an early reversed pupillary match effect along with a reduction in the normal match effect, whereas intentional encoding produced a reduction in the early reversed match effect along with a reemergence of the match effect observed under distributed learning. These results suggest that the magnitude of pupil dilation reflects the nature of encoded traces and its interaction with retrieval cues during recognition memory. The pupillary match effect following distributed learning may reflect enhanced contextual recollection, while the early reversed match effect following incidental, massed learning may reflect reduced retrieval effort associated with increased processing fluency, and this implicit processing effect may be reduced following intentional massed learning.

36.327 NEURAL DIFFERENTIATION UNDERLYING PERCEPTUAL GROUPING BENEFITS IN VISUAL WORKING MEMORY

Qianru Xu^{1,2} (psy.qianru.xu@outlook.com), Chaoxiong Ye^{1,2}, Weizhen Xie³; ¹School of Education, Anyang Normal University, China, ²Department of Psychology, University of Jyväskylä, Finland, ³Department of Psychology, University of Maryland, USA

Visual working memory (VWM) is essential for temporarily storing and manipulating visual information, yet its limited capacity raises the question of how it can be used more efficiently. One potential mechanism is perceptual organization, such as grouping identical objects to help encode and retain more items. In this study, we used multivariate pattern analysis (MVPA) to see when and how perceptual redundancy affects the neural patterns underlying VWM. Participants performed a lateralized change-detection task with memory arrays containing all-same, partial-different, or all-different orientations. Time-resolved decoding revealed early neural differentiation (around 200 ms post-stimulus) between the all-same and other conditions, suggesting that grouping benefits arise already during the perceptual encoding phase, an effect that was not observed in our event-related potential (ERP) analysis. However, this effect was specific to the all-same condition, as partial redundancies yielded no intermediate neural benefit. Cross-temporal decoding further showed that although the all-same condition produced a stable and distinguishable neural pattern across the memory interval, the early neural differentiation did not generalize to later VWM stages, indicating a shift in the underlying representational code over time. Cross-condition decoding also revealed that the partial-different condition was represented more similarly to the all-different arrays than to the all-same ones. Together, these findings indicate that perceptual grouping in VWM operates through a separation-based mechanism of neural differentiation: full redundancy enables early compression into compact neural codes, whereas partial redundancy preserves separable representations to reduce interference. This separation-based compression account refines current theories of perceptual organization and highlights the value of MVPA for revealing representational dynamics that traditional ERP measures may miss. Future work should examine whether similar redundancy-driven mechanisms generalize to other visual features, such as color.

This work was supported by the Research Council of Finland (former Academy of Finland) Academy Research Fellow project (#355369 to Chaoxiong Ye) and Finnish Cultural Foundation (#231373 to Chaoxiong Ye, #240135 to Qianru Xu).

Spatial Vision: Natural images, texture

SUNDAY, MAY 17, 2:45 – 6:45 PM, BANYAN BREEZEWAY

36.328 HIGH-DIMENSIONAL STRUCTURE OF A PERCEPTUAL SPACE OF VISUAL TEXTURES

Jonathan D. Victor¹, Mary M. Conte¹; ¹Weill Cornell Medical College

Processing of visual texture plays an important role in separating figure from ground, and recognizing objects and materials. Visual texture is also a convenient domain for probing the structure of complex perceptual spaces, as texture appearance is the result of different types of low-level features, including contrast, orientation, edges, and corners – and textures can be synthesized in which these features can be independently controlled. Previously we used these texture properties to characterize the nonlinear transformation between the perceptual space supporting texture discrimination thresholds and the space used for suprathreshold similarity judgments. The former is well-modeled by a Euclidean space whose coordinate axes correspond to low-level features and the nonlinear transformation warps those axes and alters the angles between them. Nevertheless, this nonlinear transformation had a dimension-preserving property: for a family of textures with k parameters (typically $k=4$, with parameters including first, second, and higher-order image statistics) suprathreshold perceptual distances could be largely accounted for by Euclidean distances in a k -dimensional space. To test this property stringently, we obtained texture similarity judgments in 6 subjects and 9 sets of texture parameters. For individual subjects and parameter sets, while the similarity judgments were accounted for by a Euclidean distance, the dimension-preserving fit was imperfect. We analyzed deviations from the fit via shuffle tests that assessed whether addition of further dimensions resulted in a statistically significant decrease in the unexplained variance. This analysis revealed that deviations from the dimension-preserving model were highly consistent both across subjects and texture parameter sets, and supported the addition of at least two dimensions to the space. These findings suggest a distinctive combination of characteristics for the representation of suprathreshold texture similarity: parametric families of textures occupy a low-dimensional subset in a higher-dimensional space, and within this space, similarity corresponds to Euclidean distance.

NIH EY07977, Fred Plum Fellowship in Systems Neurology and Neuroscience

36.329 DIAGNOSING BOTTLENECKS IN DATA VISUALIZATION UNDERSTANDING BY VISION-LANGUAGE MODELS

Alexa Tartaglino¹ (alexart@stanford.edu), Daniel Wurgaf¹, Satchel Grant¹, Christopher Potts¹, Judith Fan¹; ¹Stanford University

Understanding any data visualization requires integrating multiple types of visual input (i.e., text, numerals, shapes, curves) to draw appropriate inferences. Current vision-language models (VLMs) are, in principle, promising candidates for modeling the fundamental computations that underlie data visualization understanding. However, these models still struggle on basic data visualization understanding tasks (Verma et al., 2025), and the causes of failure remain unclear. Are VLM failures attributable to limitations in how visual information in the data visualization is encoded, how information is transferred between the vision and language modules, or how information is processed within the language module? We developed FUGU, a suite

of data visualization understanding tasks, to precisely characterize potential sources of difficulty in these models (e.g., extracting the position of data points, distances between them, and other summary statistics). We used FUGU to investigate three widely used VLMs (LLaMA-3.2, LLaVA-OneVision, InternVL3). To diagnose the sources of errors produced by these models, we used activation patching and linear decoding to trace information flow through each model component. We found that some models fail to generate the coordinates of individual data points correctly, and these initial errors often lead to erroneous final responses. When these models are provided with the correct coordinates, performance improves substantially, suggesting that the downstream mathematical reasoning steps performed in the language module are sound. Moreover, even when the model generates an incorrect response, the correct coordinates can be reliably decoded from latent representations in the visual encoder, suggesting that the source of these errors lies not in limitations within the visual encoder, but in the vision-language handoff. Together, these findings point to some of the opportunities and challenges presented by VLMs as a basis for developing input-computable models of data visualization understanding.

This work was supported by NSF CAREER Award #2436199, NSF DRL #2400471, and awards from the Stanford Human-Centered AI Institute (HAI) and Stanford Accelerator for Learning.

36.330 NOT SEEING STRAIGHT: BIASES IN THE PERCEPTION OF PATH LENGTH

Sixuan Chen¹, Heiko Schütt², Weiji Ma³; ¹Brown University, ²University of Luxembourg, ³New York University

Mathematically, the length of a path is simply the sum of the lengths of its parts. Human perception of path length is more interesting. Previous studies have claimed biases in the perception of the length of zigzag paths, but have been contradictory on the direction of the bias: some claim that zigzag paths are perceived as longer than equally long straight ones, others as shorter. To resolve this contradiction, we conducted a psychophysical experiment in which we systematically varied the length, the number of segments, and the sum of angles of a zigzag path. On each trial, a participant compared the lengths of a zigzag path and a straight line. We analyzed the point of subjective equality (PSE) and the just noticeable difference (JND) as a function of the number of segments, the sum of angles, and other summary statistics. Surprisingly, PSE showed a U-shaped relationship with two different measures of zigzagness: participants overestimated length at low zigzagness but underestimated it at high zigzagness. To account for these results, we constructed a model in which the observer estimates path length using three noisy cues: the true length, the Euclidean distance between the path endpoints, and the square root of the convex hull area of the path. This model qualitatively and quantitatively accounts for the trends in PSE and JND. In a second experiment, we replicated a previous study that used only right angles and further validated the model; the stimuli in this experiment had low zigzagness and consequently, data and model both only show

overestimation of length. Together, these findings resolve the contradiction in the literature and suggest that the visual system utilizes global shape statistics for path length estimation.

36.331 DENSITY AFTEREFFECTS DEPEND ON RETINAL (AND NOT REAL WORLD) ADAPTER DENSITY

Frank Durgin¹ (fdurgin1@swarthmore.edu), Alexander Huk²; ¹Swarthmore College, ²University of California Los Angeles

It has been hypothesized that large aftereffects in the perception of visual density (and thus of number) may be primarily due to adaptation of early, density-sensitive channels of visual information. Most studies measuring density aftereffects are conducted on displays at a fixed distance from the observer, leaving it ambiguous as to whether the key representation is real-world (i.e., object-centric) density or retinal density. Here we dissociated these by measuring density aftereffects to adapting textures presented at a different distance than the tests. Two monitors were positioned at 60 cm and 120 cm from the observer, with the far display positioned above the near display. After establishing baseline density matches to 2 dots/deg² on the near display, adapting textures (6 flashes prior to each test) were presented around a fixation mark on the far display, with fixation changing to the near display prior to the presentation of the test patches. The retinal density of the adapter on the far display was 4 times denser than the near display, but due to the distance difference, appeared to be a larger patch of similar world density. In a control condition, two virtual displays at the same distance (top and bottom of the near display) were employed using a retinally-matched adapter (also 4 times the test density, but now appearing much denser) in the top display with test patches in the bottom display, so that vertical refixation was again required between adapter and test. Observers (n=2) showed equally strong downward shifts in perceived density following adaptation to the distant patch (-46%) as to the nearer patch relative to baseline (-43%), with Weber fractions of 5 to 10% in all cases. Thus, an early representation, corresponding to retinal density, is the likely site of adaptation, with downstream effects on perceived density.

Swarthmore College Faculty Research Grant

36.332 EFFICIENT CODING PREDICTS VISUAL COMFORT RATINGS FOR NATURAL IMAGES AND THEIR MANIPULATIONS BUT FAILS TO EXPLAIN TRYPOPHOBIA

Christopher DiMattina¹ (cdimattina@fgcu.edu), R. Nathan Pipitone¹; ¹Florida Gulf Coast University

Images whose statistical properties deviate from those typical of natural images are often uncomfortable to view. One proposed explanation for this discomfort is that unnatural images over-activate the primary visual cortex, a hypothesis known as the Cortical Hyper-metabolism Hypothesis (CHH). We tested whether the CHH could explain visual comfort ratings for different natural image categories and natural images with amplitude and phase spectrum manipulations. Observers rated the visual comfort of images from three different

categories: 10 landscape images (LSC), 10 control textures from the Brodatz database (BDZ), and 10 tryphobic textures (TRY). Each image was presented in three versions: (1) its original form (or), (2) with its phase-spectrum randomized (ps), and (3) with its amplitude spectrum set to $1/f$ (1f). These same images were passed through biologically plausible population models of V1 neurons, from which we measured the population kurtosis and population activity as measures of coding efficiency. Restricting our analysis to non-tryphobic images, we found that consistent with the CHH, the landscape images were more efficiently coded by the V1 models and more comfortable to view than the Brodatz textures. Also consistent with the CHH, we found for both categories (LSC, BDZ) the original images were more efficiently coded by the V1 models than their phase-scrambled controls, and were also more comfortable to view. For the tryphobic imagery, we found a completely different pattern of results. Tryphobic images were much more uncomfortable in their original form than when phase-scrambled, despite the fact that the original images were more efficiently coded. This suggests that one cannot fully explain tryphobic visual discomfort in terms of the CHH and low-level statistical properties of images. We suggest that the visual discomfort elicited by tryphobic imagery may represent an over-generalization of a disgust response to visual patterns resembling parasitism and infection, like skin diseases.

36.333 PHYSICAL INFERENCE WITH UNCONSCIOUS INFORMATION

Yilong Zhang^{1,2} (zhangyilong18@mails.ucas.ac.cn), Sheng He^{1,2}, Jiedong Zhang^{1,2}; ¹State Key Laboratory of Cognitive Science and Mental Health, Institution of Biophysics, Chinese Academy of Sciences, Beijing, China, ²University of Chinese Academy of Sciences, Beijing, China

Sensory information can be processed in the human brain without consciousness. But to what extent can subliminal sensory information influence cognitive processes? Inference is one of the high-level cognitive processes relevant to human intelligence. It is unclear whether it could be triggered by subliminal information. In this study, we designed a physical inference scenario based on collision rules to examine whether individuals could predict the physical outcomes when a visual stimulus is under consciousness. Specifically, a ball was falling onto a tilted bar, and the direction of the bar (cue) could be used to predict the direction of bounce. Continuous flash suppression (CFS) was used to mask the cue stimulus, and a discrimination task was used to assess sensitivity to the ball's subsequent trajectory. Both the behavioral performance and the electroencephalography (EEG) signal of human participants were recorded to investigate whether they could infer the physical outcome even when the cue was invisible. Behavioral results indicated that the invisible cue significantly improved the ability of discriminating the ball trajectory compared to the no-cue condition. Furthermore, task-relevant feature of the ball trajectory could be decoded from the EEG signal before the ball's actual trajectory, indicating the existence of inference based on the

invisible cue. Overall, this study demonstrates that unconscious information can trigger the intuitive physical inference process, with underlying neural mechanisms potentially related to internal forward mental simulation.

36.334 A NOVEL VR-EEG FRAMEWORK FOR STUDYING ADAPTIVE PROCESSES IN NATURAL VISION

Nawal Yahiaoui¹ (nawal.yahiaoui@ens.psl.eu), Peter Neri^{1,2}; ¹Ecole Normale Supérieure, PSL University & CNRS, Paris, France, ²Italian Institute of Technology, Erzelli campus, Genoa, Italy

The visual system maintains stable perceptual representations despite constant changes in sensory input produced by self-motion and environmental dynamics. In humans, it has been proposed that perceptual stability arises from adaptive mechanisms that flexibly encode sequences of sensory observations rather than relying purely on spatial maps. Yet how these mechanisms unfold over time and interact with behavior to support persistent visual representation remains unclear. To address this question, we developed a novel experimental framework combining Virtual Reality (VR) with frequency-tagged wireless EEG to examine visual invariance during naturalistic exploration at both behavioral and neural levels. Participants navigated a controlled, achromatic 3D environment in which objects differed only in position, orientation and size. Motor and vestibular cues were minimized by restricting physical movement. Exploration occurred either actively, through self-initiated teleportation, or passively, by following replays of each participant's own exploration trajectory. This design ensured that the two interaction modes provided nearly identical visual stimulation. Following exploration, participants completed a memory-based decision task after being teleported to a 2D bird's-eye view perspective, identifying the object to which a transformation had been applied. Our framework enabled robust decoding of stimulus-specific information via frequency tagging, despite the complexity and noise inherent to VR. Model-based decoding reliably predicted stimulus identity from EEG, and prediction accuracy differed between active and passive exploration. This difference was supported by behavioral performance, demonstrating that exploration mode modulates the encoding of scene features and the formation of invariant representations. These effects could not be attributed to differences in memory performance or attentional engagement, suggesting that exploration mode modulates perceptual encoding rather than post-perceptual processes. Together, these findings highlight the potential of combining VR and EEG to explore the adaptive mechanisms underlying visual invariance and perceptual persistence in naturalistic conditions.

Spatial Vision: Neural mechanisms

SUNDAY, MAY 17, 2:45 – 6:45 PM, BANYAN BREEZEWAY

36.335 ADAPTATION SHOWS THAT VISUAL SNOW IS RETINOTOPIC

Aislin A. Sheldon¹ (sheld173@umn.edu), Samantha A. Montoya^{2,3}, Anna J. Hillstrom⁴, Hannah R. Moser¹, Tylyn A. Page¹, Carter B. Mulder¹, Abby I. Metzler⁵, Carrie E. Robertson⁶, Michael S. Lee⁷, Stephen A. Engel⁴, Michael-Paul Schallmo¹; ¹Department of Psychiatry and Behavioral Sciences, University of Minnesota, Minnesota, ²Graduate Program in Neuroscience, University of Minnesota, Minnesota, ³Perelman School of Medicine, University of Pennsylvania, Philadelphia, ⁴Department of Psychology, University of Minnesota, Minnesota, ⁵Department of Neurology, University of Minnesota, Minnesota, ⁶Department of Neurology, Mayo Clinic, Rochester, Minnesota, ⁷Department of Ophthalmology and Visual Neurosciences, University of Minnesota, Minnesota

Visual snow syndrome (VSS) is a neurological disorder involving persistent flickering static seen across the whole visual field. VSS can interfere with daily living tasks, and the estimated prevalence is ~2% of the population. Effective treatments are limited and the neural basis is unknown. In particular, it is unclear where in the visual system the snow percept is generated. Using adaptation to dynamic noise, it has been shown that snow percepts were attenuated on the side of the display where the adapter was presented (Montoya et al, 2023). To test the spatial tuning of adaptation affecting visual snow, in this experiment we used an adaptation task that included two dynamic noise gratings: vertically oriented on one side and horizontal on the other, with contrast modulated sinusoidally over space. The wavelength of adapter gratings varied across trials. If visual snow is retinotopic, we predicted that adaptation would produce oriented bands of attenuated snow, and participants with VSS were instructed to respond when they no longer perceived attenuated stripes of snow. We expected that if the adapter wavelength was smaller than the adapted neuronal receptive fields, both left and right circles would be uniformly adapted, and participants would respond immediately. In initial testing, participants with VSS responded almost immediately for adapters with shorter wavelengths ($\leq 0.5^\circ$), and reported that the percept of snow was reduced uniformly across the entire adapter region (i.e., no stripes). Adapters with longer wavelengths produced stripes of attenuated snow for greater durations. This supports the hypothesis that the neural activity underlying visual snow is retinotopic, and that adaptation of snow operates on a spatial scale up to $\sim 0.5^\circ$. Our findings may help identify neural populations that participate in generating the percept of visual snow, based on their receptive field properties.

R01 EY036005, P41 EB015894, UL1 TR002494, and the President's Postdoctoral Fellowship Program, University of Minnesota.

36.336 SUB-VOXEL SUPPRESSIVE INTERACTIONS EXPLAIN POPULATION RECEPTIVE FIELD COMPRESSION

Reeбал W. Rafeh¹ (reeбал.rafeh@gmail.com), Geoffrey N. Ngo¹, Matthew Liaw¹, Lyle E. Muller¹, Ali R. Khan¹, Ravi S. Menon¹, Taylor W. Schmitz¹, Marieke Mur¹; ¹Western University

Population receptive field (pRF) mapping is a powerful fMRI tool for fine-grained mapping of cortical responses to visual stimulus features. Over the past decade, multiple studies have shown a non-linear relationship between pRF responses and simultaneous stimulus presentation (Kay et al., 2013; Kupers et al., 2024). Simultaneous stimuli presented within a pRF drive sub-additive responses relative to individually presented stimuli. However, there is no clear mechanistic explanation for this compression. We propose that the normalization model (Reynolds & Heeger, 2009) provides a parsimonious mechanistic account of pRF response compression, where suppressive interactions between differentially tuned populations within fMRI voxels drive response non-linearities. In this 7T frequency-tagged fMRI (ft-fMRI; Ngo, Rafeh et al. 2024) experiment, we investigate sub-voxel dynamics by stimulating distinct neural subpopulations within pRFs at temporally dissociable frequencies. Participants (n=7) distributed their attention across two simultaneously presented visual checkerboard stimuli oscillating at 0.125 Hz and 0.2 Hz. Our analyses revealed that distinct populations of visual cortical neurons synchronized to either one (singular) or both (multiplexing) frequencies. A separate pRF mapping experiment confirmed that multiplexing populations had visual field preferences that overlapped with both stimuli, whereas singular populations overlapped with only one of the two stimuli. Adapting the normalization model to our dynamic frequency-tagging paradigm, we predicted that mutual suppression between neuronal subpopulations within multiplexing pRFs should generate stronger non-linear intermodulation components in their timeseries. Confirming this prediction, multiplexing populations exhibited significantly stronger intermodulation amplitudes than singular populations. Furthermore, directing attention to either of the two stimuli significantly reduced intermodulation amplitude, corroborating our dynamic normalization model. Together, our findings support the normalization model as a mechanistic account of pRF response compression and introduce a novel ft-fMRI marker for suppressive interactions at a sub-voxel scale. In addition, our dynamic normalization model offers a framework for understanding neuronal computations in ft-fMRI paradigms.

Natural Sciences and Engineering Research Council

36.337 POLAR ANGLE ASYMMETRY OF PRF PROFILE IN FAR-PERIPHERAL VISUAL CORTEX

Pei-Yin Chen^{1,2}, Atsushi Wada^{1,2}; ¹Center for Information and Neural Networks (CiNet), NICT, Osaka, Japan, ²Graduate School of Frontier Biosciences, Osaka University, Osaka, Japan

Polar angle asymmetries in the cortical organization of human V1 have been previously revealed by fMRI retinotopic measurements in central vision. Building on our previous work, which demonstrated that population receptive field (pRF) analysis can be used to effectively

characterize retinotopic organization across both central and far-peripheral vision, this study investigates polar angle asymmetries in the far-peripheral visual cortex. Specifically, we examined how pRF parameters vary with cortical distance along horizontal and vertical meridians in areas V1 to V3. Seventeen participants viewed the retinotopic stimuli through a wide-view binocular visual stimulation system in the MRI scanner, which allowed the visual presentation to cover approximately 90 degrees of the visual field. We applied a 2D Gaussian pRF model to the BOLD activations of voxels in areas V1 to V3 to estimate the pRF profiles. Our data show that pRF eccentricity generally increased with cortical distance from the visual center. However, this increase occurred more rapidly along the horizontal meridian compared to the vertical meridian in area V1, especially in far peripheral visual field. This asymmetry may be related to the difference in horizontal and vertical extent of the visual field due to the presence of monocular region. Additionally, pRF size also increased with cortical distance and then reached a saturation point, but the saturation occurs at a larger value for the horizontal meridian compared to the vertical meridian. Interestingly, these polar angle asymmetries in pRF eccentricity and pRF size were not observed in areas V2 and V3.

JSPS KAKENHI (Japan) 24K16880, 19K12745

36.338 MODEL-BASED ESTIMATION OF THE POPULATION CONTRAST RESPONSE FUNCTION IN HUMAN VISUAL CORTEX

Ilona M Bloem^{1,2}, Louis N Vinke^{3,4}, Sam Ling^{5,6}; ¹Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, the Netherlands, ²Netherlands Institute for Neuroscience, KNAW, Amsterdam, the Netherlands, ³Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA, ⁴Harvard Medical School, Boston, MA, USA, ⁵Psychological and Brain Sciences, Boston University, Boston, MA, USA, ⁶Center for Systems Neuroscience, Boston University, Boston, MA, USA

Contrast determines how strongly visual stimuli drive neural responses and, in turn, how well we can see. This contrast-response relationship is nonlinear, with gain control mechanisms acting to compress responses to high contrast stimuli, effectively setting the contrast range to which neural responses are most sensitive. Yet, capturing these saturating responses noninvasively in human fMRI has proven challenging. We previously showed that sustained contrast adaptation can reveal saturating contrast response functions (CRF) with fMRI, but this requires long interleaved top-up adaptation periods that are inefficient and limit experimental design. Here, we introduce a model-based fMRI approach to more flexibly and efficiently estimate saturating contrast responses: population contrast response function (pCRF) mapping. First, we benchmarked pCRF mapping against traditional deconvolved BOLD responses for stimuli presented with a fast event-related design. Participants viewed stimuli at nine fixed contrast levels (2.7–96% Michelson Contrast, 2 s duration), interleaved with 16% contrast top-up periods. Our results revealed tight correspondence in parameter estimates between the

deconvolution-derived BOLD CRF and pCRF estimates, validating the model-based approach. We then tested the durability of pCRF mapping with a nonstandard design that allows for more flexibility and efficiency, but does not lend itself to traditional fMRI analyses. Participants were scanned while viewing stimuli of various contrast levels sampled from a mixed uniform and Gaussian distribution centered on 16% Michelson contrast. Stimuli were shorter (0.5 s) and no top-up adaptation periods were used, resulting in a continuous sequence of contrast events. Sampling from this biased distribution preserves coverage of the full contrast range while maintaining adaptation near the adapter contrast. Voxel-wise pCRF estimates could still be recovered from this unconventional design, demonstrating that model-based pCRF mapping technique broadens available experimental designs while still capturing the saturating nonlinearities that are fundamental to neural population responses.

NIH-EY028163 and NSF-1625552

36.339 VISUAL PROFILE ANALYSIS IN A CONTRAST DISCRIMINATION TASK

Farhan Abdul Vaheed¹, Allison B. Sekuler^{1,2,3}, Patrick J. Bennett¹; ¹Department of Psychology, Neuroscience, and Behaviour, McMaster University, ²Rotman Research Institute, Baycrest Academy for Research and Education, Toronto, Canada, ³Department of Psychology, University of Toronto

How do we discriminate visual patterns? One view is that discrimination is based on changes in the distribution of responses across multiple local filters tuned to different spatial frequencies, orientations, and directions of motion. However, some evidence suggests that spatial-interval (Morgan & Ward, JOSA, 1985) and relative phase (Badcock, Vision Res, 1984) discrimination utilizes representations that explicitly represent the spatial characteristics of visual stimuli. Here, we examine if observers use changes in the luminance spatial profile to discriminate patterns in a putative contrast discrimination task. The stimuli were $f + 2f$ compound gratings. The contrast of f and $2f$ were 0.1 and 0.05, respectively, and a 3IFC task was used to measure discrimination thresholds for a contrast increment added to $2f$. Thresholds were measured in a fixed baseline condition and in conditions in which we randomized stimulus contrast (± 0.3 log units), spatial frequency (± 0.3 log units), or orientation ($\pm 15^\circ$). In the compound random conditions, randomization was applied to the compound stimulus, where both components changed by equal amounts, and therefore the stimulus luminance profile remained unchanged. In the component random conditions, randomization was applied to only one grating component, and therefore, the spatial profile was altered. We found that thresholds in the component random conditions, but not the compound random conditions, were significantly greater than the threshold in the fixed, baseline condition. In other words, discrimination thresholds were nearly constant in conditions where observers could base discrimination on changes in the luminance spatial profile, but were much higher when that cue was not available. These findings suggest

that if discriminations are mediated solely by spatially localized filters, then those filters must encode aspects of local waveform shape, and that the integration of information across channels is disrupted when the pattern's luminance profile is changed.

NSERC

36.340 RADIAL-TANGENTIAL ANISOTROPY OF POSITIONAL ACUITY IN THE PARAFOVEA

Susana Chung¹, Kathlyn Bako¹, Rijul Soans¹; ¹University of California, Berkeley

Our ability to detect small positional differences between objects is very exquisite. The precision of positional acuity depends on many factors. In this study, we investigated how positional acuity changes with stimulus orientation and locations in the visual field. Positional acuity was assessed using a three-dot bisection task. In each trial, a target dot appeared at one of seven positions between two reference dots. Observers (N=8) judged whether the target dot was closer to either of the reference dots. From the psychometric function relating the percent-response toward one reference dot's direction to the target dot location, we derived positional acuity (the positional shift of the target dot corresponding to a change in response probability from 50% to 84%) and the point of subjective equity (PSE), representing judgment bias. Measurements were made with stimulus dots aligned horizontally, vertically or obliquely (45° or 135°) at the fovea, and at 4° and 8° eccentricities in the nasal, inferior, and inferior-nasal visual fields. Dot size and separation in the parafoveal locations were scaled according to the cortical magnification factor. Each condition was repeated 3-4 times. As expected, positional acuity was best at the fovea and declined progressively with increasing eccentricity. At the fovea, positional acuities were similar regardless of dot alignments. At 4° and 8° eccentricities, positional acuities were worst when dots were aligned radially relative to fixation and best when dots were aligned tangentially. Across observers, the radial-tangential ratio of positional acuity was 1.01 at the fovea (implying isotropy), 1.79 and 2.14 at 4° and 8° eccentricities (implying anisotropy), respectively. PSE did not vary systematically with any conditions. The shift of positional acuities from being independent of stimulus orientations at the fovea to demonstrating strong radial-tangential anisotropy in the parafovea is consistent with how visual processing differs between the fovea and the parafovea.

NIH Research Grant R01-EY036054 and Training Grant T35-EY007139

Spatial Vision: Crowding, eccentricity

SUNDAY, MAY 17, 2:45 – 6:45 PM, BANYAN BREEZEWAY

36.341 REDUNDANCY MASKING EXPLAINS UNDERESTIMATION AND THE RADIAL-TANGENTIAL ANISOTROPY OF NUMEROSITY PERCEPTION

Li L-Miao¹ (miao.li@ens.psl.eu), Y. Emre Turkmen¹, Bilge Sayim¹; ¹École Normale Supérieure, PSL University, CNRS

A central question in numerosity perception is whether the visual system has a dedicated “number sense” that extracts numerosity directly from visual input. Numerosity perception is subject to a radial-tangential anisotropy: items arranged radially are perceived as less numerous than items arranged tangentially. Numerosity is usually underestimated. Redundancy masking (the reduction of reported numbers in repeating patterns; RM) has been proposed as a mechanism underlying both underestimation and this anisotropy; however, a direct link between RM and typical estimation-range numerosity perception has not been established. We generated radial and tangential arrangements by placing discs within an angular region extending from fixation, with the opening angle determining numerosity. Five reference numerosities (9, 13, 18, 24, 30) with probes at $\pm 1-4$, and a small-numerosity condition (3, 4, 5) with ± 1 probes, were tested. Discs formed local triplets oriented radially or tangentially, with slightly larger spacing to the remaining discs. Each local triplet reproduced the classical RM stimulus, and larger numerosities were constructed from multiple triplets. On each trial, a radial and a tangential arrangement were presented to the left and right of fixation (one reference, one probe). Participants judged which side was more numerous. As expected, small numerosities showed a robust radial-tangential anisotropy. Importantly, large numerosities showed the same pattern: radial arrangements were consistently perceived as less numerous than tangential arrangements. The persistence of the anisotropy when multiple triplets mapped onto a larger space suggests that the numerosity difference (relative under/over-reporting in radial and tangential arrangements, respectively) observed for small numerosities generalizes to larger numerosities. Compression of visual information emerges with as few as three items and persists across larger numerosities, suggesting that limitations in spatial vision fundamentally shape perceived numerosity. While a dedicated number-sense mechanism may contribute, RM may be the primary mechanism driving underestimation and radial-tangential anisotropy in numerosity perception.

ANR-24-FRAL-0016-01

36.342 NON-MONOTONICITY OF THE RADIAL VS. TANGENTIAL ANISOTROPY IN CROWDING

Daniel R Coates¹ (drcoates@uh.edu), Susana TL Chung²; ¹University of Houston College of Optometry, ²UC Berkeley

Precisely determining anisotropies and eccentricity-dependencies remains a crucial method to characterize phenomena in spatial vision. Visual crowding yields numerous such quantities, such as the change in critical spacing (CS) vs. eccentricity, and differing effects of radial and tangential flanker arrangements. Six observers identified the

orientation of a Tumbling E in the fovea and in the right visual field at 2, 5, 10, 20, and 30 degrees from the fovea. Targets were presented alone or flanked by two radial or tangential flankers in a variety of sizes and spacings controlled by an adaptive psychophysical procedure. The resultant threshold size vs. flanker spacing curves provided measures of the critical spacing as well as unflanked acuity. As expected, both flanker arrangements yielded monotonically increasing CS (radial = $\sim 0.2\text{--}0.3 \times$ eccentricity, tangential = $\sim 0.125\text{--}0.16 \times$ eccentricity), but neither curve had a constant change vs. eccentricity. The slope of the radial CS near the fovea was 0.175 degrees of visual angle (dva) for each degree of eccentricity, peaked between 5&10 degrees at approximately 0.3 dva/deg eccentricity, before decreasing to ~ 0.2 dva/deg at 30 degrees. The slope of the tangential CS near the fovea was 0.1 dva/deg, growing monotonically to ~ 0.15 dva/deg at 30 degrees. The consequence of these differing rates is a non-constant, non-monotonic radial vs. tangential anisotropy of ~ 1 near the fovea that grows to a maximum of ~ 2.2 at 5-10 degrees, and decreases back down to ~ 1.5 at 30 degrees. These findings help constrain theories of the substrate for crowding, whether governed by cortical receptive fields, anatomical architecture, or other explanations.

36.343 FMRI DECODING REVEALS THAT VISUAL CROWDING LEADS TO INFORMATION LOSS IN THE HUMAN PRIMARY VISUAL CORTEX

Irem Yildirim¹, Hojin Jang², Frank Tong¹; ¹Vanderbilt University, ²Korea University, South Korea

Visual crowding refers to impairments in perceptual discrimination that arise when a peripheral stimulus is flanked by neighboring stimuli. fMRI studies have strongly implicated extrastriate areas V2 through V4 in the suppressive interactions associated with visual crowding, but the potential role of V1 in crowding has remained controversial. Prior studies have largely focused on the modulations in fMRI amplitude caused by flanking stimuli, but such measures are susceptible to spatial spreading of the BOLD signal from the flankers and might not reflect changes in neural processing per se. Other studies have found evidence of crowding-related modulation in V1, but in a manner that is dependent on top-down attentional selection of the crowded target. Here, we investigated whether crowding may lead to loss of target information in V1 in the absence of directed attention. Participants attended to stimuli at central fixation while white or black digit symbols were presented in the near periphery on a mean-luminance gray background. The severity of visual crowding was manipulated by presenting a target digit, with two flanking digits on either side that either matched or differed in visual polarity. We used 6 number digits as stimuli, flickering black/white rapidly for 12 seconds, while participants engaged in a color discrimination task at the central fixation. Even though target saliency did not alter the BOLD amplitude in the voxels selective for the target location in V1, V2, V3 or hV4, decoding accuracies were higher for the salient target digit in all areas. The salient target could be successfully decoded in V1 and V2 while the non-salient target was also decodable in V1. These results

indicated that V1 shows loss of target information due to strong crowding even in the absence of attention, suggesting that the presence of strong flankers results in competitive interactions as early as V1.

This research was supported by NEI grants R01EY035157 to FT and P30EY008126 to the Vanderbilt Vision Research Center.

36.344 MODELLING INFORMATION LOSS IN THE PERIPHERY

Ece Yatikci¹, Ben Lonnqvist¹, Clementine Levy-Fidel¹, Marco Bertamini², Michael Herzog¹; ¹Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, ²Department of General Psychology, University of Padova, Padova, Italy

Peripheral vision is often assumed to be a low-resolution version of the fovea. While blur clearly contributes to the loss of fine detail in the periphery, it is unclear whether other mechanisms are also involved in peripheral perception. The Honeycomb illusion illustrates this loss, as small lines (barbs) that are visible in the fovea disappear in the periphery, while the underlying grid texture remains visible (Bertamini et al., 2016). This disappearance is often attributed to low-pass filtering. To test the role of blur, we tested performance with the Honeycomb illusion and with a visual pop-out search task in human observers (N = 20). We fitted parameters of a blur model to the Honeycomb illusion data and evaluated the same model on the search task. While blur captured key aspects of barb disappearance in the Honeycomb illusion, it failed to account for human performance in visual pop-out search. To test whether representational compression could explain peripheral perception, we trained autoencoders with latent bottlenecks, reflecting reduced cortical resources in the periphery. Two of the four models reproduced barb disappearance in the Honeycomb illusion, but all failed on pop-out search as the odd man out target was not preserved in the latent representation. Our findings show that no single model is sufficient to account for peripheral perception. Phenomena that appear compatible with simple low-pass filtering, as in the Honeycomb illusion, may depend on additional mechanisms, including compression.

36.345 ONE INTENSITY, DIVERGENT OUTCOMES: VISUAL FIELD ASYMMETRIES DICTATE RESPONSE TO HIGH-FREQUENCY TRANSCRANIAL RANDOM NOISE STIMULATION (HF-TRNS)

Irem Akdogan^{1,2} (hergunerirem@gmail.com), Simay Uner^{1,2}, Hulusi Kafaligonul^{1,2,3}; ¹Department of Neuroscience, Aysel Sabuncu Brain Research Center, Bilkent University, Ankara, Türkiye, ²National Magnetic Resonance Research Center (UMRAM), Bilkent University, Ankara, Türkiye, ³Neuroscience and Neurotechnology Center of Excellence (NÖROM), Department of Anatomy, Faculty of Medicine, Gazi University, Ankara, Türkiye

Peripheral visual processing is inherently non-uniform, characterized by robust asymmetries in polar angle. The Horizontal-Vertical Anisotropy (HVA) reflects superior performance along the horizontal meridian, while the Vertical Meridian Asymmetry (VMA) often favors the lower over the upper visual field. While the anatomical and functional bases of these inhomogeneities are well-documented, the degree to which they can be modulated under targeted neuromodulation protocols remains unclear. This study systematically investigated whether tRNS alters the contrast sensitivity function (CSF) across the visual field and whether such effects depend on the electrical stimulation frequency. Participants performed an orientation discrimination task at four isoeccentric visual field locations (left, right, upper, and lower) under either high-frequency (hf-tRNS) or low-frequency (lf-tRNS) stimulation conditions (1.5 mA), as well as baseline controls. CSF data confirmed the expected asymmetry pattern via maximum contrast sensitivities, indicating the presence of HVA and VMA, both at baseline and tRNS conditions. Moreover, tRNS applications revealed modulation patterns that were strictly specific to the electrical stimulation frequency. Hf-tRNS significantly enhanced contrast sensitivity across the spatial frequencies in the lower visual field. However, this was location-specific, as it attenuated sensitivity in both horizontal locations—where anatomical and functional processing is naturally strongest—independent of spatial frequency. The upper visual field showed no reliable modulation. Conversely, lf-tRNS failed to induce significant changes in any visual field location. These results provide the first comparative mapping of peripheral CSF under different tRNS types. A fixed stimulation intensity elicits divergent effects—facilitation in the lower field versus reduction in the horizontal meridian—illustrating that the potential for plasticity and neural gain varies significantly across the periphery, while asymmetries were preserved. Together, these findings underscore intrinsic biological constraints of polar angle asymmetries and suggest that effective neuromodulation of peripheral vision may require frequency- and location-specific tuning of intensity rather than a uniform stimulation approach.

This work was supported by The Scientific and Technological Research Council of Türkiye (ARDEB 424K278, BİDEB 2211 Program) and the BAGEP Award of the Science Academy-Türkiye.

36.346 PERSISTENT POLAR-ANGLE ASYMMETRIES IN SYSTEM-LEVEL COMPUTATIONS DESPITE M-SCALING

David Tu¹, Shutian Xue¹, Jonathan Winawer¹, Marsia Carrasco¹; ¹NYU

[Introduction] Performance is location-dependent: Typically, it decreases with eccentricity and is higher along the horizontal than vertical meridian—horizontal-vertical anisotropy (HVA)—and along the lower- than upper-vertical meridian—vertical meridian asymmetry (VMA). Both the eccentricity effect and polar angle asymmetries have been related either to cortical surface area or to system-level computations—gain and internal noise. Here we used an equivalent noise protocol to investigate whether and how scaling stimuli size to equate cortical surface area across locations (M-scaling) modulates

spatial patterns in performance and system-level computations. [Method] Observers performed an orientation-discrimination task on a Gabor signal embedded in dynamic white noise, presented either at the fovea or 8° to the left, right, above, or below fixation. Stimuli were equated either for retinal size or for cortical size, based on individual retinotopic maps. We estimated contrast thresholds (using a functional adaptive sequential testing method) and fit a perceptual template model to estimate gain and internal noise at each location. [Results] Without external noise, M-scaling eliminated contrast threshold differences across eccentricity, but only reduced polar angle asymmetries. With equal retinal size, internal noise, but not gain, increased with eccentricity. M-scaling eliminated the eccentricity effect in internal noise and gain was higher in the periphery than at the fovea. With equal retinal size, internal noise did not vary around polar angle, and gain was higher along the horizontal than vertical meridian and along the lower- than upper-vertical meridian. With M-scaling, HVA in gain was eliminated, and VMA emerged for internal noise. [Conclusion] Visual performance differences across eccentricity and around polar angle appear to arise from different computations, the former from differences in internal noise and the latter from differences in gain. Critically, despite equated cortical surface area across locations, polar angle asymmetries in performance and in system-level computations are attenuated but persist.

NIH NEI R01-EY027401 to MC and JW

36.347 A START-SITE COST OF MICROSACCADE PREPARATION: INCREASED FOVEAL CROWDING AT THE PREFERRED LOCUS OF FIXATION

Krishnamachari Prahalad¹ (pkrishn5@ur.rochester.edu), Martina Poletti¹; ¹University of Rochester

Microsaccades frequently occur during examination of complex foveal stimuli as they shift the preferred locus of fixation (PLF) and support fine spatial vision within the central 1°. Before a microsaccade, processing resources are allocated toward its target, improving discrimination at the upcoming landing site. Here we investigated whether this preparatory state alters object recognition at the start-site by testing whether crowding strength at the PLF changes immediately before microsaccade onset. Eye movements were recorded with a high-resolution Dual-Purkinje Image (dDPI) eye tracker, providing arcminute-level precision necessary to localize gaze and deliver retinally contingent stimuli. Ten observers reported the identity (4AFC) of a central Pelli-font digit presented either alone or with horizontal flankers while maintaining fixation or preparing a cue-directed microsaccade toward a 20° off-center target. A simultaneous orientation-discrimination task at the microsaccade target location verified pre-microsaccadic allocation of processing resources. Crowding was robust in both conditions, but significantly stronger prior to microsaccade onset. Threshold elevation (flanked minus unflanked) increased from 0.06 ± 0.05 logMAR to 0.11 ± 0.03 logMAR in the pre-microsaccade interval (paired t-test: $t(9) = -2.79$, $p = 0.021$). Critical spacing also increased (center-to-center: $t(9) = -2.59$, $p = 0.029$),

indicating an expansion of foveal crowding zones. These effects occurred despite no change in unflanked acuity ($t(9) = 0.37$, $p = 0.72$) and no difference in mislocalization rates ($t(8) = 0.54$, $p = 0.60$). These results suggest that microsaccade preparation, rather than simply increasing internal noise or positional uncertainty, selectively enlarged spatial integration zones at the PLF. This work reveals the cost of microsaccade preparation at the start-site; while discrimination at the microsaccade goal improves, signal pooling at the current fixation locus increases. This redistribution of perceptual resources highlights how foveal processing is reweighted before microsaccade onset not only at its landing site but also at the PLF.

This work was funded by NIH Grant EY029788 to MP and NIH grant EY001319 to the Center for Visual Science.

36.348 HOW DOES CROWDING ALTER POPULATION SPATIAL FREQUENCY TUNING?

Feiyi Wang¹, Emily Wiecek^{1,2,3}, Sam Ling¹; ¹Department of Psychological & Brain Science, Boston University, ²Department of Ophthalmology, Boston Children's Hospital, ³Harvard Medical School

Visibility of a target can be impaired when closely flanked by other stimuli, a phenomenon known as crowding. Crowding tends to be feature-tuned, wherein suppression is more potent when the target and flankers share similar features, such as spatial frequency (SF). Here, we used fMRI to assess the visuocortical mechanisms underlying crowding, and its impact on SF processing. To do so, we leveraged population spatial frequency tuning (pSFT), a model-based technique for characterizing voxel-wise SF tuning, under crowded and uncrowded conditions. In the uncrowded condition, participants viewed targets (eccentricity: 7.85 dva; subtending 7.5 x 3.5 dva), which swept through a range of SF's (bandpass-filtered noise, 0.5 cpd-12 cpd) to map voxel-wise pSFT tuning, across V1-V3. In the crowded condition, participants were shown the same SF-filtered noise targets but surrounded by two flankers (center-to-center distance: 3.6 dva). To examine the influence of flanker SF on target tuning, flankers were SF-filtered to be either above (high-SF: 3 cpd) or below (low-SF: 1 cpd) the average preferred SF of voxels corresponding to the target (2 cpd). We found that crowding shifted population SF tuning preferences: a majority of the voxels within the target area shifted their peak tuning to lower SF's when the target was crowded by high-SF flankers. In extrastriate cortex, tuning shifted to higher SF's when the target was crowded by low-SF flankers. Our results show that crowding can shift population tuning profiles, but that this shift is dependent on match between flanker SF content and the preferences of a target, as well as visual area. This tradeoff in modulatory effects as a function of area and flanker SF may be driven by the systematic drop-off in SF sensitivity along the visuocortical hierarchy.

36.349 THE EFFECTS OF CROWDING ON PERIPHERAL RELATIVE PHASE DISCRIMINATION

Emma Neto¹, Allison B. Sekuler^{1,2,3}, Patrick J. Bennett¹; ¹Department of Psychology, Neuroscience & Behaviour, McMaster University,

²Rotman Research Institute, Baycrest Academy for Research and Education, ³Department of Psychology, University of Toronto

Relative phase discriminations are made with even-symmetric and odd-symmetric visual mechanisms (Field & Nachmias, Vision Res, 1984). In central vision, even- and odd-symmetric mechanisms are approximately equally sensitive, but it is unclear how sensitivity changes in peripheral vision. Using compound gratings, Bennett and Banks (Nature, 1987) found that sensitivity of odd-symmetric mechanisms declined dramatically in the periphery; however, using stimuli consisting of single, isolated features, Morrone et al. (Vision Res, 1989) found that even- and odd-symmetric sensitivity remained approximately equal across the visual field. Neto et al. (JOV, 2025) investigated whether these different results reflected phase-specific effects of visual crowding. Consistent with this idea, they found that increasing the number of stimulus features, and decreasing inter-feature spacing, reduced odd-symmetric but not even-symmetric sensitivity. Here we examine the crowding phenomenon by measuring the effects of adjacent flankers on phase discrimination thresholds. Phase discrimination thresholds were measured in four observers. Target stimuli consisted of a single phase-reversing feature (i.e., bright/dark bars or left/right edges). We found evidence of crowding in both tasks: bar and edge discrimination thresholds increased with increasing flanker contrast and with decreasing target-flanker spacing. However, edge discrimination varied far more drastically with flanker contrast and spacing than did bar discrimination. Importantly, bar and edge detection thresholds varied little with flanker contrast and spacing, so the inability to discriminate relative phase was not due to a failure to detect the target stimuli. These results are consistent with the hypothesis that odd-symmetric mechanisms are especially sensitive to crowding in the visual periphery.

NSERC

36.350 COMBINING CROWDING AND MASKING REVEALS WHEN FRAGILE PERCEPTS BECOME ROBUST

Melissa Allouche¹ (melissa.allouche@epfl.ch), Lisa Schwetlick^{1,2}, Michael Herzog¹; ¹Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), ²École Normale Supérieure

In crowding, perception of a stimulus deteriorates when flankers are added around it. Feedforward models of vision predict that as more flankers are added, crowding increases and thus performance further decreases. However, increasing the number or complexity of flankers can improve performance, a phenomenon known as uncrowding. This occurs especially when flankers form coherent structures (e.g., a cube versus the scrambled lines of the cube). Uncrowding requires stimuli to be presented for a minimum of 160ms before it emerges. It has been proposed that this comparatively long time suggests the involvement of feedback processes. Using backward masking, a tool that has been shown to disrupt feedback, we investigate whether uncrowding can be disrupted. Participants performed a vernier discrimination task in three

conditions: vernier in isolation (baseline), vernier flanked by lines (typically crowded), and vernier flanked by cubes (typically uncrowded) where the target was presented for 120ms. The vernier was followed by a mask of randomly oriented lines. The mask reduces performance in all conditions, but this effect is particularly strong in the cubes condition, where uncrowding is clearly disrupted. A mask duration of as little as 10ms is sufficient to induce this effect. If a flanker-only preview is presented, i.e., the cubes without the vernier, ideally 60-120ms before the target, participants perform as well as in the no mask condition. Our first result demonstrates that after a presentation time of 120ms, the cube perception, and therefore uncrowding, is still fragile and can be disrupted by a mask. Our second result shows that this processing can be kick-started with 20ms flanker-only previews and suggests that once this processing is complete, the cubes perception becomes robust and immune to the mask. Together, the preview and the mask provide a toolkit to investigate the time course of object recognition.

36.351 CONTRAST SENSITIVITY AND NOISE PROCESSING IN ADULTS WITH ADHD

Hani Tsruya¹, Maria Lev, Uri Polat; ¹Bar Ilan University

Contrast sensitivity (CS) is a fundamental low-level visual function that serves as a building block for overall visual perception and underlies early visual processing. Recently, we demonstrated that adults with ADHD show increased susceptibility to visual crowding compared to controls. The present study examined whether more basic visual functions, specifically CS and signal extraction under noise, may account for these differences. Forty-five adults participated (23 controls, 22 with ADHD). CS was assessed using a contrast detection task with single Gabor patches (uncrowded condition) at spatial frequencies of 4, 6, 9, and 12 cycles/degree and stimulus durations of 40 and 80 ms. In addition, a letter identification task presented with several levels of external noise and exposure durations of 30 and 60 ms was used to evaluate mechanisms of signal extraction and noise exclusion under constrained conditions. Adults with ADHD showed reduced CS relative to controls, particularly at higher spatial frequencies and shorter stimulus durations. In the noise task, both groups showed a significant decline in performance as external noise increased. A significant main effect of duration was observed, with better identification at 60 ms compared to 30 ms; however, no significant differences were found between the ADHD and control groups. These findings indicate that perceptual processing in ADHD is compromised at early stages of visual processing, particularly in contrast detection under spatial and temporal constraints. The absence of group differences in the noise task suggests that external noise exclusion and internal noise levels may be relatively preserved, although signal strength encoding may be reduced. Overall, the results support the view that visual differences in ADHD reflect limitations in basic sensory encoding rather than solely deficits in higher-level control processes.

SUNDAY AFTERNOON POSTERS IN PAVILION

Color, Light and Materials: Neural mechanisms

SUNDAY, MAY 17, 2:45 – 6:45 PM, PAVILION

36.401 TEN MICE SUFFICE: VALIDATING SMALL SAMPLE FACTOR ANALYSIS IN VISION SCIENCE USING SAMPLING DISTRIBUTIONS AND PHOTOPIGMENT FACTORS FROM SPECTRAL SENSITIVITY DATA

David Peterzell^{1,2,3} (davidpeterzell@mac.com); ¹Fielding Graduate University, ²University of Florence, ³Color and Vision Network (CVNet)

Previously, factor analysis of spectral sensitivities from 86 bioengineered mice expressing human M and L photopigments (Jacobs, Williams, Cahill & Nathans, 2007) revealed that individual variability reflected underlying biological mechanisms rather than measurement error (Peterzell, Bloxham, & Jacobs, 2017 ECVF). Two primary factors, explaining >92% of variance, coincided with known M and L opsin absorption spectra, validating that factor analysis can extract number and tuning of known visual mechanisms from high-quality physiological (electroretinogram) data. A third factor (peak ~560 nm) explained 3.4% more variability due to tiny variability (<0.2nm) in peak wavelengths of measured spectra. The present study examined minimum sample size requirements for detecting this factor structure using systematic sampling distributions. Random samples of 3-85 mice (100 iterations per sample size) were drawn, computing correlations between each sample's correlation matrix and the full 86-mouse primary correlation matrix. Additionally, principal component analysis with varimax rotation was performed on each sample, determining component retention using 95% variance criteria. Results revealed that samples as small as 10 mice reliably reproduced the correlation structure (mean Pearson $r = 0.96$, $SD = 0.03$) and consistently extracted 3 components. Sizes 8-9s showed good reliability (mean $r = 0.93$ - 0.96), with greater variability. Below 8s, estimates were less stable but even r 's for 4s were 0.9 - 0.99 in 42% of cases. Factor loadings from standardized data showed the expected two factor M+L structure with small lambda max variability (third factor) across sample sizes ≥ 10 . These findings challenge conventional recommendations requiring large samples ($N > 100$ - 200) for factor analysis. When data exhibit high signal-to-noise ratios with substantial individual variability in known mechanisms, samples of 10-12 participants (or fewer) may suffice for correlation and factor analyses, and for reanalyzing small sample experimental data. This has important implications for re-analyzing archival data and analyzing physiological and behavioral studies where large samples are impractical.

This project was begun with the late Jerry Jacobs who provided the data.

36.402 NEURAL ORIGINS OF HUE ASYMMETRIES

Jesse R. Macyszko¹, Lindsay Peterson², Michael A. Webster¹, Erin Goddard²; ¹University of Nevada, Reno, ²University of New South Wales

Color coding in the retina and geniculate is thought to depend on two opponent dimensions encoding the differences in LvsM or SvsLM cone signals. However, color percepts cannot be accounted for by these independent dimensions, suggesting later transformations of chromatic information. A striking example is that hues change more rapidly with chromatic angle around yellowish directions (one quadrant of the LM vs S cone-opponent plane) than bluish, magenta, or greenish quadrants. The basis for this asymmetry and where it arises in visual coding remain uncertain. We investigated the magnitude of these hue differences across different cortical regions using fMRI. Stimuli consisted of two different chromatic angles presented as circular 1 c/deg gratings in a 10-deg tapered aperture, and counterphased at 1 Hz. The component colors were equal in chromatic contrast but varied from 0 to 90 degrees in angular separation, or "hue contrast," centered on each of the four diagonals in the LvsM and SvsLM plane. We compared three models using representational similarity analysis (RSA) to determine which models best predicted pairwise classification of the stimuli from neural responses: color quadrant, where the quadrants rather than the contrasts predicted responses; cone-opponent contrast, as predicted by the LvsM and SvsLM space; and perceived contrast, where we used behavioral measures of the hue differences to rank contrasts between quadrants. Color quadrant predicted responses similarly across all cortical ROIs. However, perceived rather than cone-opponent contrast predicted responses better in early cortical regions. This suggests that the asymmetry in hue percepts in color coding arise as early as V1.

Supported by EY-010834 to MW and the Australian Research Council (ARC) (DP220100747 to EG)

36.403 DO CORTICAL RESPONSES MEASURED BY VISUAL EVOKED POTENTIALS CAPTURE ASYMMETRIES IN CONE-OPPONENT MECHANISMS?

Ilgin Cebioglu¹, Jasna Martinovic², Laura K. Young¹, Gabriele Jordan¹, Quoc C. Vuong¹, Anya Hurlbert¹; ¹Newcastle University, ²University of Edinburgh

Sorting socks, picking peaches, or following footballers are tasks that typically require differentiating colours. Chromatic discrimination ability is constrained by low-level factors, e.g. spectral sensitivities of the retinal cone types, yet influenced by higher-level cortical processes. At which level and to what extent these might compensate for sensory weaknesses are open questions. Here we determine whether differences in behaviourally measured chromatic discrimination thresholds along the L-M ('red- green') and S-(L+M) ('blue-yellow')

cone-opponent axes are captured by differences in cortical visual evoked potentials (VEP). Seventeen participants with normal trichromacy (aged 18-32; 11 females) were assessed on the Colour Assessment and Diagnosis (CAD) test. Separately, participants completed an EEG task, in which they viewed chromatic sinusoidal gratings modulated along one of four unipolar directions (red, green, blue and yellow) at four chromatic contrast levels, at individually-determined isoluminance levels. Participants completed 60 trials per direction in randomised blocks. The amplitude and latency of VEPs were determined from the Oz, O1 and O2 electrodes. CAD thresholds varied with colour direction ($p < .001$), with higher thresholds for blue and yellow vs. red and green, and for blue vs. yellow. The amplitude and latency of the N1 component varied with chromatic contrast, in line with previous literature; the cortical contrast response gain was determined from the slope of these changes. Both amplitude and latency slopes varied significantly with colour direction ($p < .001$ and $p = .006$), with higher slopes for red and green vs. blue and yellow. Unlike CAD thresholds, slopes did not differ significantly between blue and yellow directions. The results suggest that changes in amplitude and latency of the N1 component capture asymmetries in suprathreshold chromatic sensitivity between the two cone-opponent axes, corresponding to asymmetries observed at threshold, supporting their future use in studies of potential cortical compensations for sensory weaknesses.

36.404 CHROMATIC EVOKED POTENTIALS PRODUCED BY SINGLE-OPPONENT AND DOUBLE-OPPONENT CELLS IN HUMAN V1: SPATIAL FREQUENCY ANALYSIS

Valerie Nunez^{1,2}, James Gordon^{2,3}, Robert Shapley²; ¹Renaissance School of Medicine, Stony Brook University, ²Center for Neural Science, NYU, ³Hunter College, CUNY

This is about testing the hypothesis that color-responsive neurons in Human primary visual cortex, V1, can be partitioned into two groups: single-opponent (SO) and double-opponent (DO) as were found in monkey V1. The experimental test was measurements of Chromatic Visual Evoked Potential (cVEP) responses vs spatial frequency in 16 human observers. Observers viewed equiluminant "red-green" checkerboard patterns modulated from 0 to 0.16 Michelson contrast in the L-M/M-L direction of DKL color space. Temporal color modulation was a rectangular wave, 0.5s on, 1.5s off. The checkerboard was equiluminant with a gray 18deg by 10deg Background (luminance 30 cd/m²); during stimulus-on, checkerboard replaced a circular region of Background roughly 2.5deg in diameter. Check size was varied to vary the dominant spatial frequency (SF) of the checkerboard from 0.6 up to 9.0 c/d. Observers also viewed color-modulation of spatially uniform fields. We identified the SO cVEP as a rapid transient (peak time ~80 ms) that was present only for checkerboards at low SF or for uniform fields; the SO waveform was not observable for SF>1.7 c/d. Separately, there was a large negative peak (peak time ~ 120-150 ms) that we and many others have assigned to the DO population. The cVEPs of DO cells were SF-tuned with a peak response ~ 3 c/d. The

color responses of Human SO and DO cells have distinct temporal and spatial signatures suggesting they are indeed independent.

This project was supported by the US National Science Foundation (grant number 1753846)

36.405 CORTICAL REPRESENTATION OF COLORS DURING HUE-SCALING AND COLOR-CATEGORIZING TASKS

Lee Lin¹ (lee.kevin.lin@gmail.com), Shunsuke Takano¹, Takato Arima¹, Kenichi Ueno², Ichiro Kuriki¹; ¹Graduate School of Science and Engineering, Saitama University, Saitama, Japan, ²Center for Brain Science, RIKEN, Wako-shi, Saitama, Japan

Humans can perceive millions of distinct color shades, yet they typically describe these shades using just tens of color terms. Understanding how these two distinct representations emerge in the brain remains an open question. Here, we investigated the neural representations underlying color appearance and color categories, a fine-grained representation and a coarser representation that often engages terms, respectively, using functional magnetic resonance imaging (fMRI) technique. We used checkerboard patterns with 12 hues evenly sampled from the isoluminant plane of cone contrast space. Participants viewed this identical set of stimuli while performing one of three tasks. In the color categorization task, participants pressed one of five buttons to report a perceived color category. In the hue-scaling task, participants reported the color appearance of the stimulus by pressing buttons three times to indicate the relative component of the four primary colors: red, green, blue, and yellow. In the baseline task, participants were asked to ignore colors and conduct an attention-demanding task using an RSVP (rapid visual stimulus presentation) letter stream at the stimulus center. Each voxel's responses across the 12 hues were fit with multiple von Mises functions to increase signal-to-noise ratio. Fitted responses at the 12 hues were then used for representational similarity analysis (RSA) and multidimensional scaling (MDS). Using the categorical clustering index in Brouwer and Heeger (2013) to the 2-dimensional MDS results, we found that early visual areas exhibited higher categorical clustering during the hue-scaling (appearance) task than during the category task. This suggests that appearance-based processing may rely on relatively small number of hue-selective mechanisms than previously assumed. Together with other groups' reports, our findings indicate that categorical information is not restricted to higher-level visual areas but is evident even in early visual cortex, and the task demands shape how color representation is expressed across the visual areas.

The Japanese Society for the Promotion of Sciences (JSPS) KAKENHI No. JP24H00702 to IK.

36.406 UNCOVERING THE NEURAL REPRESENTATIONS OF COLOR APPEARANCE

Ramith Ambati^{1,2} (ambati.10@osu.edu), William Narhi-Martinez^{3,4}, Angela Brown^{1,2}, Julie Golomb³, Delwin Lindsey^{2,3}; ¹The Ohio State

University College of Optometry, ²Graduate Program in Vision Science, The Ohio State University, ³Department of Psychology, The Ohio State University, ⁴Department of Psychology, Yale University

Humans can perceive a rich set of colors, but the neural mechanisms underlying color appearance are not fully understood. One influential theory, Hering's "color-opponent" theory, asserts that hue appearance is governed by neural responses signaling elemental redness, greenness, blueness, and yellowness sensations, which are organized into two color-opponent processes: redness-greenness and blueness-yellowness. Despite abundant behavioral evidence consistent with Hering's theory, no one has found corresponding hard-wired neural processes in cerebral cortex. Inspired by these conflicting findings, we used multi-voxel pattern analysis and Representational Similarity Analysis (RSA) to investigate the neural representation of hue in human cortex. Ten participants performed a series of Hering-style hue-scaling tasks in the fMRI scanner. On each trial, they viewed a rotating spiral colored with one of 8 hues: their predetermined personal unique red, yellow, green, and blue hues, and their binary-composite hues: orange, lime, cyan, and purple. During each block of trials, participants performed one of four hue-scaling tasks (redness, yellowness, greenness, or blueness), where they rated with 4-alternative button press ("none", "some", "mostly", "all") the amount of the corresponding elemental sensation present in each colored spiral. Our analyses included whole-brain searchlight analysis using RSA to compare neural similarity patterns from cortical responses to hypothetical similarity patterns based on the hue-scaling task and/or spiral hue. These analyses revealed color-selective cortical areas in and around hV4, where color information was represented according to hue similarity, regardless of task. By comparison, dorsal regions of frontal (FEF) and parietal (IPS) cortex, which were non-color-selective under passive viewing, showed Hering-like representational patterns of hue-scaling that were shaped by both hue and task. Our results have interesting implications for Hering's theory, and hue appearance more broadly, and also demonstrate the power of individualized hue stimuli and RSA analyses to dissociate task-dependent from task-invariant representations of hue in the brain.

36.407 JOINT REPRESENTATION OF COLOR AND ORIENTATION IN HUMAN V1: INSIGHTS FROM SINGLE-NEURON RECORDINGS AND POPULATION GEOMETRY

Xizi Gong¹, Guanpeng Chen¹, Ruolin Yang¹, Shenshen Zhang¹, Xiongfei Wang², Qian Wang¹, Fang Fang¹; ¹Peking University, ²Sanbo Brain Hospital, Capital Medical University

Integrating color and form into unified object representations is a fundamental challenge for human vision. However, the cortical stage at which this binding occurs remains debated. While classical theories posit that the V1 processes color and orientation via largely segregated neuronal populations, recent macaque studies suggest these features are jointly encoded at this early stage. Yet, the critical lack of single-neuron evidence in humans leaves the mechanisms of this integration unknown. Here, we report findings from a unique large

dataset of over 600 neurons recorded from human V1. We demonstrate that a substantial subset (~20%) of neurons exhibits co-tuning to both color and orientation. Going beyond linear co-tuning, we identify significant non-linear interactions at the single-neuron level, indicating that chromatic context dynamically modulates orientation tuning. Furthermore, by characterizing the population geometry, we show that the interaction between hue and orientation accounts for a substantial fraction of the total signal variance, suggesting that feature conjunctions are explicitly represented rather than independently processed. These results provide the first single-neuron evidence in humans that color and orientation are jointly represented in V1.

36.408 MEASURING AND CONTROLLING MAGNO- AND PARVO-SYSTEM CONTRIBUTIONS TO GRATING ACUITY

George Sperling¹ (sperling@uci.edu), Lingyu Gan²; ¹University of California, Irvine, ²University of California, Irvine

Gan and Sperling, *Journal of Vision*, 2025, tested foveal grating acuity limits for red/yellow and green/yellow chromatic and for yellow/yellow monochromatic gratings, all > 30 cycles/deg. They found three classes of subjects: Some almost exclusively used the red stimulus component for acuity judgments, some used the green, some equally red and green components. Yellow/yellow luminance judgments were generally inferior to chromatic judgments except for red dominant subjects judging red-yellow gratings in which the component stripes were of similar intensities. For the three classes of subject, when there was a clear acuity minimum as a function of the intensity of variable color stripes, it occurred when the energy of the red-or-green color stripes equaled the energy of only that same color component in the yellow stripes. That acuity minima occurred at minimal selective color differences despite large between-stripe luminance differences was evidence for superior color-selective versus luminance foveal acuity processing. Here we test whether similar procedures in peripheral viewing can reveal magno-selective versus parvo-selective processing. One subject from each color-dominance class viewed peripheral high-spatial-frequency gratings (8.9-11.7 cycles/deg). To avoid eye movements, acuity gratings appeared randomly 6.5 deg to right or left of fixation for 100 msec. Subjects judged grating orientation for a full range of contrasts of the variable-color grating stripe. In a related task, subjects judged the motion direction of similarly-colored gratings moving at 12.5 Hz, a pure magno task. Unlike the foveal data, the peripheral motion contrast minima and grating minima perfectly coincided, demonstrating magno-system processing for 100 msec high-resolution peripheral gratings and for 12.5 Hz high-frequency motion stimuli for all three subject types. Conclusion: The grating acuity paradigm reveals parvo-system properties when the parvo system is used (0.5 sec, foveally) and magno-system properties when it is used (0.1 sec, 6.5 deg peripherally).

36.409 NEURAL MECHANISMS OF COLOR ENCODING IN MACAQUE V1 AT THE CENTER OF GAZE

Isabela Tellez¹, Ramon Bartolo¹, Bevil R. Conway^{1,2}, Felix Bartsch^{1,2,3}, Daniel A. Butts²; ¹Laboratory of Sensorimotor Research, Section on

Perception, Cognition, and Action, National Eye Institute, National Institutes of Health, ²Department of Biology, University of Maryland, ³Neuroscience of Attention and Perception, Neuroscience Institute, Princeton University

Color responses of foveal V1 cells are largely unknown. Because color appearance and cone distributions vary with eccentricity, the known color-tuning properties of parafoveal V1 cells may not be instructive. Here, we measured foveal V1 receptive fields (RFs, 0 to 1.5°) using chronic Utah arrays and linear v-probes in fixating alert macaques. We obtained: 1) color-tuning responses to briefly flashed color discs (1° diameter, larger than classical RFs) of varying hue and constant saturation and luminance; 2) high-resolution spatiotemporal RFs using a dynamic chromatic stimulus ("clouds", colors defined by cone-opponent mechanisms) with power across the frequency spectrum. Precise measures of eye position were obtained, providing near photoreceptor resolution. We used Fourier analysis to determine if the color tuning curves for each neuron were captured by the first or second harmonic of a 360° cycle over the color wheel, corresponding to unimodal or bimodal tuning, respectively. Of 595 neurons, 110 were color tuned (107 unimodal, 3 bimodal). The distribution of preferred hues showed a strong bias towards the blue-orange axis and green quadrant of color space, consistent with documented color tuning of parafoveal neurons. Color-tuning curves were related to the RFs by fitting linear models, with spike-triggered averages in three chromatic channels (L, M, S) as linear filters. Projecting the disc stimuli through these filters and regressing the output against spike counts explained, on average, 61±24% of the variance ($R^2 \pm SD$) of disc responses. For 21 neurons, the model captured >80% variance, for 11 neurons, >90%. These results show that a substantial fraction of foveal V1 neurons are color tuned and that a linear transfer of the RF can account for some of the color tuning, but implicate nonlinear operations to fully explain foveal color tuning.

Supported by NIH, 1ZIAEY000558; NSF, IIS-2113197; NIH, R01 EY037347-01. NIH author contributions are Works of the U.S. Government. Findings and conclusions are those of the authors and do not necessarily reflect the views of NIH or the U.S. Department of Health and Human Services.

36.410 THE EFFECT OF TASK ON NEURAL REPRESENTATION OF MATERIALS

Vatanak Lavan¹ (vl2400a@american.edu), Masataka Sawayama², Laurie Bayet¹, Bei Xiao¹; ¹American University, ²Hokkaido University

Neural representations of material perception are typically studied as solely stimulus-driven, focusing on the mapping between stimulus features and brain responses. However, people perform different tasks with the same objects rather than simply viewing them passively. Recent works show that task demands systematically constrain how stimuli are encoded in the brain. Here, we investigated whether task context modulates the neural representation of material categories. We recorded scalp EEG using a 128-channel HydroCel Geodesic

Sensor Net while six participants (ages 18-25) performed binary judgments on three different material attributes - glossiness, softness, and edibility - of objects from identical images. Our stimulus set included 144 photographs of real-world objects across 8 material categories, with 6 exemplars per category and 3 viewpoints per exemplar, under controlled lighting. Stimuli were presented across two sessions, each with three randomized blocks. Each trial consisted of a task cue (750ms), stimulus viewing (750ms), and a response period. We performed time-resolved binary decoding of event-related scalp potentials using a linear support vector machine with 41ms sliding windows (± 20 ms) and leave-one-block-out cross-validation. Behavioral results showed significant task effects on reaction time in Session 1 (edibility: 437ms < softness: 477ms < glossiness: 500ms) that disappeared by Session 2. Representational dissimilarity analysis of behavioral judgments revealed that edibility and softness judgments were category-specific, whereas glossiness judgments varied within and across categories. Subject agreement was highest for edibility ($\kappa = 0.651$) but lowest for glossiness ($\kappa = 0.366$). Neural decoding showed that task and material category were decodable above chance during the task cue and stimulus-viewing periods, respectively (~60-70% peak accuracy). However, cross-task decoding of material categories achieved accuracy similar to within-task decoding throughout stimulus presentation. Our initial results suggest that although task information is present in the brain, it exerts minimal effect on the neural representation of materials.

36.411 OCULOMOTOR CONTRIBUTIONS TO COLOR VISION

Alina Neverodska¹, Y. Howard Li¹, Eli Brenner², Jonathan D. Victor³, Michele Rucci¹; ¹University of Rochester, ²Vrije Universiteit Amsterdam, ³Weill Cornell Medical College

Previous research with achromatic stimuli has shown that eye movements contribute to a coarse-to-fine progression in visual processing during post-saccadic fixation (Boi et al., 2017): saccade-induced transients preferentially enhance low spatial frequency information, whereas improvements at high spatial frequencies emerge with prolonged post-saccadic exposure supported by fixational drift. Recent work with chromatic stimuli, however, revealed perceptual improvements at both low and high spatial frequencies as exposure increases, suggesting that color may follow different temporal dynamics than luminance. It is unknown whether these differences reflect distinct influences of eye movements on chromatic and achromatic processing. Here we examined whether eye movements contribute to the dynamics of color perception. Observers (N=6) judged the orientation ($\pm 45^\circ$) of an isoluminant red-green grating (1 or 10 cpd) embedded in a $1/k^2$ chromatic noise field following a 6° saccade. Retinal stimulation was controlled with a high-resolution digital DPI eye-tracker coupled to a system for real-time gaze-contingent display. Trials interleaved three conditions. Normal: the stimulus appeared immediately upon saccade detection and remained visible for 500 ms, followed by a new noise field. Stabilized: after the

saccade, the stimulus moved on the display to remain immobile on the retina, eliminating drift-related modulations. No-Transient: the stimulus was presented at saccade end but slowly ramped on over 1 s, eliminating the saccadic transient, before remaining visible for 500 ms. The three conditions yielded complementary outcomes, consistent with previous luminance findings: low spatial frequency performance was selectively impaired in the No-Transient condition, whereas high spatial frequency performance declined in the Stabilized condition. These results indicate that both saccade-induced transients and fixational drift support chromatic perception, as they do for luminance. Thus, the distinct post-saccadic dynamics observed for chromatic and luminance stimuli are unlikely to originate from different contributions of eye movements.

NIH EY18363

36.412 LUMINANCE CONTRAST AND JND THRESHOLD SHOW CONSCIOUS AND UNCONSCIOUS IMPACT OF LUMINANCE ON ILLUSORY COLOR SPREADING AND FIGURE/GROUND ORGANIZATION

Tiffany Little¹, Iris Wright¹, Ralph Hale¹; ¹University of North Georgia

The watercolor illusion (WCI) is a visual phenomenon in which two thin, contrasting-colored contours create an illusory spread of color into an adjacent region where no physical color is present. This effect is closely linked to perceptual heuristics supporting surface completion and figure-ground (F/G) organization, with the WCI typically biasing the region bordered by the lighter contour to be perceived as the figure. Previous research has shown that luminance differences between adjacent regions can influence the WCI's ability to bias F/G perception (Hale et al., 2025). The present study sought to determine the luminance contrast levels at which luminance begins to override the WCI, including whether effects emerge at luminance differences that appear to fall below, near, or above a just noticeable difference (JND) threshold based on observers' responses. Participants viewed images containing two regions divided by a vertical wavy contour. Each display appeared in one of three WCI conditions: illusion on the left, illusion on the right, or absent. Across trials, one of six luminance contrast levels was applied to either side of the image, producing all combinations of WCI condition and luminance contrast. In each trial, a probe appeared on the left or right region, and participants indicated which region they perceived as the figure relative to the probe. This allowed for evaluation of how luminance differences, some of which produced performance patterns consistent with subthreshold or near-threshold discriminability, modified F/G perception. Results revealed that luminance contrast significantly modulated both WCI strength and F/G assignment. Performance patterns suggested that even luminance contrasts that were likely below an observers' JND could influence F/G judgments, while higher contrasts consistently overrode the WCI's typical figure bias. These findings help clarify the luminance levels at which unconscious and conscious luminance differences differentially impact illusory color spreading and F/G organization.

36.413 LUMINANCE GRADIENTS ACT AS ILLUMINATION CUES RATHER THAN PHYSICAL CONTRAST IN AN AMBIGUOUS FIGURE/GROUND AND ILLUSORY COLOR SPREADING PARADIGM

Iris Wright¹, Tiffany Little¹, Ralph Hale¹; ¹University of North Georgia

The watercolor illusion (WCI) is a visual phenomenon in which two thin, contrasting borders create the appearance of color spreading into an adjacent region where no physical color is present. Previous research has shown that strong luminance differences between regions can compete with the WCI and disrupt its influence on F/G perception (Hale et al., 2025). The present study investigated whether the WCI would continue to bias F/G perception when luminance gradients extended beyond the contours of the stimulus, or whether luminance would dominate as the stronger cue under these conditions. Participants viewed six reversible figures, each containing left and right regions with luminance gradients traveling upward, downward, leftward, or rightward. Each stimulus appeared with the WCI on the left, right or absent, and included a probe on the left or right region. Participants indicated which region they perceived as the figure. This design allowed us to determine whether extended luminance gradients would affect the WCI's ability to influence F/G organization. Contrary to earlier findings, luminance gradients did not significantly influence F/G perception in this study. Instead, the WCI continued to bias F/G judgments as though no luminance contrast was present. We speculate that the gradients may have been perceptually discounted as illumination cues: although they produced physical left-right luminance differences (bottom-up), observers may have unconsciously inferred that the gradients indicated lighting cues rather than surface properties (top-down), effectively treating the two regions as equivalent. This would allow the WCI to exert its typical influence on F/G organization. Stimuli without the WCI were perceived as reversible regardless of gradient direction, whereas stimuli containing the WCI consistently biased perception toward the side containing the illusion. These findings challenge assumptions about the interaction between luminance and the WCI and contribute to a clearer understanding of how perceptual heuristics shape F/G organization.

Eye Movements: Models, remapping

SUNDAY, MAY 17, 2:45 – 6:45 PM, PAVILION

36.414 TASK-CONDITIONED GAZE PREDICTION FROM LANGUAGE-BASED SCENE SEMANTICS

Michelle Greene¹, Bruce Hansen²; ¹Barnard College, Columbia University, ²Colgate University

Human gaze in real-world environments is profoundly task-driven (Yarbus, 1967), yet most computational gaze models assume stimulus-driven attention—whether through bottom-up saliency (Itti & Koch, 2001), target-template matching in visual search (Zelinsky, 2008), crowdsourced localized meaning (Henderson & Hayes, 2017),

or deep neural networks trained directly on fixation data (Kümmerer et al., 2017). Because task demands are not directly observable in the image, we begin with language as a structured means of generating interpretable, top-down priority signals. We modeled three forms of semantic guidance from human image descriptions: object-based descriptions, capturing semantic relevance similar to meaning maps; navigation descriptions, capturing action-oriented guidance; and aesthetic descriptions, capturing value-based priorities motivated by foraging accounts of attention. Each description was encoded using MPNet and used to train a dedicated convolutional neural network (CNN) to predict its embedding. Using deconvolution, we obtained three interpretable semantic priority maps for each image. These maps, together with a bottom-up saliency map and center bias, served as predictors in a spatial Poisson GLM over fixation counts using five-fold cross-validation. We tested these models on two datasets: 135 sessions from the Visual Experience Dataset (Greene et al., 2024), in which participants freely navigated a campus pond, and the COCO-Search18 dataset (Chen et al., 2021), in which observers searched for objects. The navigation semantic map was the strongest predictor of gaze in the navigation dataset (mean cross-validated normalized scanpath saliency (NSS) = 0.43), outperforming object and aesthetic semantics and low-level saliency. In contrast, in the visual search dataset, the object semantic map was most predictive across both target-present (mean NSS = 0.81) and target-absent (mean NSS = 1.23) trials. These results demonstrate that scene semantics extracted from linguistic descriptions offer a powerful, task-specific account of cognitive guidance in natural images.

NSF 2522311/2 to MRG and BCH.

36.415 HOW PREDICTIVE RECONSTRUCTION AND FIXATION INTEGRATION CREATE HIGH-RESOLUTION VISION FROM SPARSE SAMPLES

Akihiro Maruya¹ (am6949@columbia.edu), Hossein Adeli¹, Tian Zheng², Nikolaus Kriegeskorte¹, Ning Qian¹; ¹Zuckerman Institute, Columbia University, ²Department of Statistics, Columbia University

Human vision operates with sparse sampling: only the fovea provides high-resolution input, and each saccade provides a partial glimpse of the scene. Yet we experience an illusion of uniformly high resolution, suggesting that the brain predicts peripheral structure and integrates information across fixations, though the underlying computations remain unclear. We introduce two modeling approaches that test how predictive reconstruction and fixation-to-fixation integration may support this perceptual stability. We trained vision Transformers with self-supervised objectives on CIFAR-10 in two ways: masked auto-encoding where a certain ratio of image patches is made blank (ViT-Blank), and a biologically inspired one with blurred versions (ViT-Blur). Both models learned to reconstruct full images, and generalization was measured by reconstruction error on rescaled ImageNet-1K images. Unsurprisingly, across ratios and depths of Transformer blocks, reconstruction MSE was consistently lower for ViT-Blur, with especially large differences at high masking. At mask ratio 0.99, MSE

was ~ 0.06 for ViT-Blank but only ~ 0.01 for ViT-Blur, indicating substantially better recovery of global structure and color. We then evaluated the learned representations with linear CIFAR-10 classification. Classification showed the same pattern: mean accuracy was 0.749 for ViT-Blur, 0.707 for ViT-Blank, and 0.626 for controls, with the largest differences at high mask ratios and shallower depths. To examine integration across saccades, we developed a fixation-based model trained on ImageNet-1K. Each fixation produces an image with blur and down-sampling from fovea to periphery according to cortical magnification; the model receives the current view and previous ones, processed by a ViT. A Conv-GRU integrates these glimpses using corollary discharge signals, and a decoder reconstructs the head-centered image. A reinforcement-learned policy selects fixation sequences that maximize reconstruction improvement. Reconstructions improved over four fixations, reaching validation MSE ≈ 0.008 . Our framework suggests that change blindness arises when a modification remains statistically consistent with the brain's internal reconstruction.

36.416 A POPULATION CODING APPROACH TO ESTIMATING INTER-OBSERVER CONSISTENCY IN EYE FIXATIONS

Matthias Kümmeler¹ (matthias.kuemmerer@bethgelab.org), Matthias Bethge¹; ¹Tübingen AI Center, University of Tübingen

How consistently do people look at the same things in a given image? The answer matters fundamentally for understanding visual attention, but standard methods systematically underestimate human consistency. We propose a population coding approach: rather than estimating fixation distributions from spatial data alone, we combine information across multiple sources—spatial, semantic, and contextual—to better capture the true structure of where observers look. The problem arises from data limitations: typical eye-tracking studies collect 10-20 fixations per image. Standard kernel density estimation worked well until recently. However the newest generation of image-computable saliency models outperform these methods, sometimes substantially. This results in unreliable identification of images where computational models fail to predict human behavior. Our approach combines three innovations. First, we employ adaptive bandwidth kernel density estimation (Abramson's method), which automatically adjusts spatial precision to capture both tight fixation clusters and broader distributions. Second, we incorporate semantic scene information through a mixture model, leveraging the fact that similar image features attract fixations across different images. Third, we optimize estimation parameters on a per-image basis. Results across multiple free-viewing datasets (MIT1003, CAT2000, COCO-Freeview) demonstrate substantially improved inter-observer consistency estimates, 10% on average and more than 50% on individual images. This demonstrates that saliency models still have potential to improve, and enables us to reliably identify where current saliency models fail to capture human viewing behavior. Such cases provide the clearest targets for investigating missing mechanisms in

computational models of attention. We demonstrate extensibility by incorporating task context (free viewing vs. visual search) as an additional mixture component. In principle, task-specific fixation estimates should be more precise than task-averaged ones. However, conditioning on task further reduces the available data per image. By combining task-specific and task-general KDE components, we leverage the precision of task-specific information while benefiting from the statistical power of pooling across tasks.

This work was supported by the German Research Foundation (DFG): SFB 1233, Robust Vision: Inference Principles and Neural Mechanisms, project number: 276693517.

36.417 A GENERATIVE CONTROL-THEORETIC MODEL OF PURSUIT-SACCADE COORDINATION: BAYESIAN INFERENCE AND CLINICAL RELEVANCE

Todd Hudson^{1,2}, Mahya Beheshti^{1,2}, JR Rizzo^{1,2}; ¹NYU Grossman School of Medicine, ²Tandon NYU School of Engineering

From a control-theoretic perspective, smooth pursuit is an inherently unstable tracking problem: even small temporal errors accumulate until the continuous controller fails and generates a discrete corrective reset. Adaptive catch-up and backtracking saccades are therefore computational consequences of imperfect tracking. In healthy observers, they reveal visuomotor prediction, delay compensation, and uncertainty; in clinical populations they become exaggerated, delayed, or misdirected, providing a functional biomarker of impaired controller dynamics, processing, and/or models. Quantifying these events, however, requires solving a difficult inverse problem: the observed data trace reflects the sum of outputs from the continuous pursuit controller and saccadic corrective mechanism - a combination that is not identifiable without a generative model. We develop a generative Bayesian framework for joint inference of latent pursuit dynamics and saccadic correction. Pursuit is represented as a 3-state linear dynamical system (position, velocity, acceleration), capturing the smooth but uncertain internal trajectory. A heteroskedastic observation model increases variance during saccadic intervals, enabling the forward estimator to preserve the continuity of latent pursuit while avoiding contamination from large saccadic deviations. Saccades arise from a mechanistic phase-error policy: instantaneous pursuit-target phase discrepancy drives a probabilistic hazard for generating adaptive saccades, yielding a complete generative account of pursuit failure. We conducted a stress test of the model at the frequency of our control and patient data (0.25 Hz), with simulations spanning observation noise from near-zero to a significant proportion of pursuit amplitude along with a range of positive/negative imposed phase errors. For each condition, we computed timing error distributions, timing SD, amplitude gain, and detection fidelity. Results reveal large identifiability regions where Bayesian inference reliably reconstructs pursuit and recovers saccade timing. We will present these computational results alongside human pursuit data, demonstrating how this generative-inference approach yields

calibration-robust, model-based biomarkers of visuomotor control in both healthy control and clinical populations.

36.418 CONJUNCTIVE TUNING AND COMPENSATORY CORTICAL DISTORTIONS SUPPORT A LOW-DIMENSIONAL MECHANISM OF PREDICTIVE VISUAL REMAPPING

Xize Xu¹, Sachira Denagamage², Anirvan Nandy¹, Monika Jadi¹; ¹Yale University, ²Columbia University

Perceptual continuity across saccades depends on pre-saccadic receptive fields (RF) remapping of visual neurons, a process driven by self-motion signals – corollary discharge (CD) – from eye-movement planning areas. Despite prior computational efforts, key gaps remain in our understanding of the underlying mechanisms: (1) What is the role of pre-saccadic suppression (PS)—the temporal reduction in neuronal firing rates during saccade planning—in remapping? (2) Do neurons remap independently or in a coordinated manner consistent with low-dimensional population dynamics? (3) Most importantly, visual areas that perform RF remapping exhibit highly retinotopic organization, characterized by eccentricity-dependent cortical magnification (CM) and contracted representation of polar angles from the visual field. How do these non-uniform geometric characteristics impact the accuracy of remapping? In this study, we present a network mechanism whereby CD inputs guide RF forward remapping in a recurrent population model of 2-D cortical space, represented by neurons conjunctively selective to planned saccade direction and visual field location. We reveal a regulatory role of PS to counteract the adverse effects of saccade direction selectivity during saccade planning. By yoking the responses of neurons through structured recurrent connectivity, our model predicts conserved cell-cell RF relationships across remapping. We show that this prediction and the assumption of saccade direction selectivity are supported by simultaneous population recordings in area V2 from macaques performing a cued saccade task. Cortical transformations such as eccentricity-dependent CM and polar angle contraction systematically distort remapping in our model, which is again confirmed by the empirical data. Finally, we show that remapping distortion due to CM is mitigated by distortion due to polar angle contraction in our model, a phenomenon explained by the counteracting nature of distortion from the visual field to cortical representation by each factor.

Eye Movements: Pupillometry

SUNDAY, MAY 17, 2:45 – 6:45 PM, PAVILION

36.419 ASYMMETRIES IN THE SEMANTIC PUPILLARY LIGHT RESPONSE TO BRIGHT AND DARK WORDS

Olessia Jouravlev¹, Veronica Chiarelli¹, Dalia Serik¹, Tamaya Murray¹; ¹Carleton University

Introduction: The semantic Pupillary Light Response (sPLR)—pupil constriction to semantically bright words and dilation to dark words—has been proposed as evidence for embodied language processing, reflecting activation of visual sensory representations during word comprehension. Prior work shows robust sPLR effects (Mathôt et al., 2017), yet the contribution of individual differences in visual mental imagery and language background remains unclear. We investigated whether (1) semantic processing of visually bright and dark words elicits sPLR, (2) imagery vividness modulates this effect, and (3) native versus non-native English status influences sPLR strength during English word reading task. **Methods:** Seventy participants completed a visual go/no-go categorization task while pupil diameter was recorded with an EyeLink 1000. Stimuli included semantically bright, dark, matched control words, and animal names (go-targets). Brightness/darkness stimuli were matched to controls on lexical frequency, arousal, valence, and visual intensity. Pupil dilation ratio was computed as proportional change from a pre-stimulus baseline. After the task, participants completed the Vividness of Visual Imagery Questionnaire (VVIQ-2). **Results:** There was a significant main effect of word condition. Pupil constriction was greater for semantically bright words ($M=99.23$) relative to dark ($M=100.06$) and control words ($M=100.21$), which did not differ. No interaction was observed between condition and language status or VVIQ-2. However, a follow-up analysis including only bright versus control conditions showed a significant interaction between condition and imagery scores, suggesting imagery ability selectively modulates constriction to bright words. **Conclusion:** These findings partially replicate the sPLR for semantic brightness but not darkness, providing qualified support for embodied accounts. Results further indicate that individual differences in visual imagery may influence sPLR, whereas language status did not modulate the effect in this sample.

36.420 INVESTIGATING PUPIL LIGHT RESPONSE UNDER INTEROCULAR SUPPRESSION

Hsin-I Liao¹, Yung-Hao Yang²; ¹NTT Communication Science Laboratories, NTT, Inc., ²Graduate School of Informatics, Kyoto University

Pupil light response (PLR) is traditionally considered a reflexive function of adjusting retinal luminance. However, recent studies indicate that this low-level response also reflects the subjective conscious perception or attended brightness. This shift means the question of whether the pupil responds to brightness when the stimulus is invisible is no longer straightforward, but rather perplexing. The current study addressed the issue by presenting visual stimuli with different luminance levels under interocular suppression. In Experiment 1, human participants discriminated, or guessed, the luminance (dark or bright) or orientation (clockwise or counterclockwise) of an invisible Gabor patch. Experiment 2 involved presenting two invisible Gabor patches, one black and one white, side by side, while attention was directed to one location using a rapid serial visual presentation (RSVP) task. Experiment 3 featured the

simultaneous presentation of one visible and one invisible Gabor patch, with participants judging whether the two patches had the same or different luminance. We adopted both subjective visibility and objective judgment criteria to exclude ineligible trials and participants who showed potential awareness of the suppressed stimuli. The results showed a reliable, but attenuated, PLR to invisible brightness compared to the visible control. Similar amounts of PLR were found regardless of whether the participant focused on luminance or orientation. Crucially, paying attention to the invisible stimulus's location did not induce PLR, unlike in the visible condition. Finally, when there was competition between the invisible and visible stimuli, invisible brightness affected PLR only when the visible stimulus was dark but not when it was bright. The overall results suggest that PLR is modulated by visual awareness. While PLR to invisible brightness remains intact and resistant to attention modulation, the response varies depending on the visible brightness. This highlights a dissociation between the mechanisms underlying awareness-sensitive and attention-modulated components of the PLR.

36.421 THE PUPIL AND THE MIND'S EYE(S): DIVERSE PROFILES OF REACTIVITY TO ENDOGENOUS BRIGHTNESS

Michael McPhee¹ (mcphee.mi@northeastern.edu), Morgan McCarty¹, Seth Schallies¹, Michael Young^{2,3}, Jorge Morales¹; ¹Northeastern University, ²Massachusetts General Hospital, ³Harvard Medical School

The pupillary response is driven by endogenous signals of brightness in addition to direct light. The pupil's diameter adjusts in response to the brightness of imagined objects, the semantic brightness evoked by words, and the illusory exaggerated brightness of images depicting stereotypically bright objects like the sun. While these effects have been individually demonstrated in isolated experiments, their reliability remains uncertain and the relationship between them is unknown. Across four within-subjects experiments, we systematically investigated these effects using a head-mounted display equipped with an eye tracker, which provided a tightly controlled lighting environment for pupil size recording. Subjects imagined light or dark shapes (Exp. 1), read (Exp. 2) and listened (Exp. 3) to words evoking lightness (e.g. "flame") or darkness (e.g. "abyss"), and examined a curated set of landscape photographs containing a clearly visible sun (Exp. 4). Words were categorized as light or dark based on ratings we collected beforehand, and the words' valence, arousal, and length were matched across categories. Each sun image was presented upright and upside down across trials, as perspective inversion disrupts construal of the sun as exceptionally bright. Critically, our design allows us to compare the effects of these different endogenous brightness cues within each subject. While we replicated previously reported effects of illusory (Exp. 4) and semantic (Exps. 2 & 3) brightness on pupil size, we did not find a significant pupillary response to imaginary light (Exp. 1), in contrast with previous findings. Within individuals, however, we found that the magnitudes of semantic and

imaginary brightness effects on pupil size were correlated – but that each was unrelated to the magnitude of the illusory brightness effect. Together, our findings suggest that pupillary reactivity to endogenous brightness signals may be governed by multiple distinct mechanisms rather than a single system sensitive to high-order sensory representations.

36.422 SACCADIC REMAPPING IN DYNAMIC ROAD SCENES

Ido Zivli¹, Benjamin Wolfe¹; ¹University of Toronto

Saccadic remapping enables stable perception of the world by providing precise information about the object targeted by the impending saccade prior to the start of the eye movement. Work with static stimuli (Wolfe & Whitney, 2015) demonstrated that, compared to covert attention, remapping provides more object-level information. Here, we examined whether remapping helped drivers detect and localize immediate hazards in dynamic road video. Eleven licensed drivers participated in a dashcam video-based hazard localization paradigm, initially fixating 13° below video centre (approximating a dashboard-mounted smartphone) and making a saccade towards the road scene on receipt of an auditory cue. Videos were shown peripherally for 1-3s and were removed from the screen either when the cue was given (covert attention), on saccade initiation (saccadic remapping) or 100 ms after the saccade was completed (first fixation). After removal, the video was replaced with a rectangular frame, and participants used a cursor to report where the hazard was. Click errors for each trial were calculated based on the smallest Euclidean distance from click position to the annotated hazard area. For each condition and participant, we computed the mean error, expressed as a z-score relative to a permuted null (i.e., errors produced by random clicking). We observed a significant main effect of condition ($p < .001$), with the lowest accuracy in the covert attention condition (mean $z = -3.257$, $SD = 0.98$), followed by the saccadic remapping condition (mean $z = -3.85$, $SD = 1.19$). As expected, accuracy was highest in the first fixation condition (mean $z = -5.375$, $SD = 0.84$). Our results demonstrate that saccadic remapping contributes to perceptual stability even in dynamic, changing scenes, and may be particularly useful for distracted drivers as they look back to the road.

This work was supported by an NSERC Discovery Grant to BW

36.423 PERSONALIZED EYE TRACKING REVEALS STATE-DEPENDENCE AND VIEW INVARIANCE IN FREELY GAZING REPTILES

Mark Shein-Idelson^{1,2} (sheinmark@tauex.tau.ac.il), Nimrod Leberstein^{1,2}, Milan Becker¹; ¹School of Neurobiology, Biochemistry and Biophysics, Tel-Aviv University, Tel-Aviv, Israel, ²Sagol School of Neuroscience, Tel-Aviv University, Tel-Aviv, Israel

Around 320 million years ago, stem amniotes emerged as the first vertebrates fully adapted to terrestrial life, initiating the evolutionary lineage that gave rise to mammals, reptiles, and birds. This transition

to land profoundly expanded visual opportunities and was accompanied by major innovations in both the visual apparatus and the neural circuits controlling it. These innovations included the evolution of a soft oval lens and ciliary eye muscles enabling an enhanced range and accuracy of accommodation, a steep-walled fovea combined with a clearer cornea for greater visual acuity, and a specialized craniovertebral complex permitting flexible head movements and gaze shifts. These changes were paralleled by a marked expansion of visual regions in the forebrain and their innervation by thalamic nuclei. To investigate the evolution of visual processing, we focus on reptiles—the sister clade of mammals. We developed a flexible, open-source framework for 3D printing customized eye-tracking head stages for recording eye movements in small, freely viewing animals, and validated it in mice, turtles, and lizards. Using this system, we show that reptiles exhibit a rich oculomotor repertoire characterized by conjugate and partially independent monocular saccades, amplitudes that scale with velocity, and flexible coordination with head movements. Long-term recordings revealed a biphasic visual acquisition pattern alternating between active scanning epochs and quiescent intervals lasting tens of seconds with limited eye motion. Pupil dynamics displayed an inverse relationship to mammalian arousal, with constriction during active states and dilation during quiet states. Furthermore, by combining eye movement recordings with cortical electrophysiology during oddball paradigms, we observed enhanced responses to deviant stimuli that were invariant to viewing angle. Together, these findings establish reptiles as a valuable model for studying the evolution of visual acquisition in vertebrates and illuminate early neural processing schemes in the amniote brain.

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement no. 949838; the Israel Science Foundation (ISF) grant no. 1597/25; and the Minerva fellowship of the Minerva Stiftung Gesellschaft fuer die Forschung mbH

Visual Search: Features, scenes, real-world stimuli

SUNDAY, MAY 17, 2:45 – 6:45 PM, PAVILION

36.424 FEATURAL CONTINGENCIES OVERRIDE SPATIAL FACTORS IN CONTEXTUAL CUEING EFFECT

Ananya Mishra¹, Tony Thomas¹; ¹Indian Institute of Technology Roorkee

INTRODUCTION Contextual cueing involves implicit learning of the inter-spatial relationships among search items, enhancing search efficiency. The effect is considered pre-attentive and argued to be rapid, automatic and effortless. The present study investigates the robustness of the contextual cueing effect under conditions of featural level disruptions. **METHOD** 30 participants completed a letter search

task that had 10 search configurations repeated (unbeknownst to participants), and each configuration was presented 30 times in random order (Training Phase). Each letter was in white font and size ($0.57^\circ \times 0.47^\circ$). The Testing phase had the configurations used in the Training Phase presented again. However, items appeared in a different colour (red and green) or size (20pt and 50pt; $0.38^\circ \times 0.28^\circ$ and $0.95^\circ \times 0.66^\circ$, respectively). The featural level changes were not expected to affect retrieval of the learnt invariant spatial relationship of the search items, provided the implicit learning was robust. **RESULTS** Training phase demonstrated robust learning $F(5,145)=21.90$, $P<.001$, $\eta^2=.43$. In testing phase, a significant RT cost (vis-à-vis repeated (Old) configuration) was obtained: when items appeared randomly in green and red (Rand_Color: $t(29)=-2.65$, $p=.013$), in two different sizes (Rand_Size: $t(29)=-6.98$, $p<.001$), perceptually grouped in size (Percept_Size: $t(29)=-6.38$, $p<.001$); suggesting disruptive effects of featural changes on Contextual Cueing effect. **CONCLUSION** Extraction of regularities in contextual cueing effect seems to be dependent on bottom-up factors, and not a generic phenomenon as suggested previously. Featural-level violations (for instance, colour and size in the present case) stand out perceptually and seem to override the advantage offered by the implicitly learnt patterned spatial information. Features get perceptual prioritization earlier in time than the advantages offered by the contextual cueing effect.

36.425 THE EFFECTS OF LIGHTNESS DIFFERENCES AND LUMINANCE DIFFERENCES ON VISUAL SEARCH EFFICIENCY

Howard Jia He Tan¹, Simona Buetti¹, Cathleen Moore², Alejandro Lleras¹; ¹University of Illinois Urbana Champaign, ²University of Iowa

According to Reverse Hierarchy Theory, processes like visual search first operate on the end-products of visual processing. One such end-product of the visual processing system is perceptual constancy. Perceptual constancies provide us with invariant representations of objects out in the world. In the current study, we investigate the level of visual representation used in visual search with the phenomenon of lightness constancy and phenomenal scission. Here, we used displays that invoked lightness constancy illusions and controlled for local luminance differences between search stimuli and their backgrounds. We hypothesized that search processes would use lightness differences between items ("post-constancy" representations), when they are informative, to guide search. Subjects searched for the perceptually brightest target semicircle among other semicircles. We controlled luminance differences ("pre-constancy" representations) and defined the target by a uniquely bright luminance (easy) or luminance-matched with that of distractors in a different region of the display (hard). There were two background conditions, one induced a perceived filter (Adelson snake illusion) and the other a flat surface. Set size was manipulated (Target only, 4, 8, 16) and distractors were distributed equally across the high- and low-luminance regions. RTs increased logarithmically as a function of set size, indicating that participants had performed the search in parallel across the display.

When perceived lightnesses were equal, search was efficient and no significant differences were found between background conditions. More importantly, when targets were luminance-matched with distractors in the surface condition, search became significantly harder. However, and somewhat surprisingly, in the filter condition, search was just as easy when the target was luminance-matched as when it was a unique luminance, even when the luminance information available should have resulted in a more difficult search. These results suggest that early visual information is only used when necessary and are in line with the Reverse Hierarchy Theory.

36.426 THE GUIDANCE OF ATTENTION BY STATISTICAL CONTINGENCIES WHEN MULTIPLE GUIDANCE MECHANISMS COMPETE

Mark W. Becker¹ (becker54@msu.edu), Morgan Dodd¹, Tristan Janisse¹; ¹Michigan State University

Research suggests that attention can be biased toward frequent target locations or features through implicit statistical learning. However, most prior work has examined only a single biasing contingency. We investigated whether individuals can simultaneously learn and implement two distinct contingencies: one based on location and another on color. To do so participants searched a clock-face array of 16 Landolt Cs for the C which had a break on the left or right, among Cs with breaks on the top and bottom. Each quadrant had four stimuli - one of each of four colors. The target appeared in the dominant quadrant 50% of the time, and in each of the other quadrants 16.67% of the time. Similarly, targets appeared in the dominant color 50% of the time, and in each of the other colors 16.67% of the time. Results indicate that both contingencies were learned implicitly and exerted additive effects on attentional allocation. Next, we examined whether shifting the bias for one feature to the volitional system via endogenous cues affected the implicit learning and implementation of statistical contingencies for the other feature. We found that moving one feature's bias to the top-down volitional system did not block implicit learning of contingencies for the other feature. However, the impact of that learning on attentional allocation depended on the validity of the volitional cues: statistical contingencies biased attention when volitional cues were neutral or valid, but not when invalid. This pattern suggests that the implicit and volitional systems operate hierarchically, with the top-down volitional system gating the items over which implicit statistical learning can exert its influence.

36.427 CUE INFORMATIVENESS MODULATES ATTENTION BENEFITS FOR BOTH TARGET AND DISTRACTOR CUEING DURING VISUAL SEARCH

Douglas Addleman¹, Viola Störmer²; ¹Gonzaga University, ²Dartmouth College

During visual search, information about upcoming targets is highly beneficial: Knowing that your next target is likely red helps you find it much faster. Information about distractors is also helpful, but less so, even when compared to similarly informative target cues. How do

these effects depend on how reliably cues predict upcoming features? We addressed this in four experiments (N=50 each) manipulating the format (target vs. distractor) and informativeness (75% vs. 92%) of cues to the color of upcoming search items. In all experiments, participants searched for a target C (left or right gap) among seven distractor Cs (top or bottom gap). Each trial contained eight items (four in each of two colors) and was preceded by a colored cue that predicted the upcoming target color (Exp. 1, 75% informativeness; Exp. 2, 92% informativeness) or non-target color (Exp. 3, 75% informativeness; Exp. 4, 92% informativeness). The specific cue color changed every 108 trials, with four different colors used across the experiment. Participants knew the precise informativeness and format of their cues. In all experiments, accurate cues led to faster responses than inaccurate ones ($p < .01$). However, cue format and informativeness strongly influenced cueing benefits. When comparing validly cued trials to invalid ones, people were 236ms faster for 75% valid target cues and 385ms faster for 92% valid target cues, but only 48ms faster for 75% valid distractor cues and 162ms faster for 92% valid distractor cues. All across-experiment cueing benefits were significantly different ($p < .01$). These results show that the greater effectiveness of target compared to distractor cueing holds at multiple levels of cue informativeness. Perhaps more importantly, they show that increasing informativeness of cues makes cueing effects stronger, even though participants have strong incentives to use cues to benefit performance even at moderate levels of informativeness.

This research was supported in part by NSF Grant BCS-1850738 to V.S..

36.428 ASSESSING THE DETECTION OF PELVIC FRACTURES IN IMMERSIVE VIRTUAL REALITY BY NOVICES

Alex Detrich¹ (andetrich@gmail.com), Katrina Visintainer², K. Hudson Guttman¹, Dongyu Zhang², Jason Orlosky², Megan Mills¹, Jeanine Stefanucci¹; ¹University of Utah, ²Augusta University

Radiologic interpretation error rates are estimated to occur in 3–5% of cases and can lead to misdiagnoses. In radiology practice, complex three-dimensional (3D) anatomy is typically viewed in two dimensions (2D). Translating what is seen on 2D images to the 3D anatomy may require mental effort and draw on limited cognitive resources similar to other spatial reasoning tasks. These factors may contribute to errors in perception and interpretation. Immersive virtual reality (VR) may reduce the mental load by providing wholistic 3D representations of the anatomy. This study compared novices' performance across three VR methods that portrayed anatomical structures in 2D and 3D for detecting acetabular pelvic fractures. Participants (n=55) completed 15 trials (five per condition) in VR. They annotated fractures of the pelvis from computed tomography (CT) and rated confidence after each trial. Three visualization conditions were tested: 2D volumetric (CT scans on the wall), 3D (static rendered pelvis), and combined 2D volumetric + 3D. Most participants had prior VR experience (71.4%) and little CT experience (94.6%). Repeated-measures ANOVAs

examined confidence, reaction time, and accuracy across conditions. Trial-by-trial confidence did not differ. Response times differed significantly, $F(2,108) = 20.90$, $p < .001$, $\eta^2G = 0.10$. Bonferroni-corrected pairwise comparisons showed no difference between 2D volumetric and 2D volumetric + 3D, $t(54) = 0.78$, $p = 0.438$, but 3D was faster than both 2D volumetric, $t(54) = 5.79$, $p < .001$, and 2D volumetric + 3D, $t(54) = 5.38$, $p < .001$ (3D: $M = 90.0$, $SD = 45.5$; volumetric: $M = 132.0$, $SD = 67.8$; volumetric + 3D: $M = 126.0$, $SD = 54.9$). Accuracy did not differ among conditions. Overall, 3D viewing enabled faster detection of pelvic fractures compared to 2D or combined 2D+3D viewing conditions in novices. Future work will extend this research to participants with CT viewing experience.

36.429 LONG-TERM MEMORY AND VISUAL WORKING MEMORY TARGETS EXHIBIT DISTINCT SENSITIVITY AND CRITERION IN HIGH AND LOW TARGET PREVALENCE SEARCH

Younha Collins¹ (yo547698@ucf.edu), Joseph Schmidt¹; ¹University of Central Florida

Individuals fail to recognize low prevalence (LP) targets, even when fixated directly (Hout et al., 2015), resulting in increased misses and a conservative criterion shift. Likewise, high target prevalence (HP) increases false alarms and shifts criterion liberally (Wolfe & VanWert, 2010). Decades of target-prevalence research focused on long-term-memory (LTM) targets, despite clear differences between LTM and visual-working-memory (VWM) systems (Goldstein & Beck, 2018). The current study examined whether the prevalence effect interacted with memory system by asking participants to search for VWM or LTM targets at low (10%) and high (90%) prevalences. Before the LTM search, participants memorized six targets and their categorical names. At the start of each search, the eye tracker detected central fixation (200ms) before two cues appeared (1s; text cues for LTM and images of real-world objects for VWM), followed by a 1s delay before an 18-item search array appeared (half from each cued category). Replicating prior work, LP produced higher misses than HP, and HP produced higher false alarms than LP. Additionally, the miss rates did not significantly differ between memory systems, suggesting similar low prevalence effects. However, false alarms were more frequent for VWM than LTM targets, suggesting VWM exacerbates the high prevalence effect. Signal detection analyses demonstrate that HP and VWM produced more liberal criteria than LP and LTM, respectively. Sensitivity was positively associated with prevalence for LTM but negatively associated with prevalence for VWM. Criterion correlated with correct-rejection RTs for LTM (replicating prior work), but not for VWM. Sensitivity correlated with correct-rejection RTs in LP but not HP in both memory systems. Taken together, this suggests that each memory system produces independent sensitivity and criterion thresholds and VWM exacerbates the high prevalence effect, demonstrating that there is not a singular prevalence effect, but rather one that interacts with target memory location.

36.430 WHY SCRAMBLING SOMETIMES FAILS: TARGET SPECIFICATION MODULATES CONTEXT EFFECTS IN VISUAL SEARCH

Antje Nuthmann¹, Franz Faul¹; ¹Kiel University, Germany

Scene context is known to facilitate visual search in images of real-world scenes by constraining where a target object is likely to appear. These contextual benefits are commonly investigated by comparing search performance for objects located in typical versus atypical positions, or in intact versus scrambled scenes. Unexpectedly, Foulsham, Alan, and Kingstone (2011) reported no effect of scene scrambling on search time when scenes were subdivided into a 4×4 grid. This result seems to challenge the widespread assumption that disrupting global scene structure reliably impairs search. We suspected this null effect arose from using whole grid cells as target cues, a method that may bypass the need for object identification and render contextual information unnecessary. To test this, we conducted an extended replication where targets were cued either by naming the object (word cue) or by the content of the grid cell containing the object (image cue). These cue types were crossed with scene scrambling (normal vs. scrambled) and object position typicality (typical vs. atypical), allowing us to investigate how target specification interacts with contextual disruption. Targets were always present, and we collected behavioral and eye-movement data. The results support our hypothesis: Scene scrambling and atypical object position both produced substantial increases in search time and error rate in the word-cue condition, whereas these effects were minimal in the image-cue condition, where the cue effectively removed the need for object identification. With word cues, atypical object positions were particularly disruptive, suggesting that contextual expectations play a crucial role in locating objects. In this case, the lack of context sometimes prevented target discovery altogether. These findings demonstrate that scene context supports visual search only when semantic information about the target is available, offering an explanation for prior discrepancies in the literature regarding the effects of scrambling.

36.431 VISUAL SEARCH FOR REAL OBJECTS: HOW SPATIAL CONSISTENCY FACILITATES PERFORMANCE.

M Pilar Aivar¹ (mariapilar.aivar@uam.es), Laura Cepero¹, Enrique Heredia-Aguado², Alejandro Rujano³, Rocio Asperilla Ruiz¹, Victoria Plaza¹; ¹Universidad Autónoma de Madrid, ²Universidad Miguel Hernández de Elche, ³Universidad Nacional de Educación a Distancia

In everyday life we continuously search for specific objects in our environment. However, much of what we know about visual search comes from studies performed using computer-based tasks. To test whether our knowledge regarding the characteristics of visual search applies to natural situations we performed a series of experiments in which participants searched for real 3D objects. As targets and distractors, we used a collection of plastic figures of different shapes and colors that were placed on a rectangular frame. The task we employed was repeated visual search: the display was kept constant

over trials, but a different piece was presented as target in each trial. Participants searched for 10 different pieces a total of six times. They had to point to the correct piece in each trial, and we recorded their eye movements during the complete task. Results showed that target repetition led to a decrease in the time needed to find the different pieces, replicating the effects that have been seen on computer versions of this task. This suggests that participants rapidly acquire and rely on stable spatial information to guide their search. After 3 blocks of trials the complete frame was inverted, so that all pieces changed location. This disruption showed that learning was location-specific: when spatial consistency was lost, performance dropped, indicating that participants' strategies depended on memory for object locations rather than purely on visual features. The recovery of performance once the original configuration was used again further supports the idea that spatial knowledge is robust and can guide visual search efficiently. Overall, these findings extend classic repeated-search effects to real-world settings, showing that natural search behavior is strongly shaped by learned spatial regularities.

Research supported by grant PID2024-155632NB-I00

36.432 STUDY OF THE EFFECTS OF DEPTH, SET SIZE AND ECCENTRICITY ON REPEATED VISUAL SEARCH IN REAL ENVIRONMENTS

Federico Sorcini¹, Jun-Ming Yu¹, Simona Buetti¹; ¹University of Illinois Urbana-Champaign

Repeatedly searching for a target through the same layout of objects should yield some improvement in one's ability to find a target. Previous work has suggested that this intuition is incorrect and that when performing a repeated visual search, participants prefer to use vision rather than memory to guide their search, even in complex, realistic scenes. However, much of the work that has been done has involved presenting the search stimuli at relatively small eccentricities on a 2-D monitor, both of which ensure that all stimuli are easily accessible by the visual system. Here, we investigate the extent to which memory guides repeated search when visual information is not always easily accessible: in a real 3D space. Participants were eye-tracked as they repeatedly search through sets of real objects placed on a desk in front of them. Four, eight and twelve objects were placed in a wide range of eccentricities (up to 45 degrees to the left and right of the participant), either on the same depth plane as an initial fixation point (horopter fixation condition) or on a different plane (near fixation condition). Participants were instructed to direct their gaze to the target object as quickly as possible. We hypothesized that observers would use memory to a greater extent for objects located at the farthest eccentricities, and perhaps more so in the near fixation condition because objects would appear initially blurrier than in the horopter fixation condition. Results showed that initial fixation depth, set size, and eccentricity impacted saccadic initiation times and time to foveate the target. Learning was increasingly evident as set size increased and at farther eccentricities, but performance leveled in all conditions. Surprisingly, participants were only about 70% correct at recalling the

location of the near and farthest targets, with worse memory for targets at intermediate eccentricities.

36.433 HOW CONSTRUCTED ENVIRONMENTS SHAPE VISUAL SEARCH PERFORMANCE

Yoshiyuki Ueda¹, Kikuno Iizuka¹, Ryusuke Nakai¹; ¹Kyoto University

Visual search studies have traditionally investigated how the cognitive system operates within a given environment, implicitly assuming that the observers play a passive role with respect to their surroundings. However, in real-world contexts, environments are not simply given; rather, they are often actively constructed by people. In this study, we investigate the relationships between environment construction and subsequent visual search processes within a single experimental framework. To address this issue, we developed a constructing environment-visual search task. In the constructing environment task, participants arranged twelve visual objects in ways they believed would make searching easier. Under the same instruction, each participant created four types of search displays. Afterward, participants performed a visual search task in which the target appeared in displays constructed by (i) themselves (self-constructed condition), (ii) another person (other-constructed condition), or (iii) randomly (random condition). Displays in the self- and other-constructed conditions were presented repeatedly to examine learning efficiency (i.e., the contextual cueing effect). The results showed that reaction times decreased more strongly in both the self-constructed and other-constructed conditions than in the random condition, although no difference was observed between the self-constructed and other-constructed conditions. Moreover, accuracy was significantly higher in the other-constructed condition than in the self-constructed and random conditions. These results suggest that constructing environments may interfere with efficient visual processing or introduce memory-related confusion, whereas environments constructed by others may facilitate more straightforward processing. Preliminary findings from a brain-activation study further support that self-constructed displays elicit broader activation, particularly occipital regions.

JSPS Kakenhi #25K00900

36.434 A MICROGENETIC ANALYSIS OF VISUAL SEARCH REVEALS A DIFFERENCE FOR TARGETS WITH KNOWN VERSUS UNKNOWN IDENTITIES

Patrick Wu¹, Howard Egeth¹, Jonathan Flombaum¹; ¹Johns Hopkins University

Consider basic variations of visual search: (1) pop-out, where a target is defined by a unique feature, (2) directed conjunction search, where a target defined by a combination of a color and a shape is shown prior to the trial, and (3) undirected conjunction search, where the displays are identical to a directed search, but no information is given before the trial. These likely demand different computations to perform. But it can be difficult to identify how and why by looking only at reaction times

and error rates. We conducted six experiments, modifying typical search by limiting stimulus exposure. By requiring participants to click as close as possible on a blank screen to the position of the target we could compute pixel distance from click to true location as a function of exposure. This approach produces curves that describe the microgenesis of a trial: the rate at which knowledge accumulates. Two key findings emerged. First, directed search curves are indistinguishable from pop-out, suggesting that knowledge of the target allows for suppression of visual features and detection of salient signals despite limited exposure. Undirected search curves were distinct from the others and did not show similar properties; knowledge accumulation (distance error as a function of stimulus exposure duration) in undirected search fit a parabolic model as compared to a logarithmic model in the others. Second, increasing the number of nontargets greatly impaired undirected search with hardly a measurable effect on directed search. After increasing the nontargets from 7 to 11, undirected search performance did not reach an asymptote even at 3 seconds of stimulus exposure. Undirected search appears to rely on processes that are different in kind from those that support pop-out and directed search.

Johns Hopkins University Provost's Undergraduate Research Award

36.435 FROM BROAD GOALS TO SPECIFIC FEATURES: HIERARCHICAL TARGET GUIDANCE IN NATURALISTIC VISUAL SEARCH

Shiyang (Harry) Ren^{1,2} (syhren@ucdavis.edu), Qianqian Wan², Shea Duarte^{1,2}, Joy J. Geng^{1,2}; ¹Department of Psychology, University of California, Davis, ²Center for Mind and Brain, University of California, Davis

Traditional visual search paradigms involve a target in a static display. However, natural search tasks typically involve a hierarchical process in which the target has to be navigated to before search and localization can occur. Here, we investigated how the information guiding behavior changes over time in a naturalistic search task. Participants (N = 80) searched for 10 objects (e.g., a small yellow lamp) sequentially within a 4,216 sq. m. virtual reality furniture store, navigating via teleportation. We found that the information used to reduce the functional search space occurred only periodically at "decision stops", defined by high degrees of head- and eye-movements indicative of information seeking (i.e., searching), which alternated with action-based movements. These navigational decision stops occurred at path junctures and were characterized by looking at, and rejecting, alternative paths. The path chosen was looked at proportionally longer just before movement initiation ($\beta = 0.04$, $p < 0.001$). The number of navigational decision stops decreased over time, indicating spatial learning ($\beta = -0.22$, $p < 0.001$). Once in the target room, decision stops were characterized by looking at clusters of target-similar objects. The highest proportion of looking times (M = 38.5%) and likelihood of approaching objects (M = 78.9%) were for clusters sharing one target feature (e.g., shape). After approaching the cluster, participants looked for the other target feature (e.g., color) to

localize the target. The number of search-based decision stops did not change over time because each target was new ($\beta = 0.00$, $p = 0.91$). We conclude that in a naturalistic task, searching for information occurs only periodically and alternates with action-based priorities. Moreover, the target template used at each stop evolves in accordance with hierarchical goals to reduce the functional search space, with behavior-guiding information becoming more specific to the target over time.

James S. McDonnell Foundation, NSF BCS 2446371

36.436 CURIOSITY IGNORED THE CAT: HOW EXPLORATION PROMOTES CONTEXT-DEPENDENT DISTRACTOR SUPPRESSION IN VISUAL SEARCH THROUGH REAL-WORLD SCENES

Justin Frandsen¹, Brian Anderson¹; ¹Texas A&M University

Statistical regularities between objects and their locations can guide attention during visual search through scenes, even when the objects lack semantic meaning. Previous work demonstrates that target-location associations can facilitate visual search; however, it is unclear if distractor regularities can develop in the same way. In the present study, participants completed a training task where they would search for a target shape while a critical distractor shape appeared in the scene. Four target shapes were each associated with one of three possible locations within the scene (floor, counter, or wall) where they would appear 75% of the trials on which they were encountered; similarly, four critical distractor shapes were also each given an association with one of the three regions, where they would appear 100% of the trials on which they were encountered. Each trial, participants were cued to search for one of the four possible targets and respond to the orientation of a small "T" appearing within the target. After their response, participants were given a post-search duration where they were able to freely explore the scene. In a subsequent test phase, participants continued with the same task, but the critical distractors became the targets and appeared in all possible regions equally, and the post-search period was removed. In the training phase, participants were faster to respond and were more likely to initially fixate the target when it appeared in its associated high-probability region. In the test phase, we found a reversal of this effect where participants were slower to respond and less likely to initially fixate the target when it appeared in its previous high-probability location, suggesting that participants learned to inhibit the processing of that shape in its associated location during training. Additionally, we found that amount of exploration predicts stronger inhibitory effects.

36.437 ATTENTION WEIGHTING OF FEATURE DIMENSIONS DEPENDS ON BOTH FEATURE CONTRAST AND SUBJECTIVE SIMILARITY JUDGEMENTS

Jun-Ming Yu¹ (jy87@illinois.edu), Zoe(Jing) Xu², Simona Buetti¹; ¹University of Illinois, Urbana Champaign, ²University of Washington

When targets differ from distractors along multiple feature dimensions, the attention system can adaptively assign greater weight to the dimension providing stronger guidance signals. Earlier, we quantified this dynamic weighting process with distractor colors sampled from both directions on the color circle relative to the target. Distractor colors were twice as likely to be picked from one direction than the other, introducing potential weighting biases. Here, we selected distractor colors from only one direction. Search performance was modeled when shape had a stronger guidance signal than color (shape easier condition) and vice-versa (color easier condition), with distractors differing from the target in both color and shape. The strength of the shape and color guidance signals was quantified in single-feature-dimension search tasks. In the shape-easier condition ($N = 20$; color search slopes 190 – 318 ms/log-unit, shape search slopes 101 – 140 ms/log-unit) we found a shape weight of 1.56 and a color weight of 0.44, indicating that participants prioritized shape over color. Unexpectedly, in the color-easier condition ($N = 20$; color search slopes 108 – 176 ms/log-unit, shape search slopes 263 – 365 ms/log-unit), we found a color weight of 0.53 and a shape weight of 1.47, indicating that participants did not prioritize color over shape. We hypothesized that the lesser reliance on color arose from participants' subjective similarity judgments, perceiving target and distractor colors as overly similar and therefore choosing not to rely on color information. As a follow-up, we increased the color guidance strength (color much easier condition; $N = 21$; color slopes 39 – 78 ms/log-unit, shape slopes 101 – 140 ms/log-unit). The color weight increased to 1.00, indicating that participants shifted attention towards color when the subjective dissimilarity was sufficiently large. The results suggest that participants are not necessarily sensitive to search efficiency, rather favoring subjective similarity judgements.

Visual Search: Neural mechanisms, models, eye movements

SUNDAY, MAY 17, 2:45 – 6:45 PM, PAVILION

36.438 CORRELATIONAL APPROACH TO UNRAVEL DISTINCT MECHANISMS IN VISUAL SEARCH

Sandali *Liyanagoonawardena*¹
(sandali.liyanagoonawardena@epfl.ch), Werner Schneider², Michael Herzog¹; ¹Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, ²Neuro-Cognitive Psychology & CITEC, Bielefeld University, Germany

Performances in basic visual tasks correlate very little with each other. For example, Landolt C acuity correlates very little with vernier and bisection acuity, despite high test–retest reliability on each test. Similar results were found for studies with visual illusions and in older participants. Here, we asked whether similar results hold true for visual search. We tested 32 healthy young adults with five conjunction visual search tasks with line and shape stimuli varying in shape, color, and

orientation. In addition, we tested a letter search task (upright T among Ls). Each task was conducted under three distractor levels (2, 4, and 8 items), across two sessions, with two repetitions per session to assess test–retest reliability. Test-retest reliability was good in all but one condition, with ICCs ranging from 0.5 to 0.9. In contrast to visual acuity and illusion studies, moderate to very high correlations were observed between the different conjunction searches, varying across three of the primary dimensions: shape, color, and orientation. However, search efficiencies with these line and shape displays correlated only weakly with the letter search. We speculate that letter search is driven by integrated object representation rather than distinct target features, as in the first four tasks. Our results show how correlative studies can be used to unravel distinct mechanisms in visual search.

36.439 PROACTIVE FEATURE-SPECIFIC SUPPRESSION IN A LEARNED SUPPRESSION TASK

Matthew Tong¹, Ziyao Zhang², Nancy B. Carlisle¹; ¹Lehigh University, ²University of Texas at Austin

An important aspect of performing efficient visual search is the ability to modulate attention to ignore distractions. This attentional suppression can occur as a result of selection history. Through repeated experience with a salient distractor over time, individuals can implicitly learn to ignore it and find their target more efficiently. Research examining the underlying mechanisms of this learned suppression effect leaves some open questions. The signal suppression hypothesis suggests that learned suppression can occur proactively and specific to feature values. Alternatively, stimulus-driven accounts suggest that learned suppression occurs reactively, after an initial attentional capture and followed by a rapid disengagement of attention. In addition, second-order suppression models suggest that suppression may not be specific to individual features. To clarify these open questions, we measured the power of steady-state visually evoked potentials (SSVEPs) throughout a learned suppression task. In this task, participants had to find a target shape in a search array with a distractor color singleton appearing on half of the trials. We used SSVEPs as a neural correlate of feature-specific attention by manipulating the color of the SSVEP-eliciting stimuli. We also analyzed time periods both before and after the onset of the search array to compare proactive and reactive mechanisms. We observed a significant decrease in SSVEP power corresponding to the color of a learned distractor singleton compared to both target and neutral baseline colors in the proactive time period. In addition, this reduction in SSVEP power for the distractor color was maintained into the reactive time period. Thus, we provide support for the signal suppression hypothesis by finding neural evidence for proactive, feature-specific learned suppression. These findings highlight important neural mechanisms underlying learned distractor suppression, address open questions in the literature, and more broadly, advance our current understanding of attention during visual search.

36.440 FEATURE-BASED LANGUAGE GUIDANCE FACILITATES VISUAL SEARCH IN HUMANS AND FOVEATED VISION-LANGUAGE TRANSFORMER MODELS

Parsa Madinei¹, Sana Shehabi, Miguel P Eckstein; ¹University of California, Santa Barbara, ²University of California, Santa Barbara, ³University of California, Santa Barbara

Introduction: In everyday life, language specifying target features facilitates visual search (e.g., Wolfe & Horowitz, 2017). Here, we investigate how including features in language pre-cues improves target detection performance and guides eye movements in humans and a deep neural network-based foveated language-guided search model (FLGSM; Madinei & Eckstein, 2025). **Methods:** We generated 30 images (15 target-present, 15 target-absent) using Gemini 3 Nano Banana Pro, each containing five objects of the same category varying along two feature dimensions: one with higher peripheral visibility/discriminability (HPV-feature, e.g., color, shape, orientation) and one with lower (LPV-feature, e.g., brand, text label, subtle feature). Language pre-cues used expressions referring to the target's LPV-feature (e.g., "clock with time 5:54") or included its HPV-feature (e.g., "star-shaped clock with time 5:54"). Each human participant (N=6) completed a target search detection task (50% presence) using one of the language pre-cue types per image. FLGSM combined multi-modal transformers with foveated processing (Freeman & Simoncelli, 2011) and vision-language cross-attention heatmap fixation selection. FLGSM was evaluated on the same images presented to humans. **Results:** Human target detection improved significantly with linguistic cueing of target features visible in the periphery (Area under the ROC, AUC= 0.74 for LPV-feature vs. AUC= 0.89 for HPV-feature, $p<0.01$), and the distance of the closest fixation to the target decreased (3.0° for LPV-feature vs. 1.9° for HPV-feature, $p<0.05$). FLGSM showed similar patterns to humans: detection accuracy increased (AUC= 0.55 for LPV-feature vs AUC=0.75 for HPV-feature) and closest fixation distance decreased with HPV-features (2.4° vs. 3.1° for LPV-features). **Conclusions:** Our findings suggest that human feature-based linguistic search guidance can be accounted for by a model with visual-linguistic representations and the interaction between the visual features and foveated processing.

36.441 IF YOU KNOW IT'S THERE, DO YOU KNOW WHERE? DETECTION AND LOCALIZATION IN VISUAL SEARCH

Alex J. Hoogerbrugge¹, Jeremy M. Wolfe^{2,3}, Johan Hulleman¹; ¹University of Manchester, ²Brigham and Women's Hospital, ³Harvard Medical School

Wolfe's Guided Search 6.0 (2021) posits that covert attention is deployed serially to individual items within a Functional Visual Field (FVF). In contrast, Hulleman & Olivers (2017) proposed an item-free model, where target presence within the FVF is detected without those serial shifts of covert attention. Here, we tested these contrasting viewpoints. In two online experiments ($n=122$ & $n=120$), participants

searched for a T amongst Ls (Experiment 1: set sizes 6 and 8; Experiment 2: set size 18). Display durations were 67–333 ms. Participants reported either target presence (2-AFC), target location (m-AFC), or both. In an item-based search account, target detection implies knowledge of its location – it was specifically attended, after all. Therefore, across presentation durations, detection accuracy should roughly match localization accuracy, and both should increase in a linear fashion (e.g., one item per ~50 ms). In an item-free account, more search time allows further evidence accumulation, but this is spread across the FVF. Localization is a second step which may not be executed if insufficient time is available. Consequently, targets may be detected without their exact location being known. Across presentation durations, localization accuracy should therefore lag detection accuracy, with both increasing non-linearly. We found in both experiments that localization performance was worse than detection, especially for the shortest presentation durations. Although chance-level performance is inherently higher for 2-AFC detection than for m-AFC localization, none of our chance-correction methods completely removed the performance deficit for localization. Furthermore, no clear evidence about the (non-)linearity of performance increases was found in either experiment. These results seem to place the theoretical burden on an item-based theory to explain how observers might know that a target is present without them knowing where that target was located.

AJH and JH were supported by UKRI grant ES/X000443/1. JMW was supported by NIH-NEI: EY017001, NSF: 2146617, and NIH-NCI: CA207490.

36.443 CONTEXTUAL CUEING INFLUENCES EARLY AND LATE DECISION PROCESSES IN VISUAL SEARCH

David A. Tomshe¹ (dtomshe1@lsu.edu), Jason Scimeca¹, Melissa R. Beck¹; ¹Louisiana State University

Contextual cueing occurs when memory of the spatial context for repeated search displays improves visual search times. However, it is unclear how contextual cueing influences the cognitive processes involved in early (attentional guidance) and late (target identification) stages of visual search. To better characterize the effects on these latent cognitive processes, we applied hierarchical drift diffusion modeling (HDDM) to accuracy and response time data from three experiments. Experiments 1 and 2 consisted of a typical T among L array embedded over scene backgrounds. Both distractor locations and the repeated background scene served as context for the target location. In Experiment 3, participants searched for a symbol within aeronautical maps with varying levels of visual clutter while undergoing eye-tracking. In Experiments 1 and 2, initial results favored a model in which the decision threshold for making a response, but not the rate of accumulate evidence towards the decision, differed between repeated versus non-repeated displays. In Experiment 3, the results reveal a critical distinction between the early and late stage of visual search: contextual cuing effects were linked to changes in decision threshold during late stages and were linked to evidence accumulation

during early stages. This finding extends prior work to explore the underlying processes of contextual cuing. By combining eye-tracking and behavioral modeling, we find evidence that contextual cuing facilitates performance by influencing distinct processes engaged during early versus late stages of search.

36.444 WHEN TO DRILL AND WHEN TO SCAN: INSIGHTS FROM A FOVEATED IDEAL SEARCHER IN 3D IMAGE STACKS

Anqi Zhang¹ (anqizhang@utexas.edu), Miguel P. Eckstein¹; ¹Department of Psychological and Brain Sciences, UC Santa Barbara

Introduction: Radiologists regularly search for disease in slices acquired by 3D imaging techniques. Their navigation strategy varies (Drew et al., 2013; Aizenman et al., 2017; Ba et al., 2020). Some scroll through slices while maintaining fixation (drilling), whereas others deploy multiple eye movements within one slice before advancing (scanning). However, it remains unknown when and why either strategy is more effective than the other, and what the performance-maximizing strategies might be. **Methods:** We derived an image-computable Foveated Ideal Searcher in 3D (3D-FIS) that accounts for foveated visual processing and noise correlations across fixations and slices. We compared the 3D-FIS to model searchers with systematic drilling and scanning for a larger target visible in the periphery (mimicking a mass in breast images) and a smaller target difficult to detect in the periphery (mimicking a microcalcification). We assessed localization accuracy of human observers (N=12) for each target using drilling and scanning strategies with gaze monitoring and auto-scrolling display. On each search trial, observers gave an early localization response after the 21st fixation and a late response after the 80th. **Results:** The 3D-FIS shows emergent drilling predominance for the larger target, and alternates scanning and drilling for the smaller target. The drilling model approximates 3D-FIS accuracy for the larger target. The scanning model attains closer accuracy to the 3D-FIS than the drilling model for the smaller target. These results predict our psychophysical findings. With short search time, human drilling outperformed scanning for the larger target; with long search time, human scanning outperformed drilling for the smaller target (both $p < .0001$). Furthermore, the foveated models with recorded eye movements yield the same pattern of performance differences. **Conclusion:** We establish a new theoretical foundation for understanding optimal visual search strategies in 3D volumes, explaining radiologists' search behavior, and predicting the effectiveness of heuristic strategies.

Supported by NIH grant R01EB026427.

36.445 A COMMON LOOK-AWAY STRATEGY: EYE-MOVEMENT EVIDENCE FROM CUED AND LEARNED DISTRACTOR IGNORING

Woongjin Seo^{1,2}, George R. Mangun^{1,2}, Joy J. Geng^{1,2}; ¹University of California, Davis, ²Center for Mind and Brain, University of California, Davis

Cuing and implicit learning of likely distractor locations have both been used to study distractor suppression in visual search, but some evidence suggests that the underlying mechanisms differ. We directly compared these two methods during singleton-search mode using eye tracking to assess how suppression operates in each case. In separate studies, the salient distractor location was either pre-cued on every trial, or uncued in a statistical-learning paradigm in which the distractor appeared more often at a single high-probability location (HPL). Manual RTs were faster when distractors appeared in the expected location (cued: 1.06s vs. 1.09s; learned: 1.18s vs. 1.25s both $p < .01$). However, the proportion of first saccades to the singleton-distractor was significantly higher than non-singleton-distractors, suggesting a lack of suppression (cue: 25.9% vs. 5.5%, $p < .001$; learning: 16.9% vs. 6.0%, $p = .025$). Yet, for both cued and learned conditions, the proportion of looking to the singleton distractor was lower for the valid/HPL than the invalid/LPL conditions, (Valid: 25.9% vs. Invalid: 31.7%, $p < .05$; HPL: 16.9% vs. LPL: 38.6%, $p < 0.001$), showing reduced capture when the salient distractor appeared in the expected location, although the effect was larger in the learned case ($p < .05$). Interestingly, the reduction in capture in all conditions was accompanied by looking-away. In both experiments, there were significantly more looks to the location opposite the cued/HPL location compared to all other locations; the other locations did not differ from each other (Cue-to look-distance 1–4: 9.2%, 10%, 13.9%, and 33.9% $p < .001$; HPL-to-look-distance 1–4: 11.5%, 11.1%, 13.0%, and 28.8%, $p < .001$). In conclusion, while neither cue-based nor learning-based distractor control provided strong evidence for spatial suppression below the nonsalient-distractor baseline, they both produced a similar look-away gaze strategy to avoid capture by the salient distractor.

36.446 DETERMINISTIC MISS ERRORS IN VISUAL SEARCH ARE DUE TO THE TARGET

Johan Hulleman¹ (johan.hulleman@manchester.ac.uk), Katerina Sergi¹, Jeremy M. Wolfe^{2,3}; ¹University of Manchester, ²Brigham and Women's Hospital, ³Harvard Medical School

Visual searchers miss targets. Every search task produces stochastic errors, where searchers miss a target due to random bad luck. Some search tasks also produce deterministic errors, where properties of specific scenes or specific stimuli make the target hard to find. If a search display is presented twice, a deterministic error would be likely to occur both times. The chance of a stochastic error occurring twice would simply be the square of the probability of missing once. Do properties of individual targets or of the search display as a whole cause deterministic errors? We investigated by having 22 participants search for Ts amongst Ls, with zero, one, or two targets. Critically, each 24-item display was presented twice (576 trials). Ts in one-target displays were either high or low contrast. In two-target displays, one T was always high contrast; the other T was either high (high-high) or low contrast (high-low). Eye movements were tracked. For high-low trials, 17.8% of search displays yielded a miss error on both presentations. For high-high trials, this only occurred for 4.8% of

displays. The double miss rate for high-high is consistent with purely stochastic misses, but the elevated high-low rate indicates a substantial deterministic component. If participants miss a target on first presentation of a high-low trial, which target do they miss -if they miss- on second presentation? Cases where the missed target changed from high to low or vice versa were in line with the stochastic prediction (4.9% vs. 4%). In contrast, double high and, especially, double low misses occurred almost twice as often as predicted stochastically (12.8% vs. 6.8%). So, the deterministic component of miss errors in high-low trials arises from double misses of the same target. In these tasks, it is the camouflaged item, rather than the camouflaging scene that causes miss errors.

JH and KS were supported by UKRI grant ES/X000443/1. JMW was supported by NIH-NEI: EY017001, NSF: 2146617, and NIH-NCI: CA207490.

36.447 HOW DOES AWARENESS IMPACT THE EXPRESSION OF THE EXPECTANCY EFFECT IN VISUAL SEARCH?

Natalie Paquette¹, Joseph Schmidt¹; ¹University of Central Florida

Violating visual search expectations shifts performance towards the expected search (e.g., Cox et al., 2021). Expecting a difficult search but performing an easy search can decrease guidance, increase RTs and object (target/distractor) dwell times, whereas expecting an easy search but performing a difficult search, can decrease accuracy, RTs and object processing times, relative to when expectations are accurate (Paquette & Schmidt, 2025). Awareness of an expectation violation often facilitates task performance (Pinquart et al., 2021), resulting in reduced expectancy effects. However, if search history drives search performance (Wolfe, 2021), awareness of the upcoming difficulty shouldn't impact expectancy driven changes to search performance. Using a mixed design (between subjects: awareness, predictive/non-predictive cue; within subjects: violation type, harder/easier than expected and stimulus type: Landolt-C's/real-world objects), this study examined effects of awareness of expectation violations on search performance. Participants pictorially previewed (200ms), then after a 1000ms delay, searched for target Landolt-C's or real-world objects in blocks of easy or difficult search (low vs. high target-distractor similarity, respectively). Blocks contained 25% of trials from the unexpected difficulty level to assess performance changes when expectations were violated relative to when they were accurate. A predictive or non-predictive difficulty cue was presented during the delay and maintained during search. Early results show a significant effect of expectancy in both awareness conditions, violation types, and stimulus types in distractor dwell time (all $p < .05$) and reaction times in Landolt-C search (all $p < .05$). Real-world objects generated an additional effect in harder than expected guidance ($p < .05$) and Landolt-C's generated effects in easier than expected target dwell time in the predictive condition. These data suggest that awareness of an expectation violation does not mitigate expectancy

effects, suggesting they may be primarily driven by recent search history rather than trial-by-trial expectations.

36.448 THE EFFECT OF COGNITIVE LOAD ON FULL-BODY GAZE CONTROL DURING 3D SEARCH

Zachary Kelly¹, John Franchak¹; ¹University of California, Riverside

Search in the real world is three-dimensional, and thus requires the use of the head and body to help move the visual field around the environment. Previous research has established that increased cognitive load can alter movement of the eyes during search. Cognitive load also disrupts motor coordination in various tasks. However, it is unclear how cognitive load affects the eye, head, and body movements used to shift gaze during three-dimensional visual search. We gathered head-mounted eye-tracking data from 44 participants who searched an array of 27 images distributed across 3 separate walls in a 270° space and reported the number of images that matched a specified color-object type pairing. At the same time, participants completed a backwards counting task used to manipulate the level of cognitive load. Similar to past work, we found that as cognitive load increased, participants made fewer and longer looks at stimuli, and increased their rate of returning to look at previously viewed stimuli. In addition, we found that participants decreased the amount of head and body movements as cognitive load increased, evidenced by a decrease in wall switches (shifting the gaze from one wall to one of the other two walls). These findings suggest that when cognitive load was high, participants decreased their head and body movements in order to lessen the expenditure of cognitive resources spent on coordinating motor effort. This in combination with the changes in eye movements may suggest that visual search behavior in 3D is not only moderated by cognitive load, but the motor effort required to move the head and body.

36.449 I'LL ASK YOU AGAIN: WHAT HAPPENS WHEN YOU REPEAT A MISSED SEARCH TARGET ON THE NEXT TRIAL?

Daniel Ernst¹, Jeremy M. Wolfe²; ¹Bielefeld University, ²Brigham & Women's Hospital/Harvard Medical School

Why do people miss clearly visible targets in visual search? Such misses can have serious consequences, e.g., in radiology or airport security. Others have shown that observers can respond with above chance accuracy to forced-choice questions about objects that they deny seeing in inattention blindness paradigms (Nartker et al. 2025, eLife, 13, RP100337). In inattention blindness, the missed item is typically irrelevant to the primary task. Here, we seek evidence for partial knowledge about missed targets that observers have been actively seeking. We used a mixed hybrid search design where observers search through displays of pictures of objects. In a block of trials, a target could be one of three specific pictures or any picture from three object categories. This design induces relatively high numbers of miss errors for known targets, especially for categorical targets. Whenever observers missed a target, the identical target was

presented again at a random location among random distractors in an immediate “repetition trial”. Participants were not informed about this manipulation and received no feedback after miss errors. Participants had 85% and 94% accuracy in standard target present trials with categorical and specific targets, respectively. In repetition trials, participants found just 62% of the categorical targets and 80% of the specific targets that they had missed in the previous trial. Interestingly, when the target was found on a repeated trial, it was fixated more quickly than it was fixated in a normal trial. This suggests that some partially processed information from the previous, missed trial might prime the search for the repeated target. On the other hand, the high error rates on repeated trials indicate that whatever led to the original miss still makes a subsequent miss quite likely.

Daniel Ernst was supported by grant ER 962/3-1; Jeremy Wolfe was supported by grants NIH EY01700, CA207490, and NSF2146617

36.450 IS GAZE BIASED TOWARDS LOCATIONS WHERE A TARGET WAS RECENTLY FOUND? EVIDENCE FROM ENVIRONMENTS APPROXIMATING REAL-LIFE SEARCH

Daniel Toledano¹, Tamer Abd Alrazaq¹, Dominique Lamy¹; ¹Tel Aviv University

Laboratory studies consistently show that attention and eye movements are strongly biased toward the location where a previous target was recently found. However, it remains unclear whether this robust location priming effect generalizes to more naturalistic search environments, which differ from highly controlled laboratory environments in important ways: they contain numerous distractors, complex objects, non-uniform backgrounds and spatial layouts, and strong context-related statistical regularities. To address this question, we conducted a preregistered re-analysis of a publicly available dataset (Zelinsky et al., 2019). Participants searched for either microwaves (N= 27) or clocks (N = 27) in real-world scene photographs taken from the COCO2024 dataset while their eye movements were recorded. On each trial, participants indicated whether the target was present or absent. To quantify location priming we measured both the absolute distance and the angular distance between the first saccades’ destinations and the previous target location. We then compared these distances to two baseline measures. We found a bias not only towards the previous target location but also towards the target location from two trials back. In addition, we found that participants’ eye movements dwelled more on locations that were closer to the previous target location. Overall, these findings provide the first evidence that overt attention is strongly biased towards previous target locations in visually rich environments that closely mimic real-world search.

36.451 STATISTICALLY LEARNED COLOR-TARGET ASSOCIATIONS BIAS ATTENTIONAL GUIDANCE AND FIXATIONS FOR SEARCH DECISIONS

Catherine Halpern¹ (cahalpern@ucdavis.edu), Joy Geng¹; ¹UC Davis

The ability to localize and select targets during visual search depends on how well target-defining features are incorporated into the search template in memory. However, heuristics regarding probable target features (that do not define the target, but predict target identity) may also aid search guidance and decision making. We tested this hypothesis by having participants perform a visual search task for a target tilted $\pm 15^\circ$ from vertical among five distractors tilted $\pm 15^\circ$ from horizontal, with eye-tracking. On each trial, the six Gabors were split evenly between two colors (i.e., purple–yellow, green–purple, yellow–green). For each participant, one of the colors was randomly designated as the probable target color, which was predictive of target identity; this contingency was not disclosed. In the two color-pair displays where the probable color was present, the target always appeared in that color (100% color-orientation contingency), as did two distractor items; on neutral trials, (those containing the two non-probable colors) the target appeared in either color on 50% of trials. Preliminary analyses suggest that statistical learning of the predictive color strongly influenced attentional guidance. On displays including the probable color, initial fixations were disproportionately directed to items in that color first ($\mu = 76.5\%$, $\sigma = 8.1\%$), a bias that strengthened throughout the session (7.9% increase from the first half of the session to the second half). Additionally, fixations on non-probable colored distractor items (on probable color displays) had shorter durations during target evaluation, suggesting that participants more rapidly rejected these distractors. Together, the results suggest that statistically-learned, color-target contingencies can be used to guide attention by reducing the effective set size of the search display, and can bias target evaluation for decision making, even when the color that predicts the target does not define it.

UCD, JSMF

Attention: Inattention, attentional blindness

SUNDAY, MAY 17, 2:45 – 6:45 PM, PAVILION

36.452 CRITERION SHIFTS, NOT DECLINES IN SENSITIVITY, CHARACTERIZE THE VIGILANCE DECREMENT IN A CONTINUOUS PERFORMANCE TASK.

Henri Etel Skinner^{1,2}, Tabitha Cook¹, Kamya Mahendru¹, Janani Aiyer¹, Barry Giesbrecht^{1,2}; ¹University of California Santa Barbara, ²Institute for Collaborative Biotechnologies

The vigilance decrement has been long characterized as a decline in performance as individuals continuously maintain attention to rarely occurring vigils. Often vigilance tasks rely upon detection rate to characterize this decline, but studies that measure false alarm rates are also able to compute signal detection metrics of sensitivity and criterion. One blindspot of these methods is that reliance on responses indicating a target is present makes them unable to disambiguate

between true changes in decision making parameters (sensitivity and criterion) and failures to respond outside of these processes (e.g. due to lapses in responding or lapses in attention). Here subjects viewed a continuous stream of images each appearing for short (1640 ms, 90% of trials) or long (2040 ms, 10% trials) durations. Participants were instructed to make a short/long judgement for each trial. Results showed across time (~5 minute blocks), detection rate and false alarm rate (indexed by “long” responses) declined. Miss rate increased and correct rejection rate remained stable (both indexed by “short” responses). Criterion (c) increased over time ($b = 0.0676$, $se = 0.0156$, $p < 0.001$) while sensitivity (d') did not change ($b = -0.0224$, $se = 0.0198$, $p = 0.259$). Lastly, while omission rate (failure to respond on any trial) also increased, this was not the driving force behind criterion shifts. These results suggest the driving factors for declines in detection rate appear to be conservative shifts in criterion. While the continuous nature of this task is different from standard vigilance tasks requiring rare responding and other sustained attention tasks requiring frequent responding and rare withholding (i.e., SART), it adds critical information to the study of sustained attention. Specifically the present results indicate declines in detection rate that are characteristic to the vigilance decrement can be driven primarily by shifts in criterion.

Research was sponsored by the U.S. Army Research Office and accomplished under cooperative agreement W911NF-19-2-0026 for the Institute for Collaborative Biotechnologies.

36.453 FEATURE-SPECIFIC EFFECTS OF MOTION ON CHANGE DETECTION

Rachel Pitman¹, Daryl Wilson¹; ¹Queen's University

In prior research, we found that motion impaired change detection for orientation but not for colour changes – an unexpected finding, as previous research has shown that motion can interfere with colour perception (Suchow & Alvarez, 2011) and memory (Chung et al., 2023). The absence of any motion effect on colour suggested that motion's impact on change detection may not stem from disruptions to attention or memory, but rather to characteristics of the features themselves. Therefore, we conducted follow-up experiments to determine whether motion affects change detection differently across features. Across experiments, participants viewed an array of task-relevant triangles and task-irrelevant circles that were either stationary, moving synchronously, or moving asynchronously. In each trial, one task-relevant stimulus gradually changed while participants attempted to identify the change target. In Experiment 1, we used a staircase procedure to replicate our previous findings and determine the change magnitude required for 50% detection accuracy for colour and orientation changes under high (3 task-relevant stimuli) or low (9 task-relevant stimuli) perceptual load. In Experiment 2, we tested whether motion selectively disrupts change detection for spatial features by testing detection accuracy for size (spatial) versus luminance (non-spatial) changes. Detection accuracy was measured directly (no staircase design) as a function of Motion Type (Stationary, Synchronous, Asynchronous) and Load (3, 6, 9, 12 task-relevant

stimuli). Across experiments, motion consistently impaired detection for spatial features (orientation and size) under high load, producing both a Pure Motion Effect (Synchronous > Static) and a Chaotic Motion Effect (Asynchronous > Synchronous). In contrast, motion had no impact on non-spatial features (colour and luminance), although detection of both declined with increasing load. These results challenge the assumption that motion impairs change detection by disrupting attention or memory, instead supporting a feature-specific mechanism in which motion selectively impairs detection for spatial-related features.

36.454 ATTENTION IN CRISIS: CLARIFYING FLICKER CHANGE BLINDNESS

Justin George¹, Jonathan Flombaum¹; ¹Johns Hopkins University

Computational advances, combined with a better understanding of stimulus pooling in the visual system, suggest that processing limits attributed to attention may be structural rather than resource-limited. Across three behavioral experiments and simulations with a pooling model, we demonstrate that indeed, change detection limitations in a flicker paradigm can be well-characterized without appealing to a limited attentional capacity. We found that, in addition to obscuring changes outside fixation, pooling creates a threshold level of “change noise” across an image, which legitimate change signals must overcome. Participants searched for changes in flickering photographic images. Initially, we found that image-wise reaction times were correlated for scenes presented upright versus inverted, implicating image structure over attention. We then derived a model-based change-magnitude by subtracting the model's pooled feature vectors for each base and changed image across multiple fixations. This isolates the altered region because unchanged areas cancel. This metric did not predict detection latencies, even after several reweightings, indicating that low-level pooling corruption alone cannot explain performance. The model pools features over progressively larger, receptive field-like regions with distance from fixation, enabling synthesis of metamers that match its pooled statistics. Using metamered versions of each base and changed image, we computed pixel-wise differences to obtain a change-magnitude that includes pooling “change noise” in unchanged regions; this measure correlated strongly and significantly with detection latencies. We conclude that change blindness in this context arises from a combination of undetectable differences when a change region is outside fixation and, crucially, with how much other regions of the image may appear to change. We further demonstrate through simulation how attention can be conceptualized as capacity-free spatial selection such that it reduces the impact of general change noise, producing novel predictions for cueing experiments.

36.455 THE GORILLA IN THE MACHINE: REVERSE-ENGINEERING INATTENTIONAL BLINDNESS WITH A NEW 'PERCEPTUAL REFRAMING' HYPOTHESIS

Mario Belledonne¹ (mario.belledonne@yale.edu), Ilker Yildirim^{1,2}; ¹Yale University, ²Wu Tsai Institute

Inattentional Blindness demonstrates sophisticated tradeoffs realized on-the-fly by visual cognition: When focused on a goal, we are often unaware of even salient events that involve irrelevant elements. What is the computational nature of these tradeoffs, and how do patterns of selective processing ultimately impact visual awareness? Here, we present the *Perceptual Reframing Hypothesis*, in which inattentional blindness arises from adaptive processes within visual cognition that balance the computational complexity of a perceptual “frame” (e.g., how to represent and organize objects in the scene) with the information necessary for decision-making. We realize this hypothesis with a novel computational mechanism, Multigranular Optimization (MO), which dynamically builds and manipulates increasingly more efficient frames. Every several hundred milliseconds, the MO model uses an online stream of task-relevance signals to reframe object representations — on a ladder from fine-grained individuals to coarse ensembles. Under bounded resources, the model maintains task-relevant objects as fine-grained individuals to sufficiently inform decision-making and reframes task-irrelevant elements into coarse, cheap ensemble summaries to exponentially decrease computational complexity. This process yields humanlike patterns of awareness, as reframed ensemble representations preclude the explicit encoding of visually similar, irrelevant objects by “explaining away” their corresponding sensory signals. Across several studies, the MO model not only recapitulates classical awareness patterns (e.g, Most et al., 2001) but also predicts new awareness phenomena that we confirm in a new preregistered experiment. The model also explains, for the first time, detailed, trial-level variance in humans’ awareness responses. Alternative ablation models do not recover humanlike patterns, and crucially, detailed comparisons establish that MO realizes resource-rational performance benefits that outweigh any algorithmic overhead. By revealing the algorithmic basis of inattentional blindness, this work begins to ground the striking phenomenology of awareness in precise and interpretable computational terms.

AFORS, NSF CAREER

36.456 TWO STAGES OF PROCESSING FOR VISUAL AWARENESS IDENTIFIED IN A NOVEL INATTENTIONAL BLINDNESS VIDEO-GAME PARADIGM

Michael Pitts¹, Gennadiy Belonosov², Oscar Ferrante³, Ling Liu⁴, Rony Hirschhorn², Ole Jensen⁵, Huan Luo⁴, Lucia Melloni⁶, Liad Mudrik²; ¹Reed College, ²Tel Aviv University, ³University of Surrey, ⁴Peking University, ⁵University of Oxford, ⁶Ruhr-Universität Bochum

What is the difference in neural processing when we are aware versus unaware of a stimulus? Previous attempts to identify neural signals uniquely associated with visual awareness have faced several challenges. A critical one has been to isolate neural events linked with perception from those associated with reporting one’s perception. To address this, we developed a dual-task inattention paradigm in which participants played an engaging video-game as the primary task, while face and object stimuli were occasionally presented in the background. In 1/3 of the trials, a secondary task was presented immediately after

stimulus offset (500ms post-stimulus-onset): the game was paused and participants were prompted to report whether they just saw a stimulus. We first analyzed MEG-EEG data from the report trials. We then trained a pattern classifier to differentiate reported-seen versus reported-unseen trials, and used it to label the remaining 2/3 of the EEG data (no-report trials). In the report trials, we found a visual awareness negativity (VAN; ~220-270ms) followed by a fronto-central negativity (fcN2) and concurrent late bilateral occipital positivity (LBOP) from ~320-550ms. The P3b was also evident but only at ~800-1000ms post-stimulus (~300-500ms after the report-prompts). In the classifier-labeled no-report trials, we found the VAN and fcN2/LBOP with no evidence for P3b. These results demonstrate the potency of this paradigm and classification approach for isolating neural events linked with conscious perception from perceptual-reports. They also suggest that after the VAN, there may be a second stage of processing necessary for visual awareness, indexed here by the fcN2/LBOP.

Templeton World Charity Foundation

Attention: Spatial

SUNDAY, MAY 17, 2:45 – 6:45 PM, PAVILION

36.457 SEEING THE TARGET MATTERS: STATISTICAL LEARNING OPERATES ON PERCEPTION, NOT MENTAL CONSTRUCTION

Agnes Rigo¹ (agnes.rigo@nyu.edu), Daryl Fougnie¹; ¹Division of Science and Mathematics, New York University Abu Dhabi

Our environment contains far more visual information than we can process, making prioritization critical. Environmental regularities can guide attention and improve visual processing through statistical learning. A key question is how complex the information guiding attention can be. What happens when the regularities are not directly present in the visual input, but instead must be internally constructed? To address this question, we employed a contextual cueing paradigm in which the color and shape of a pre-cue explicitly defined the target while implicitly predicting target spatial location in a subsequent visual search task. For half of the color-shape pairings, the target appeared in one of eight positions with 100% consistency (mapped), whereas the remaining pairings were unmapped and did not predict target location. Critically we manipulated the relation between color and shape information such that it was either present as a fully formed object (Experiment 1) or required mentally combining features from two distinct objects, either in separate (Experiment 2) or overlapping spatial positions (Experiment 3) (all n=20). Experiment 1 produced a robust contextual-cueing effect: mapped trials were significantly faster than unmapped trials ($t(19)=4.760$, $p<.001$). Critically, this effect remained even after excluding participants who reported noticing the mapping, indicating that the guidance happened implicitly. Experiments 2 and 3, in contrast, showed no evidence of statistical learning: reaction times were equivalent for mapped and unmapped

trials (Experiment 2, $t(19)=1.022$, $p=.320$; Experiment 3, $t(19)=.430$, $p=.672$), except among participants who explicitly reported noticing the mapping. Thus, while participants were able to explicitly combine color and shape (search performance was >90%) they showed no evidence of guidance when these features were split across objects. These findings suggest surprising limits on statistical learning: it succeeds when regularities are visible and contained within an object, but fails when they must be mentally constructed from separate objects.

This work was supported by the NYUAD Research Institute grant CG012 under the NYUAD Center for Brain and Health.

36.458 TARGET SELECTION VERSUS DISTRACTOR INHIBITION: AWARENESS ENABLES WORLD-CENTERED TARGET LOCATION LEARNING, BUT DISTRACTOR INHIBITION REMAINS EGOCENTRIC

Litian Chen^{1,2} (l.chen2@vu.nl), Shichen Jia³, Caterina Fatato³, Freek van Ede^{1,2}, Heleen A. Slagter^{1,2}; ¹Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, ²Institute Brain and Behaviour Amsterdam, ³Department of Psychology, University of Amsterdam

Both the allocation of attention to goal-relevant information and the inhibition of distracting information are fundamentally shaped by statistical learning. For instance, attention is automatically facilitated at locations where targets are likely to appear and inhibited at locations where distractors frequently occur. In the domain of target-location learning, the spatial reference frame in which this statistical learning operates appears to depend on awareness of the regularity: with explicit awareness, target location learning can be world-centered (allocentric), whereas without awareness, it remains viewer-dependent (egocentric). We and others recently extended this work by showing that implicit distractor-location learning similarly relies on a viewer-centered. Here, we investigated whether explicit distractor knowledge can similarly shift distractor-location learning to an allocentric reference frame. In six experiments (total $n=167$), participants performed an additional-singleton search task projected on a tabletop while standing. On each trial, they moved to a different side of the table, so that we could manipulate whether the high-probability target or distractor location was fixed in the world (allocentric) or stable relative to their viewpoint (egocentric). We replicated past findings that with explicit target knowledge, participants successfully learned where in the world targets likely occurred, while in the unaware condition, no target-location learning was observed in the world-centered frame. In contrast, distractor-location learning remained stubbornly viewer-centered, even with explicit knowledge. These findings reveal a clear dissociation between target selection and distractor inhibition: explicit knowledge enables world-centered learning – generalizing across viewpoints – when applied to goal-relevant information, while distractor inhibition remains viewpoint dependent, regardless of awareness. We posit that distractor inhibition may be more fundamentally anchored in immediate sensorimotor

processing, guiding the agent on how not to attentionally sample the environment from their viewpoint.

36.459 MODULATION OF VISUAL ATTENTION BY TASK-RELEVANT INFORMATION DURING SIMULATED DRIVING

Noah Britt¹ (brittn@mcmaster.ca), Juan Lupiáñez², Elisa Martín-Arévalo², Ana Chica², Hong-jin Sun¹; ¹Department of Psychology, Neuroscience, and Behaviour, McMaster University, Canada, ²Department of Experimental Psychology, and Brain, Mind, and Behavior Research Center (CIMCYC), University of Granada, Granada, Spain

Through daily experiences, we learn where and what information our spatial attention should be prioritized to. We (Britt et al., 2025; CRPI) showed that, during simulated driving, drivers attend more to roadside pedestrian targets oriented toward the road than away from it, demonstrated in an enhanced cueing effect in the form of facilitation in an exogenous cueing task with SOA of longer than 1000ms. However, such attention modulation by target orientation was not observed in the same environment when the participants' viewpoint remained stationary or during driving, but when the target was changed to a light post. The classical cueing literature has shown that if a central, attention-capturing event—known as an ‘intervening event’—appears between the cue and the target, cueing effects can shift from facilitation to inhibition of return (Martín-Arévalo et al., 2013; APP). In this study, we sought to observe, with the implementation of an intervening event, if our earlier findings of the context-dependent and orientation-dependent modulation of attention would be replicated. During the time interval between the appearance of the cue and the target, we implemented a driving-relevant stimulus (flashing the tail-light of a centrally located lead car). Participants were required to discriminate the hand position of the pedestrian (up, down) or the array position of the light within the light post (top, bottom). The results revealed that despite adding the intervening event, when the target was the roadside pedestrian, facilitation effects remained and with enhanced facilitation for inward (facing the road) compared to outward orientations. However, when the roadside target was the arbitrary light post, or participants' viewpoint remained stationary, an inhibition of return effect was observed. These results demonstrate again that the behavioural significance of roadside pedestrians changes how participants orient attention across stimuli, but importantly, in a contextually-specific manner (i.e., specifically only while actively driving).

NSERC

36.460 ALL THAT DOESN'T GLITTER CAN BE GOLD UNDER THE RIGHT CONTEXT: MOVIE-VIEWERS' EVENT MODELS INFLUENCE ATTENTION WHEN A TARGET WITH A LOW DEFAULT GAZE PROBABILITY IS PRIORITIZED.

Prasanth Chandran¹, Baile Dreyfuss², Camille Dortland², Claire Holt², Eva Naab², Karla Salto², Laurel Conner², Maddy Pilarz², Joseph P. Magliano³, Lester C. Loschky⁴; ¹Doctoral Student, Kansas State

University, ²Undergraduate Student, Kansas State University, ³Professor, Georgia State University, ⁴Professor, Kansas State University

A person's event model—their understanding of events—has been shown to influence their attention during reading and viewing static scenes. But its effect on attentional selection in films is minimal or absent, a phenomenon known as the Tyranny of Film. This study provides evidence for the event model's effect on selective visual attention in conventional film clips. Using the Scene Perception & Event Comprehension Theory (SPECT), we hypothesized two necessary conditions for a viewer's event model to impact their attention to a target: 1) The Differential Event Model Criterion: The target of attention is prioritized by a strong event model; 2) The Default Gaze Probability Criterion: The target of attention has a relatively low visually-based default gaze probability. We verified whether our selected video clips met the first criterion by examining differences in event understanding through linguistic measures and moment-to-moment spatial importance judgments. We confirmed the second criterion, that the target's default gaze probability was relatively low, using a state-of-the-art gaze prediction algorithm, DeepGaze IIE. We identified eight video clips fulfilling both criteria. We then tested whether these differences in event understanding affected attentional selection in a within-subjects eye-tracking study (N = 44) where participants first watched the videos without context, then with context. The target of attention was fixated four times more often in the Context condition than the No-context condition, which was at chance level. Additionally, over time in the Context, the likelihood of fixating on the target increased as its importance grew; this increase was five times larger than the No-context condition. Finally, a priori event-model importance had a stronger influence on attentional selection than a priori default gaze predictions based on stimulus features. In conclusion, these two criteria are sufficient for the event model to influence attentional selection during film viewing, thereby overcoming the Tyranny of Film.

Research reported in this abstract was partially supported by the Cognitive and Neurobiological Approaches to Plasticity (CNAP) Center of Biomedical Research Excellence (COBRE) of the National Institutes of Health under grant number P20GM113109.

36.461 ELECTROPHYSIOLOGICAL EVIDENCE FOR DISSOCIATION BETWEEN COVERT SPATIAL ATTENTION AND WORKING MEMORY GATING

Brecken Marome¹, Henry Jones¹, William Thyer¹, Edward Vogel¹, Edward Awh¹; ¹University of Chicago

To be successful at visual tasks, observers must selectively process information from cluttered environments using covert spatial attention, which selects specific regions of space, and working memory (WM) gating, which selectively encodes items into limited-capacity storage. We analyzed data from a basic change detection task, where participants encoded square target items from amongst rectangular

distractors. Some trials contained 2 square targets (ss2), while others contained one target and one rectangular distractor (ss1). For ss1 trials, target/distractor similarity was manipulated such that distractors were either very dissimilar (long grey rectangles), dissimilar (long colored rectangles), or similar (squat colored rectangles). Late in the delay period, spatially-selective alpha activity showed that attention was biased towards targets over distractors. However, during the first ~600 ms after stimulus onset, spatial attention was precisely oriented towards both targets and distractors, regardless of target-distractor similarity. A WM load classifier showed that gating of distractors happened as early as 180 ms contrasting with the less selective deployment of spatial attention. In addition, the classifier showed that similar distractors were more likely to be unintentionally encoded into WM. In next steps, we are examining visually-evoked ERPs at either target or distractor locations to corroborate (or call into question) the slow time course of spatial attention that is indicated by alpha activity. Taken together, our results so far provide further evidence that covert spatial attention and WM gating represent distinct modes of attentional control.

36.462 SEATED COMPUTER-BASED AND WHOLE-BODY ACTION-BASED VISUAL TASKS ONLY ALIGN UNDER SPECIFIC VISUAL COGNITIVE AND MOTOR DEMANDS

Leah Gano¹, Barry Giesbrecht¹; ¹University of California Santa Barbara

Performance in everyday tasks requires the integrated functioning of action and cognitive systems. Tasks combining physical activity with cognitive demands improve cognitive functioning more than physical activity alone (Anderson-Hanley et al., 2012; Barcelos et al., 2015). It remains unclear whether cognitive abilities demonstrated in integrated action-cognition tasks are associated with performance on traditional computer-based assessments. We examined correlations between performance on two computer-based visual cognition tasks, the Attentional Networks Task (ANT) and a visual working memory change detection task; and two integrated action-cognition tasks, a modified spatial cueing task and an n-back task. Subjects (n=39) completed all tasks in a counterbalanced order. For the computer-based tasks, subjects were seated and responded using their fingers on a keyboard. For the action-cognition tasks, subjects stood and responded using either both hands and feet (cueing task) or using one hand (n-back). Both types of tasks replicated the classic behavioral patterns typically reported in the literature. In the ANT, RTs were faster for cued vs. uncued locations ($t(38)=8.22$, $p<.001$). In the change detection task, accuracy declined with increasing set size ($F(2,76)=177.45$, $p<.001$). In the action-cognition tasks, RTs were faster for cued vs. uncued target locations ($t(38)=-11.95$, $p<.001$) and accuracy declined as "n" increased ($t(38)=14.44$, $p<.001$). Correlations between the tasks revealed an association between uncued RTs in the ANT and cueing task ($r(37)=.35$, $p=.030$). Exploratory analyses revealed that this correlation was driven primarily by hand responses, not foot responses. The correlations between the cueing effects and the

working memory load effects across task type were not significant. These findings suggest that the association between computer-based tasks and integrated action-cognition tasks emerges only under specific cognitive and motor response conditions, indicating limited transfer across modalities.

Research was sponsored by the U.S. Army Research Office and accomplished under cooperative agreement W911NF-19-2-0026 for the Institute for Collaborative Biotechnologies.

36.463 COVERT VS. OVERT SPATIAL ATTENTION EFFECTS ON SCENE COMPREHENSION.

Sana Shehabi¹, Srijita Karmakar¹, Miguel P. Eckstein¹; ¹University of California, Santa Barbara

Introduction: Covert and overt attention improve performance in a variety of basic perceptual tasks: detection, contrast and orientation discrimination, letter identification, and scene categorization (Posner, 1980; Carrasco, 2011). Scene understanding is crucial for everyday visual cognition, but we know little about how attention influences the human ability to comprehend scenes. Here, we assess whether covert and overt attention enhance real-world scene comprehension. **Methods:** In Experiment 1 (covert attention; N=19), observers completed a modified Posner task under enforced fixation. Ninety percent of randomly sampled trials followed a standard 2-location Gabor orientation-identification (± 5 deg) paradigm (80% valid, 20% invalid peripheral precues, Gabor presentation: 50 ms). The remaining 10% were “scene trials,” where Gabors were replaced with real-world images (300 ms) containing individuals performing distinct actions on each side. A response cue (250 ms) indicated which side’s scene observers should describe (50% response cue validity relative to the precue). Semantic similarity index (SSI) between observers’ descriptions and gold-standard descriptions was quantified using cosine similarity computed from text embeddings. Experiment 2 (overt attention; N=10) used the same design, except observers were free to move their eyes. **Results:** Both experiments showed significant and comparable cueing effects for Gabor trials ($p = 0.90$, covert vs. overt). Critically, covert attention did not enhance scene comprehension: SSI did not differ (bootstrap $p = .060$) between valid (0.609 ± 0.014) and invalid (0.579 ± 0.012) cued scenes. In contrast, overt attention resulted in a strong cueing effect on scene comprehension: SSI were higher for valid (0.662 ± 0.012) than invalid (0.438 ± 0.026) cued scenes (bootstrap $p < .001$). **Conclusion:** While covert attention produces benefits for basic perceptual tasks, its influence on the comprehension of real-world scenes is small and less than $\sim 1/8$ th, relative to overt attention.

36.464 RETINAL MOTION STATISTICS ON CROWDED SIDEWALKS

Mary Hayhoe¹, Youjin Oh, Daniel Panfili, Nathaniel Powell; ¹University of Texas Austin

In natural environments it is unclear how attention is directed to information needed to avoid independently moving obstacles that are largely unpredictable. It is often assumed that attention to such objects is attracted by stimulus-driven signals, such as motion. However, retinal motion patterns are actively shaped by gaze behavior, by walkers’ gait, and by the structure of the environment. Little is known about the nature of the time-varying retinal stimulus since it has been difficult to gather this kind of data, and retinal object motion is confounded with the effects of self motion, adding complexity to the signal. In addition, familiarity with the probabilistic structure of natural environments might allow a more pro-active role, implicating both bottom-up and top-down strategies. We recorded gaze, head, and foot movements, together with 3D scene data and depth estimation, while subjects walked on crowded sidewalks. We extracted retina-centered motion statistics and gaze distributions in the scene. Because gaze is stabilized by the vestibular-ocular reflex, motion is always zero at the fovea and increases with eccentricity. We calculated the statistics of these retinal patterns and found obvious systematic changes in velocity distribution with the addition of pedestrians to the scene. This might allow independently moving objects to be detected as statistical outliers, suggesting that peripheral motion contrast might attract gaze. To test this, motion contrast caused by stationary 3D structure must be accounted for. We are currently analyzing the statistics of spatio-temporal contrast patterns caused by these three sources (self-motion, object-motion, and stationary 3D structure) of retinal motion. Preliminary data suggest that the statistical patterns from the different sources are easily discriminable. In addition, examination of gaze patterns suggest that walkers frequently monitor distant regions in the scene, supporting the role of proactive (top-down) strategies for pedestrian detection and avoidance as well.

Supported by: NIH EY05729, NSF 2318065

36.465 ADAPTING THE POSNER PARADIGM TO STUDY 360° ATTENTION ORIENTING IN VIRTUAL REALITY

Chloé Deschanet¹ (chloe.deschanet@unige.ch), Loïc Kreienbühl¹, Patrik Vuilleumier^{1,3}, Joan Llobera^{2,3}; ¹Laboratory for Behavioral Neurology and Imaging of Cognition (LabNIC), Department of Fundamental Neuroscience, University Medical Center & Campus Biotech, University of Geneva, Geneva, Switzerland, ²Artanim Fondation, Geneva, Switzerland, ³Equal supervision

Orienting attention in space is crucial to modulate perception and adapt behavior. The Posner paradigm has been widely used to understand the cognitive and neural mechanisms that shift attention across locations. However, most of the literature has focused in frontal space. Here we adopt immersive Virtual Reality (VR) and adapt the Posner paradigm to study the deployment of attention both in front and rear peripersonal space using both response times and eye movements. Twenty participants in a VR storage room had to search for and discriminate as quickly as possible between 2 targets in 4 possible locations arranged circularly around them. Targets were preceded by a spatialized sound cue whose validity was manipulated

(80% Valid, 20% Invalid), and presented with varying distractors (None, One near the target, One far from the target, or Three), with the goal of validating the serial search effect in both front and rear space. Response times were comparable to the classical Posner paradigm when the target was presented both front and rear to the participant, with participants being significantly faster with Valid cues, compared to Invalid ones. Distractors also delayed significantly target detection. However, response times for rear conditions were significantly greater, possibly due to greater spatial exploration and body movements. Gaze data confirmed the impact of invalid cues and distractors on attention shifts, and were tightly correlated with response times for both front and rear space. We believe this is one of the first studies to quantify eye-tracking in 360° around the participant and validates VR as a powerful tool for quantifying attentional dynamics. Results show that orienting attention in front and rear space engages similar mechanisms of sensory competition, supporting an overlap or a continuity in space representation, while revealing how overt gaze strategies mediate performance.

This research is supported by ABC-Space, research grant #10000279 awarded by the Swiss National Science Foundation (SNSF). We would like to thank Yves Schmid-Dornbierer from the Artanim Foundation for helping in the development of the VR environment.

36.466 SPATIAL AND SOCIAL FUNCTIONS OF THE HUMAN GAZE: GAZE FOLLOWING AND MUTUAL GAZE PERCEIVED BY AN EXTERNAL OBSERVER

Maria Falikman¹ (maria.falikman@gmail.com), Rob Sullivan², Amelia Clark¹, Myers Bell¹, Sienna McFarland¹; ¹The University of The South, ²University of Virginia

Human gaze perception evolved to support social interaction and cooperation. Another person's gaze direction is used to obtain information, and gaze contact is a straightforward way to engage in interaction. Yet, it is unclear how the two types of gaze exchange are processed by an observer. To address this question, we developed a novel peripheral gaze shift detection paradigm based on the gaze cueing paradigm. We presented three faces on the computer screen, one in the center of the screen and two—at 6 degrees to the right and to the left of it, and instructed participants to respond to a gaze shift of one of the peripheral faces while fixating the central face. The central face looked towards the left or right peripheral face. In 200 ms, the corresponding peripheral face looked either towards the central face (mutual gaze) or in the same direction as the central face (gaze following). Participants gave a 4-AFC manual response about the location and direction of a peripheral face's gaze shift. Gaze following was detected significantly faster than mutual gaze, a result that may have evolutionary relevance, with the gaze acting as a spatial cue towards vital information. Post-experimental debriefing also revealed a lack of metacognitive access to the efficiency of gaze direction detection. The gaze following superiority effect was replicated in the second experiment, in which we manipulated the instruction by using two types of prompts: one emphasizing teamwork between the

observer and the agents on the screen, and the other emphasizing individual competition. We hypothesized that prioritizing cooperation can facilitate mutual gaze processing. However, responses to gaze following were again significantly faster than to mutual gaze, and there was no interaction between gaze type and instruction, probably indicating the primacy of a shared focus of attention in both competition and cooperation.

36.467 MULTI-SCALE STRUCTURAL COMPLEXITY OF PICTURES CORRELATES WITH GAZE BEHAVIOR

Anna Kravchenko¹, Andrey Bagrov¹, Veronica Dudarev²; ¹Radboud University, ²University of British Columbia

Image complexity affects a variety of aspects of processing, including recognition and search efficiency, attention distribution and the likelihood that scenes and objects will be remembered (Sun & Firestone, 2021; Hensellek, 2025; Kyle-Davidson et al., 2025; Endemann & Kamp, 2024). Yet, quantifying image complexity remains a persistent problem. Multiple measures of complexity have been proposed in the literature, from entropy estimate to more complex measures of the number of regions or visual congestion, however, no single metric is generally accepted as the standard, their performance varying widely between images of different types, e.g. artwork vs. photographs (Saraee et al., 2018; Saraee et al., 2020). Here we employ Multi-Scale Structural Complexity (MSSC), a measure that is designed to quantify the presence of structure at distinct spatial scales within images (Bagrov et al., 2020). It has shown reliable correlations with human ranking of image complexity and robust performance across categories (Kravchenko et al., 2025). To test behavioral relevance, we used MSSC to predict scanning patterns (number and duration of fixations) in 101 participants who were viewing naturalistic scenes (data published in Dudarev et al., 2025). Each scene was displayed for 5 seconds alongside image of a simple object. The scenes varied in emotional valence (negative, neutral) and presence of people. We demonstrate that MSSC has higher correlation with the number of fixations than entropy and visual congestion. MSSC also showed an indirect relationship with duration of fixations, in that the two correlated in images without people. When people were present in the picture, images were processed with shorter fixations overall (this effect is reported and discussed in Dudarev et al., 2025), despite no difference in MSSC. We suggest that this discrepancy reflects a distinction between perceptual and conceptual complexity and argue that MSSC provides a framework for disentangling the two.

36.469 RESOURCE COMPETITION IN ATTENTION: CONCURRENT TARGET PROCESSING ATTENUATES EARLY DISTRACTOR SUPPRESSION

Lou Lescuyer De Decker^{1,2}, Dirk Kerzel¹; ¹University of Geneva, ²University of Lausanne

The signal suppression hypothesis posits that the brain actively inhibits salient distractors to prevent attentional capture. While the distractor positivity (PD) is established as an electrophysiological marker of this

suppression, its functional independence from target discrimination processes remains an open question. Adapting Drisdelle and Eimer (2021)'s paradigm, we orthogonally manipulated target (shape-defined) and distractor (color-singleton) presence in a visual search task, such that they never occupied the same axis. Participants had to enumerate targets across 4 sequential displays and then report how many they detected. To disentangle target and distractor processing, we contrasted displays where both target and distractor were present against displays where either target or distractor was absent. We hypothesized that isolating these components would reveal distinct electrophysiological signatures for capture (N2pc) and suppression (PD). Using a bootstrap-based approach adapted from Sawaki et al. (2012) we observed a comparable N2pc for all target-present conditions, regardless of distractor presence. Crucially, distractors elicited an early PD (105-224ms) in both target-present and target-absent conditions. Interestingly, this PD was larger in target-absent trials. Furthermore, it was consistently followed by a distractor-elicited N2pc (234-311ms). These findings support the existence of a rapid, active inhibitory mechanism that operates before attentional selection. However, the attenuated PD with concurrent target processing suggests that suppression efficiency is resource-dependent, resulting in less inhibition. The subsequent N2pc indicates that despite early inhibition, salient distractors still trigger delayed attentional orienting. To gather definitive evidence, we extended the paradigm to a multimodal setup combining EEG with frequency-domain fNIRS. We are currently running this protocol with the hypothesis that the PD will correlate with an enhanced activation in frontal inhibitory regions (anterior cingulate cortex and prefrontal cortex) and a reduced activity in the visual cortex contralateral to the distractor, confirming the inhibitory nature of this component.

This work is supported by the Swiss National Science Foundation (SNSF) grant n°10001_386.

36.470 SEQUENTIAL SPATIAL REGULARITIES BIAS VISUAL ATTENTION AND GUIDE ANTICIPATORY EYE MOVEMENTS

Keara M. Ginell¹, Ren E. Hentz¹, David E. Osher¹; ¹The Ohio State University

Implicitly learned sequential spatial regularities bias visual attention during search. Previous research shows that memory-guided and top-down attentional biases reflect predictive eye movements. However, whether memory-guided attention, through sequential spatial regularities, also guides anticipatory eye movements is unclear. This question is imperative for understanding how attentional habits emerge through experience. We investigated whether implicitly learned spatial sequences guide pre-stimulus eye movements, using an attentional serial reaction time (attSRT) task. On each trial, a Gabor appeared briefly at one of four locations, and participants reported its tilt. Blocks alternated between learned sequences, where target locations followed a repeating spatial sequence across trials; and novel sequences, where locations were randomized. Participants were

not informed of the sequence and could freely move their eyes. Eye movements were analyzed using time-to-first fixation on the target location, and the proportion of gaze samples in target vs. non-target locations (during both the pre-stimulus period and stimulus presentation). After learning the sequential regularities, fixations tended to reach the target location earlier than when the regularities were absent. This pattern, along with faster reaction times for the learned condition, suggests that sequential regularities may provide a spatial attentional bias via efficient eye movements. The proportion of gaze samples at stimulus onset was higher within the target location than non-target locations, confirming that eye movements were made to the Gabor target. During the pre-stimulus period, the proportion of gaze samples to the target location was greater for the learned condition compared to the novel condition. Collectively, the faster time-to-fixate and higher proportion of gaze samples to the pre-stimulus target location suggest that implicitly learned sequential spatial regularities can guide predictive eye movements while boosting behavioral performance (e.g., faster response times). We conclude that attentional habits may form efficient search performance, partially through guiding predictive eye movements.

36.471 ACTIVE SAMPLING OF VISUAL STIMULI IS UNNECESSARY FOR IMPLICITLY LEARNED ATTENTION

Emma Holtz¹, Vanessa Lee¹; ¹University of Minnesota

Past experience often guides spatial attention: locations that frequently contain a visual search target are prioritized (location probability learning, LPL), even without conscious awareness. Previous studies suggest that LPL may reflect a habit of shifting attention towards target-rich locations, raising questions about whether its acquisition depends on active control over where to shift attention. To address the role of active visual sampling in implicitly learned attention, we administered a novel two-person yoked visual search task. Participant 1 (P1) actively explored the search display containing one T and 11 Ls using a mouse cursor with a 6.7deg-diameter region to uncover items until finding the target. The T was more often in one visual quadrant than the others. Participant 2 (P2) viewed a recording of P1's displays while also doing the search task; P2 had no control over the mouse or the pace of the trials. P2 later gained control over the mouse in a neutral testing phase during which the target's location was random. Results showed that P1 developed LPL, with faster RT in the high-probability than low-probability target quadrants. When performing the search task over P1's displays, P2 was moderately successful (13% miss rate; 73% accuracy on hit trials), while displaying strong correlations in search RTs with P1. Importantly, in the final neutral testing phase, P2, who now performed with 97% accuracy, showed faster RT in the previously rich quadrant than the sparse quadrants. Unlike P1, who showed greater LPL when they became aware of the rich quadrant, P2's awareness of the rich quadrant did not affect their learning. These results showed, for the first time, that active sampling of visual stimuli is unnecessary for the acquisition of implicitly learned attention.

Attention: Individual differences

SUNDAY, MAY 17, 2:45 – 6:45 PM, PAVILION

36.472 EFFECTS OF FEEDBACK ON INDIVIDUAL DIFFERENCES IN ATTENTIONAL ASYMMETRIES

Zainab Haseeb¹ (zainab.haseeb@mail.utoronto.ca), Anna Kosovicheva^{1,2}; ¹University of Toronto, ²University of Toronto Mississauga

Human observers differ substantially in how effectively they can attend to locations across the visual field. Even when gaze is fixed, some regions within the functional visual field (FVF) are consistently identified more readily than others. Recent work shows that these “strong” and “weak” regions vary as a function of polar angle around fixation. These results show stable and idiosyncratic patterns of performance within individuals, suggesting that covert attention is distributed unevenly across space. This study tested whether individual differences in these attentional asymmetries can be modified through feedback, or if they are resistant to change, and whether metacognitive insight is associated with feedback-related improvements. Participants completed a visual search task in which a briefly presented rotated T shape appeared among distractor Ls at eight isoeccentric locations. Participants reported the orientation of the shape. A baseline session measured each participant’s error distribution and metacognitive judgments of performance across spatial locations. In a subsequent session, participants received trial- and block-level feedback of error rates by location. The baseline and feedback conditions both revealed reliable idiosyncratic asymmetries in covert attention, with clear “strong” and “weak” regions for each participant. Feedback improved performance overall but it did not reliably improve uniformity; a comparison of entropy in error rates between conditions showed no significant systematic reduction in spatial asymmetries. As expected, however, metacognition improved robustly with feedback, such that participants’ confidence ratings and end-of-task spatial rankings aligned closely with actual accuracy. However, these gains did not translate into more uniform performance. Moreover, individuals with higher baseline metacognitive accuracy did not show greater feedback-related increases in entropy, suggesting that even accurate insight does not readily compensate for attentional limitations. While awareness of accuracy can be modified, the underlying attentional asymmetries are seemingly resistant to change.

This work was supported by an NSERC Discovery Grant to AK

36.473 INDIVIDUAL DIFFERENCES IN INITIAL EYE MOVEMENT BEHAVIOR AND THEIR RELATIONSHIP TO SEARCH SPEED

Carly J. Leonard¹ (carly.leonard@ucdenver.edu), Ryan V. Ringer¹; ¹University of Colorado Denver

There is no question that individual differences are prevalent across a wide range of psychological and physiological processes. Previous work has shown that this is indeed true for oculomotor behavior, with individuals varying substantially on measures such as average fixation duration and saccade amplitude during visual tasks. Recent studies in our lab have focused on the latency of initial saccades during visual search for a target defined by contour integration. Of particular interest is a speed accuracy tradeoff, such that some individuals tend towards fast first eye movements that land further from the target, while others wait longer and make a more accurate first saccade. The question of interest here is how this initial eye movement behavior relates to task completion (i.e., manual response), as previous results suggested that it may not (Leonard, 2024). To address this, eye movements during visual search were analyzed in a large sample of participants (N = 74). Results replicated previous individual difference findings, showing a correlation between first saccade latency and its landing distance from the target. Critically, the results showed that this first saccade latency did not significantly correlate with manual reaction time. This suggests that while those with slower eye movements may have higher first saccade accuracy, delaying did not directly confer benefits to task performance. More detailed analyses show nuances of how the first saccadic decision is mediated by intermediate saccade behavior and decision time to result in overall manual response speed. These results indicate the importance of considering how small decisions, like that of making a single eye movement, may differ between individuals but not always translate to more comprehensive measures of performance. More broadly, the findings are discussed in terms of how individual differences in a range of processing abilities may interact with eye movement decisions.

R15EY035056-01 to C.JL

36.474 EVALUATING THE EFFECT OF THE VALUE AND VARIABILITY OF REWARDS ON VOLUNTARY ATTENTIONAL CONTROL

Molly McKinney¹ (mollyr.mckinney@tamu.edu), Mianzhi Hu², Darrell Worthy³, Brian Anderson⁴; ¹Texas A&M University

Studies attempting to elucidate the visual search strategies individuals bring to bear in the performance of a task have identified vast individual differences; of note is the large proportion of suboptimal, or slow, visual search strategies. Our first experiment examined whether reward considerations, search efficiency, or the confluence of the two better predicted search strategy for each of 100 participants. Using the Adaptive Choice Visual Search (ACVS) task, a large and small subset of potential target items each contained one target; the target in either the smaller or larger subset was associated with greater reward and participants could choose which target they wanted to search for on each trial. Searching among the smaller subset of items is the fastest way to perform the task. Computational modelling revealed three different strategies participants used: value-based, in which they searched to maximize reward, response-time-based, in which they searched to maximize speed above and beyond reward outcome, and

a reward-per-unit-time-based strategy, which highlighted a potentially joint influence of these priorities. Given that some participants deviated from a purely reward-maximizing strategy, our second and ongoing experiment set out to evaluate whether reward variability could modulate the priority for reward and/or speed. Each stimulus subset (small and large) in the task was assigned a reward variability (high or low) and value (high or low), counterbalanced such that all combinations of value and variability were assigned across groups of participants. The results suggest that the variability of rewards does not modulate the preference for searching among the higher-value subset. Contrasted with involuntary attentional biases for cues associated with more variable reward outcomes, our findings suggest a possible dissociation between how reward variability influences voluntary and involuntary attentional control. Overall, our findings speak to the factors that motivate the choice of visual search strategy.

36.475 RELATING SELECTION HISTORY TO THE STRATEGIC CONTROL OF ATTENTION

Virginia K. Winston¹, Molly R. McKinney², Brian A. Anderson³; ¹Texas A&M University

Both prior learning and the choice of strategy influence the efficiency of behavior, including the allocation of attention. The present study examines the relationship between past experience and the strategic control of attention. Recently, the Adaptive Choice Visual Search task was used to test the effect of completing an inefficient search on the subsequent strategic control of attention. An experimental group experienced a single-target search phase, where a target randomly appeared in either the smaller or larger subset of potential target items (defined by color). Subsequently, they and a control group performed a two-target search, in which the choice to search for one of two potential targets in the smaller (optimal) subset yielded faster response times. The experimental group made more optimal choices in the two-target phase, even though the color distribution of search items was previously task-irrelevant, suggesting that a history of searching inefficiently motivated participants to search more optimally. The current study (N=52) explored the effects of selection history on the strategic control of attention. A Bias-Optimal group experienced a training phase with the target in the optimal subset 100% of the time, whereas a Bias-Suboptimal group experienced a similar training phase but with the target always in the suboptimal subset. Each group then performed a test phase with a free-choice two-target search. Both groups had high ($M = 81\%$) and significantly above chance ($p < .001$) proportion of optimal choices in the free-choice phase. Regardless of whether participants previously prioritized the smaller or larger subset during training, sensitivity to the color distribution of stimuli translated to an ability to flexibly leverage this distribution to the benefit of search performance rather than a tendency to inflexibly follow a practiced search routine, highlighting the flexible contribution of selection history in the strategic control of attention.

36.476 THE IMPACT OF DIGITAL DISTRACTIONS AND ATTENTIONAL SUPPRESSION ON ACADEMIC PERFORMANCE

Jennifer Trujillo¹ (jtrujillo4@csuchico.edu), Isabella F. Avalos², Bo-Yeong Won¹; ¹California State University, Chico, ²University of California, Riverside

Our environment is saturated with different types of distractions, yet digital distractions have emerged as some of the most prevalent in modern education, particularly for college students who are constantly interrupted by digital distractions while trying to focus on their academic workload. These distractions, ranging from smartphone notifications to multitasking on multiple devices, can detract from academic pursuits and lead to decreased academic performance. Our attention is limited and scarce, so the ability to suppress such distractions is crucial for academic success. This study aims to investigate the relationship between individual differences in engagement with digital distractions and performance on an attentional suppression task, as well as their impact on grade point average (GPA). Specifically, we assessed whether engaging in digital distractions, as measured by the Smartphone Distraction Scale (SDS), along with the ability to ignore distractors in favor of target stimuli, affect academic outcomes like GPA. We hypothesized that individuals who were more susceptible to digital distractions and who were less efficient in attentional suppression tasks would exhibit lower GPAs compared to their counterparts who manage distractions better. The analysis investigated how SDS scores and performance on an attentional suppression task collectively predict GPA among college students. Surprisingly, there was no significant relationship between SDS scores and performance on the attentional suppression task, but there was a significant result with SDS scores and GPA. While this study contributes to the growing literature on how digital distractions influence students' academic focus and offers guidance for supporting academic performance in the face of widespread digital interruptions, the findings also suggest that real-world distraction may be driven by more complex mechanisms than cognitive distractor suppression alone.

36.477 WORKING MEMORY CAPACITY MODULATES THE LINK BETWEEN LAPSES AND VOLUNTARY SWITCHING

Hyung-Bum Park¹, Monica Rosenberg², Edward Vogel²; ¹Washington University School of Medicine, ²University of Chicago

Attentional lapses are often interpreted as uniform failures of control, yet emerging accounts propose that they reflect competition among concurrently active goals. If lapses signal goal competition, they should prospectively predict voluntary task switches, and this relationship may depend on individual differences in working memory (WM) capacity. We tested this idea using a modified continuous performance task with bilateral face and scene stimuli. Participants freely chose whether to perform the face or scene task after each block, with the constraint that all 25 blocks of each task type had to be completed. Lapses in block n predicted the task chosen on block $n+1$,

and this prospective link depended strongly on WM capacity. Individuals with lower WM capacity were more likely to switch after lapse-prone blocks, whereas individuals with higher WM capacity tended to switch after successful blocks. Eye-tracking revealed matching patterns, with low-capacity participants directing early saccades toward the irrelevant distractor before switches and high-capacity participants showing the opposite tendency. Pupil dilation was elevated early in pre-switch blocks for low-capacity participants, consistent with heightened internal competition. In a forced-choice control session, externally imposed task sequences selectively reduced lapse rates for low-capacity individuals while leaving high-capacity performance unchanged. These findings indicate that lapses do not uniformly reflect disengagement or a single type of control failure. Instead, the relationship between lapses and voluntary switching depends on WM capacity, and is consistent with a goal-competition framework in which lapses mark dynamic shifts in the balance between competing goals.

This research was supported by Office of Naval Research Multidisciplinary University Research Initiatives (MURI) N00014-22-1-2123 to E.K.V. and M.D.R.

36.478 STEADFAST ATTENTIONAL BIASES: NO EFFECTS OF INCREASED PERCEPTUAL LOAD IN REJECTION SENSITIVITY

Emily A. Rancorn¹ (earancom@usf.edu), Bradley E. Buchanan¹, Liana Y. Thielecke¹, Bryce C. Romero¹, Emma Sonenblum¹, Connor Murray¹, Geoffrey F. Potts¹; ¹University of South Florida

Rejection Sensitivity (RS) is a personality trait linked to maladaptive responses to rejection and is present in various anxiety disorders, such as Social Anxiety Disorder (SAD). Individuals displaying RS, along with those diagnosed with SAD, show biases in selective attention and attentional control by primarily displaying bias towards salient social stimuli, such as faces. Perceptual Load Theory suggests that selective attention may be influenced by the amount of attentional resources needed for a task. Previous literature shows that the ability to selectively attend is reduced with heightened perceptual load. Perceptual load has been shown to affect this bias in SAD – however, previous studies have shown that perceptual load has no effect on this selective attention bias in Generalized Anxiety Disorder (GAD). Given RS is a predictor of SAD, perceptual load should impact this selective attention bias within RS. Event-related brain potentials (ERPs), described as insightful measures of cortical activity, have been shown to be representative of covert cognitive processes such as selective attention. The P3b is a posterior maximal positive deflection at 300-500 ms post-stimulus presentation and is described as an index of attention resource allocation. A sample of 86 undergraduate students were examined using a visual search task to explore electrophysiological indices (ERPs) of perceptual load effects on selective attention within individuals with high self-reported RS traits. Participants were asked to locate a target letter in low or high perceptual load conditions in which a task-irrelevant neutral or

scrambled facial image appeared at fixation. Bayesian linear mixed-effect models reveal that while perceptual load and RS impact P3b amplitudes independently, there is no interaction between RS and perceptual load. This suggests that trait RS attentional systems may map onto GAD more than they do SAD. Further literature should explore how trait RS's attentional systems are socially impacted.

MONDAY MORNING POSTERS IN BANYAN BREEZEWAY

Object recognition: Categories

**MONDAY, MAY 18, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY**

43.301 ON THE RELATIONSHIP BETWEEN DISTRIBUTED ASSOCIATION NETWORKS AND CATEGORY-PREFERRING VISUAL STREAMS IN THE INFEROTEMPORAL CORTEX

Joseph Salvo^{1,3}, Nathan Anderson^{2,3}, Rodrigo Braga³; ¹NIH, ²Brigham Young University, ³Northwestern University

The ventral visual stream in the inferotemporal cortex (ITC) comprises hierarchically-organized visual regions involved in recognition. Within ITC, contrast-based approaches reveal regions responding preferentially to images containing scenes (Epstein et al., 1998), faces (Kanwisher et al., 1997), or text (Cohen et al., 2002). These regions are arranged as streams along the long axis of the ITC's posterior half. Posterior ITC also includes a broad region of the canonical dorsal attention network (dATN), an association network implicated in top-down visual attention (Corbetta, 1998) definable using intrinsic ('resting-state') correlations. We asked whether the visual streams are confined to the canonical dATN, given the dATN's role in higher-order visual processing, or if the streams extend beyond the dATN, implying dATN involvement in earlier stages of visual processing. We collected 3T data from eight healthy adults across eight MRI sessions, allowing for individualized analyses, using a multi-echo sequence (Lynch et al., 2020). Participants performed tasks targeting language, theory of mind, mental imagery, and visual categories (scenes, faces, text). Large-scale networks were identified through functional connectivity using seed-based and data-driven clustering approaches (Braga & Buckner, 2017). The categories activated parallel ITC regions, arranged text, faces, scenes from lateral to medial. Face and scene streams clearly passed through the dATN, and the text stream did so in most participants, linking the dATN with mid-stage visual processing of multiple stimulus categories. All streams also overlapped with different large-scale networks in more anterior ITC: text regions aligned with the core language network, scenes aligned with a mental scene construction network (Default network A; DN-A), and faces tentatively aligned with the social cognition network (DN-B). Our

results suggest a conserved arrangement of parallel visual streams passing through the dATN to terminate in distinct association networks that are functionally related to the content of each visual stream.

43.302 DISSOCIABLE REPRESENTATIONS OF OBJECT CATEGORY AND STIMULUS FORMAT IN HUMAN VENTRAL TEMPORAL CORTEX

Jewelina Yao¹, Kalanit Grill-Spector¹; ¹Stanford University

A central goal of visual neuroscience is to understand how the brain extracts meaningful information from visual inputs ranging from natural scenes to symbols used for communication. Consistent with this goal, prior work in ventral temporal cortex (VTC) has examined the representations of both high-level category information and low-level stimulus properties. However, how category representations in VTC generalize across culturally invented visual formats, particularly those used for efficient communication (line drawings, emojis), remains unknown. Using fMRI, we measured brain responses in 9 adults to stimuli from 5 ecologically relevant categories (faces, words, manipulable objects, hands, and places) presented in 3 formats (line drawings, emojis, and photographs). We extracted multivoxel response patterns from anatomically defined lateral and medial VTC and examined category and format information encoded in distributed responses using representational similarity matrices (RSMs), multidimensional scaling (MDS), and cross-validated decoding. In lateral VTC, representational structure was dominated by category: RSMs showed high similarity within category across formats, MDS showed tight clustering by category, and category decoding was robust (~60-80% accuracy), exceeding format decoding (~50-60%). By contrast, medial VTC exhibited a representational structure dominated by format: RSMs showed high similarity within formats across categories, MDS revealed clustering by format, and format decoding (~60-70%) was more robust than category decoding (~50-60%). Interestingly, category and format contributed independently to representations in both VTC partitions, indicating separable information axes. Our results show that both object category (what the item is) and format category (how the item is depicted) are independently and differentially represented across lateral and medial VTC. Together, this study provides a new framework for probing how regions in high-level visual cortex support abstraction across culturally invented visual formats, and establishes a foundation for investigating how these representations might emerge with experience and schooling during development.

Funding: NIH Grant R01 EY023915

43.303 CHARACTERIZING CRITICAL BANDS AND ABSOLUTE EFFICIENCY OF NEURAL NETWORKS USING IDEAL OBSERVERS

Anne Peiris¹, Jaykishan Patel¹, Richard F. Murray¹; ¹York University

Deep neural networks (DNNs) have revolutionized computer vision and computational modelling, and they achieve human-like accuracy

in object classification tasks. However, because DNNs learn through training, little is understood about how they work. Spatial frequency is a key feature of visual stimuli, but which frequencies are critical for DNNs' performance? Previous work found that DNNs have a critical spatial frequency band 2–4 times wider than humans in an object classification task (Subramanian et al., 2023). This may be a shortcoming of DNNs, and they might perform better if they had narrower, human-like channels. Another possibility is that DNNs' channels are wider because they exploit more of the useful spatial frequencies in images. We used ideal observer analysis to test these hypotheses. We ran five ideal observers and 18 DNNs in a classification task using ImageNet's validation set (1000 categories, 50 samples each). We used a critical band masking paradigm to determine the most important spatial frequencies for object classification. We chose one representative DNN from each category of architectures available in PyTorch: one ResNet, one VGG, etc. Images were presented in filtered Gaussian noise, and object classification thresholds were measured at each noise frequency to identify critical bands. DNNs' critical bands were, on average, around two octaves wide, consistent with previous work. Interestingly, the ideal observers' critical bands were low-pass instead of band-pass, with a 1/f fall-off. Thus, DNNs do not exploit the entire range of informative spatial frequencies in natural images. Indeed, in a separate analysis, we found that object classification thresholds in white Gaussian noise were orders of magnitude lower for ideal observers than for DNNs, meaning that DNNs' absolute efficiency is extremely low ($\eta \approx 7 \times 10^{-5}$). These results demonstrate that ideal observer analysis can be highly informative for understanding fundamental properties of DNNs.

Funded by an Ontario Graduate Scholarship to AP and an NSERC Discovery Grant to RFM.

43.304 EXCITATION-INHIBITION BALANCED PRUNING YIELDS SPARSE, ACCURATE CATEGORY COMPUTING CIRCUITS WITH EMERGENT ANIMATE-INANIMATE ROUTES

Jeffery Andrade¹ (jandrade01@g.harvard.edu), Parisa Vaziri¹, George Alvarez¹; ¹Harvard University

The ventral stream maps patterns of light to untangled high-level category representations through hierarchical processing stages. What is the nature of the circuitry underlying this untangling process? We address this question in a "model organism," AlexNet, to identify computational motifs of category-computing circuits in a "white-box" system. Specifically, we ask which connections between layers must be preserved for correct classification and what the resulting minimal circuits reveal about shared wiring. Prior work extracted sparse category-selective circuits by pruning connections with low importance scores computed from forward activations and backward gradients (Hamblin et al., 2023). We improve this method using (i) "sign-aware pruning," which separately prunes positive and negative links to maintain excitation-inhibition balance, and (ii) a cross-layer pruning

scheme that uses a marginal-loss metric to allocate pruning across layers. These additions yield more consistent, sparser, and more accurate circuits than prior approaches. Dissecting 1000 circuits (one per ImageNet category), we obtain circuits that require 6% of the original network while being 2% more accurate than the unpruned network. Analyzing the circuit structure, we see a common macro-motif, retaining an average of ~93% of weights in layer 1, declining across the convolutional hierarchy to dramatic sparsity in the fully-connected layers (5.2% in layer 6). We find that shared wiring between categories declines across the hierarchy, where the median link appears in only 3/1000 circuits by layer 6. Considering the aggregate animate and inanimate circuits, we observe increasing separation through early convolutional stages and a bimodal distribution in later fully connected layers, with many links computing nearly exclusively animate or inanimate categories. Broadly, this method provides a computational approach toward quantifying the size, overlap, and distributedness of category circuits in deep neural network systems, and offers new experimental targets for predictions of competition, overlap, and priming effects in biological vision.

This work was supported by funding from the Kempner Institute and NSF-CAREER 7835425-01 (to Talia Konkle).

43.305 PERCEIVING ANIMACY IN OTHERWISE-IDENTICAL IMAGES

Tal Boger¹, Chaz Firestone¹; ¹Johns Hopkins University

Even when completely motionless, some objects look animate (e.g., dogs and elephants) while others don't (e.g., boots and sofas). A rich literature suggests that the mind automatically encodes this property, reporting striking effects of perceived animacy on visual attention and working memory. However, objects that differ in animacy tend to differ in many lower-level features (e.g., shape), and follow-up work has revealed that such confounds often account for these seemingly high-level effects. Thus, whether animacy itself drives visual processing remains unanswered and even controversial. Here, we take a new approach to this question by exploiting "visual anagrams": static images whose interpretations change radically with orientation. We used a diffusion model to generate such images in ways that varied animacy — e.g., a dog in one orientation and a boot when rotated, or an elephant in one orientation and a sofa when rotated. Each anagram image contains the exact same pixels in either orientation, such that animacy varies across the two interpretations while nearly all lower-level features remain constant. 9 experiments used this approach to demonstrate that animacy itself guides memory and attention. Experiments 1–2 found that changes to an object in a memory array were more detectable when they altered animacy (e.g., dog→sofa is more detectable than dog→elephant, even when the elephant and sofa are the very same image, just rotated). Experiments 3–6 found that animate targets in a search array were easier to find among inanimate distractors than among other animate distractors (and vice versa). Finally, Experiments 6–9 verified that differences in orientation (one of the only lower-level features uncontrolled by visual anagrams)

cannot account for these results; the effects disappeared when using silhouetted, blurred, and pixelated versions of the anagrams. Thus, visual processing extracts animacy itself, over and above its covarying lower-level features.

43.306 SEEING NEW CONNECTIONS: HOW INSIGHT TRANSFORMS VISUAL REPRESENTATIONS

Emily Van Hove¹, Hans Stuyck¹, Andrea Ivan Costantino¹, Kobe Desender¹, Hans Op de Beeck¹; ¹KU Leuven

Visual representations are often studied as stable constructs. Yet in a constantly changing visual world, the flexibility of these representations is essential to reorganize stored knowledge, and crucial to learning, creativity, problem-solving, and generalization. Nevertheless, the mechanisms underlying this dynamic reorganization remain poorly understood. In this project we study insight problem-solving, a cognitive process that requires restructuring, as a prime candidate of profound representational change in the visual system. Insight was induced through a newly developed visual version of the Remote Associates Task, previously mainly used in semantic literature, in which participants searched for a non-trivial association between three images. Each trial required integrating the representations of all images to find the solution. Behavioral representational spaces were assessed before and after the insight task, through a similarity sorting task. Using tools widely adopted in vision science, including representational geometry, RSA/PCA, and deep artificial neural network models, allowed us to understand the multidimensional similarity space and how it restructures as a consequence of insight. Results from behavioral experiments (n = 197) show a clear restructuring of behavioral similarity spaces succeeding the visual insight task. Initial category- and scene- based organization, shifts towards a geometry of insight-induced associations assessed through a multimodal encoding model. Nevertheless, the structure of the initial representational space remains partially preserved, suggesting that insight allows integration of new associations atop preexisting ones. These results provide a behavioral demonstration of flexibility in visual representations and a reference for preliminary neuroimaging results on the spatial dynamics of representational changes.

FWO

43.307 SIZE-DEPENDENCE OF OBJECT RECOGNITION IN NATURAL SCENES

Gabriel Yancy¹ (gmy225@nyu.edu), Ajay Subramanian¹, Najib Majaj¹, Denis Pelli¹, Jonathan Winawer¹; ¹New York University

Human visual object recognition relies on a narrow, 1.5-octave-wide filter (Solomon & Pelli, 1994 & Subramanian et al., 2023). For simple stimuli like isolated letters, the center frequency of the band used grows as the $\frac{2}{3}$ power of stroke frequency (Majaj et al., 2002, Vision Research). The $\frac{2}{3}$ power indicates that letter recognition is scale dependent: the observer relies on higher frequencies (in cycles per

letter) as object size increases. We extended the Subramanian et al. (2023, NeurlPS) noise-masking method to assess the scale dependence of natural object recognition. We presented 1,050 ImageNet images (16 categories) mixed with Gaussian noise at five strengths ($SD = 0, 0.02, 0.04, 0.08, 0.16$), filtered into seven log-spaced spatial frequency bands (0.2 to 14 cycles/image) to seven observers. Each stimulus was displayed for 200 ms through a square aperture of 3, 6, or 12 degrees. Participants performed a 16-way classification. Accuracy was modeled as a logistic function of noise level, with thresholds constrained to a Gaussian function of log spatial frequency. Results mirrored findings for letters: the recognition channel is about 1.5 octaves wide, and the center frequency follows a $\frac{2}{3}$ power law of stimulus size. This is evidence that the size dependence of object recognition is not restricted to simple isolated objects but extends to a diverse range of target types.

NEI core vision grant P30EY013079 / NEI grant R01EY033628 / NEI grant R01EY027401

43.308 THE EDGE STATISTICS OF DRAWINGS FROM AROUND THE WORLD

Tom Endejan¹, Linda Smith²; ¹Indiana University

The visual world is made up of a large variety of kinds of things – man-made and natural, living and nonliving, animate and inanimate, foods and tools, solids and substances. Growing research both in human neuroscience and vision indicates that people learn the predictive patterns of edges that distinguish these higher order classes. We asked 30 adults from 6 different locations around the world to draw 50 common things that would be known even by young children. The populations differ markedly in education and culture: From the Bora people in the amazon rain forest to metropolitan Japan. We measured the distributions within each sketch of line segment length, orientation, and symmetry across the medial axis. We find that the predictive edge patterns that broadly distinguish these higher order semantic categories are also generated by humans in their drawings independent of cultural background and level of education. For example, sketches of living objects exhibit lower proportions of long line segments and lower symmetry across the medial axis from sketches of non-living objects. Man-made sketches can be distinguished from natural by higher levels of horizontal and vertical segments and symmetry across the medial axis. The similarity in edge statistics between higher level semantic categories held despite the wide array of lower level sub-categories reported in the collected sketches. We also compared the edge statistics of our sketches to photographs of the same objects and found both selectivity and correspondences in the drawn images relative to photographs of their real world counterparts. The results strongly implicate universal representations of high-level predictive patterns and highlight the value of using drawings as a measure of human visual representation of high-level semantic categories.

The Cognitive Development Lab at Indiana University

43.309 VISIBILITY IS NOT PERFORMANCE: FEATURE-SPECIFIC DISSOCIATIONS IN VISUAL PROCESSING

Marjan Persuh¹, Coryanne Coryanne Mulvey¹; ¹Manhattan University

A longstanding debate concerns whether visual experience is genuinely rich or whether we consciously perceive only a few detailed objects at a time. To address this question, we used a repetition-detection paradigm with very small set sizes, deliberately staying within the commonly assumed limit of three to four consciously perceived objects. The idea was that if all items should be consciously available under such minimal load, then performance limitations would reveal something deeper about the relationship between awareness and the ability to act on visual information. In Experiment 1, participants viewed arrays containing two or three items drawn from different feature domains, color, orientation, or shape, and judged whether any item repeated. In Experiment 2, the same task was used but participants additionally rated the visibility of the display. Across both experiments, performance depended strongly on feature type: color was easiest to process, whereas orientation posed the greatest difficulty. Crucially, visibility did not decrease in parallel with performance. Even when detection became harder at larger set sizes or for more demanding features, participants continued to report similar levels of visual clarity. Most importantly, visibility and task performance were not interchangeable indicators of conscious perception. In fact, the two measures sometimes moved in opposite directions. For example, at the larger set size, performance was lowest for orientation, yet subjective visibility was actually higher for orientation than for shape. This feature-specific reversal represents a dissociation: the feature that participants saw more clearly was not the one they could use most effectively for the detection task. Together, the results provide new evidence that what we consciously experience and what we can reliably report are not the same, supporting a view of visual perception that is phenomenologically rich even when behavioral performance is limited.

43.310 CAN EURASIAN JAYS (GARRULUS GLANDARIUS) MAKE CAUSAL INFERENCES ABOUT WEIGHT?

Giuliana Bucci-Mansilla¹ (gbuccim1@jhu.edu), I. Hawkins^{2,3}, V. Ajuwon², N. S. Clayton¹, Jason Fischer^{4,1}, C. Krupenye¹; ¹Johns Hopkins University, ²University of Cambridge, Cambridge, UK, ³University of Oxford, Oxford, UK, ⁴University of Coimbra

Humans have an intuitive understanding of the physics of the world that allows them to make inferences about how a scene will unfold. While nonhuman animals may share some capacities for making inferences about the physical world, the underlying representations remain unclear. One open question is how animals represent the weight of objects and whether they understand that heavier objects produce greater impacts when dropped. To address this, we tested whether Eurasian Jays (*Garrulus glandarius*) can infer the identity of occluded objects based on how the objects affect the bearing of a balance beam. In our experiment, we hid two objects with different weights in cups and placed them on opposite sides of a balance beam.

We hypothesized that Eurasian Jays would use a representation of weight as a force to infer that the heavier object tilts the balance, allowing them to predict the desired object's location. Our design had two conditions: in Condition 1 the desired object was paired with a lighter object, causing it to be at the lower end of the beam, and in Condition 2, it was paired with a heavier object, making it end at the higher position. We found that jays differentiated between conditions: they selected the lower option more often when the desired object was the heavier of the two but chose according to chance when the desired object was the lighter of the two. These findings suggest that jays may have mental representations of weight that extend beyond inflexible responses to perceptual cues.

Johns Hopkins University, psychological and brain science; Animal behavioral society.

Object Recognition: Features, parts

**MONDAY, MAY 18, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY**

43.311 COLOR IS NOT IMPORTANT FOR OBJECT RECOGNITION, EVEN WITH LOW VISION

Nil Altinordu¹, Geoffrey Boynton², Ione Fine³; ¹University of Washington

Introduction: Recognizing objects in black and white photographs or movies is remarkably easy. Why then is there extensive neural hardware devoted to supporting color vision – what is color vision good for? One alternative to the 'fruit and berries' hypothesis is that trichromacy might compensate for optical deterioration as a result of aging in longer-lived primate species. Here we examine whether color information facilitates object recognition under simulated low vision conditions in which high frequency form information is absent. **Methods:** We measured object-identification performance in real-world scenes (Adobe Stock & BOiS database) presented in grayscale or color across a range of blur levels (0-8 diopters). Participants judged whether a cued object was present. Object presence/absence was randomized across trials, producing a 2 (color/grayscale) × 2 (object present/absent) × 9 (diopter level) within-subject design. In Experiment 1 (N=27) participants viewed scenes and were given unlimited time to search for the object and respond. In Experiment 2 (N=18) the scenes were masked after 2.5 seconds and participants were given an unlimited time to respond. **Results:** Performance was similar for both experiments. While there was a strong effect of blur, there was little difference in performance between color/grayscale conditions and no interaction between the presence/absence of color and the amount of blur. **Conclusions:** Color plays a surprisingly limited role in object recognition within complex, real-world scenes even under low vision conditions. It remains possible that we would find an interaction between color and blur for specific tasks based on nature-based

stimuli, such as picking out berries or fruit in a bush, or judging the health and emotion of a face. These results have important implications for sight-restoration technologies (e.g., retinal and cortical stimulation, and optogenetics): there may be little added utility in producing colored percepts.

Supported by R01EY014645

43.312 EFFICIENT ENCODING OF DYNAMIC VISUAL SCENES BASED ON ELEMENTARY 3D FEATURES

Serena Castellotti^{1,2}, Antonio Brau¹, Maria Michela Del Viva², Giovanni Punzi^{1,3}; ¹Department of Physics, University of Pisa, Italy, ²Department of Neurosciences, Psychology, Drug Research and Child Health (NEUROFARBA), University of Florence, Florence, Italy, ³INFN, Pisa, Italy

Previous work has suggested that, given the need to rapidly process large amounts of visual data, the visual system extracts bottom-up saliency maps of static scenes using a limited set of features. These edge- and bar-like "optimal" 2D features can be derived by applying constrained maximum-entropy principles to the frequency distribution of all possible features. It is interesting to ask the same question in the spatio-temporal domain, as motion information constitutes a crucial saliency cue in early visual analysis. However, extending the approach to real-life dynamic scenes leads to an exponential increase in the number of possible 3D features, making it too demanding to be performed with standard computing means. We address this challenge by using specialized big-data reduction algorithms (Floating Top-k) adapted for execution on FPGA devices. This new approach makes it computationally feasible to identify optimal 3D features by applying the constrained maximum-entropy method to a much larger space of potential features than previously possible. We created movie sketches using optimal 3D features (3×3pixels×3frames) extracted from a large video database and tested their effectiveness in a discrimination task. Observers' accuracy remained very high across a broad range of model parameters, supporting the robustness of the model in predicting salient motion features. As a control, we generated alternative sketches by selecting sets of 3D features from frequency ranges different from the constrained maximum-entropy selection, matching either its information content or its constraint values. In both cases, these alternative "non-optimal" selections yielded poorer discrimination performance, indicating that neither total information nor the constraints alone are sufficient to produce meaningful representations of moving scenes. Overall, these findings suggest that the visual system reduces complex dynamic inputs at an early processing stage by selecting a limited number of elementary features that maximize information transmission under computational constraints.

This project was funded by the European Union – Next Generation EU, in the context of the grant PRIN 2022 (Project: "Real time reconstruction of data from LHC experiments with a distributed FPGA system", Grant no. 2022Z3K93E, CUP: I53D23001540006).

43.313 LOW SPATIAL FREQUENCY AND THE HUMAN CHANNEL BIASES EMERGE IN NEURALLY ALIGNED MODELS BUT DO NOT EXPLAIN VISUAL ROBUSTNESS

Zhenan Shao¹, Tianyu Ren¹, Chengxiao Wang¹, Diane M. Beck¹;
¹University of Illinois, Urbana-Champaign

Despite matching or even surpassing human performance on object recognition, deep convolutional neural networks (DCNNs) are strikingly vulnerable to almost imperceptible image perturbations generated by adversarial attacks. Recent work shows that explicitly aligning DNNs to human neural representations improves adversarial robustness, raising the question of what mechanism underlies human robustness and drives those gains. One line of work has attributed human robustness to a reliance on low spatial frequencies (LSF), known to support holistic shape processing, whereas DCNNs overemphasize easily disrupted high-frequency details. Neural alignment could therefore be instilling LSF biases. However, recent work (Subramanian et al., 2023) revealed that humans instead rely on a narrow mid-frequency “human channel” for object recognition. Here we test whether neural alignment improves robustness by inducing spatial frequency (SF) biases (LSF or the human channel), and whether such a bias is sufficient for robustness. We first used adaptive adversarial attacks to probe SF reliance of DCNNs aligned to successive regions of the human ventral visual stream (VVS), which previously showed hierarchical robustness gains. We show that neural alignment produces systematic SF reliance changes: DCNNs aligned to higher-level VVS regions increases reliance on both LSF and the human channel while reducing reliance on high frequencies. We then test the sufficiency of SF bias by imposing targeted LSF or human-channel biases during training, using both SF filtering and a noise-based augmentation method that suppresses specific SF bands with controlled strength. Across all conditions, these biases fail to improve adversarial robustness or produce more human-like representational geometry. Our findings indicate that while neural alignment induces both LSF and human-channel preferences, neither alone suffices for inducing benefits observed with neural alignment. Therefore, such SF preferences may emerge from broader computational goals of the human visual system rather than serving as the primary mechanism that confers adversarial robustness.

43.314 PROBING THE LANGUAGE OF PERCEPTION: OBJECT FILES DISTINGUISH POP-OUT FROM ANCILLARY FEATURES

Avery Caulfield, Justin Halberda¹, E. J. Green¹; ¹Johns Hopkins University

Abstract: An object file is a perceptual representation that sustains reference to an object over time while retaining information about the object’s currently and recently perceived features (Kahneman et al. 1992). Here we investigate whether object files incorporate a distinction between individuating and non-individuating features of objects, and whether this distinction governs how object files interface

with the rest of the mind. In one series of experiments, subjects detected a pop-out target among homogenous distractors and then made a memory judgment about that object’s features (with the precision of each feature staircased across trials: the pop-out individuating feature and the ancillary non-individuating feature). Findings suggest that, surprisingly, participants demonstrate less precision for the individuating feature than for the non-individuating feature. We hypothesize that individuating and non-individuating features play distinct computational roles within object files. The representation of the pop-out feature functions to pick out the object to be attended, while the representation of the non-individuating feature is prioritized for higher-precision processing in visual working memory. This asymmetry between individuating and non-individuating features was mirrored in another series of experiments where subjects were given a pop-out display and were asked which English sentence best described the array. Subjects preferred descriptions placing the pop-out feature within the scope of the demonstrative and the ancillary feature in the predicative clause. For example, when shown a red tilted bar among green tilted bars, subjects preferred the description “That red bar is tilted” over “That tilted bar is red.” This pattern suggests a correspondence between genuinely discursive representations (i.e., language) and a visual language of thought.

43.315 BEHAVIORAL AND NEURAL SIGNATURES OF REPRESENTATIONAL RELEVANCE FOR OBJECT CONCEPTS

Tonghe Zhuang^{1,2,3*}, Selma Akarsu^{1*}, Malin Stymal^{1,2,3}, Martin N. Hebart^{1,2,3}; ¹Computational Cognitive Neuroscience and Quantitative Psychiatry, Department of Medicine, Justus Liebig University Giessen, Germany, ²Vision and Computational Cognition Group, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, ³Center for Mind, Brain and Behavior, Universities of Marburg, Giessen, and Darmstadt, Germany * equal contribution

Objects in everyday life differ in how much they matter for our cognition and behavior. For example, we may more often look at rocks on the ground than babies, yet in our mental and neural representation of objects, a baby will matter more, highlighting that occurrence frequency is not sufficient to capture the relevance of objects. We term this graded importance representational relevance: the degree to which an object representation is likely to matter directly or indirectly for downstream cognition and behavior. Here we aim at operationalizing representational relevance for vision and cognitive science. We hypothesize that representational relevance will be related with, yet not identical to traditional measures of mental access, including lexical decision times, familiarity, or word frequency. To measure representational relevance (RR), we developed a rank-ordering “Family Feud”-style task in which participants arranged object words according to how often they believed other people would spontaneously name them (e.g., which is mentioned more frequently, “lion” or “burrito”?). This yielded an RR score for each concept. Split-half analyses over 1,000 random partitions showed that RR scores

were highly reliable (Spearman–Brown corrected reliability = .957). RR showed strong good construct validity, correlating with lexical decision times (LDT, $r = .706$), familiarity ($r = -.558$), and frequency ($r = -.685$). To externally validate RR, we linked scores for 601 object concepts to a large-scale fMRI dataset (Hebart et al., 2023). Voxel encoding models revealed reliable RR effects in high-level occipitotemporal cortex, and RR scores were more predictive than word frequency and familiarity and led to comparable effects of LDTs, underscoring the usefulness of RR. Together, this work introduces representational relevance as a construct and highlights its potential usefulness for improved understanding of the relative weighting of objects beyond their frequency of occurrence.

MNH was supported by the ERC Starting Grant COREDIM, a LOEWE Start Professorship of the Hessian Ministry of Higher Education, Research, Science and the Arts, and the Deutsche Forschungsgemeinschaft, under the Collaborative Research Center “Cardinal Mechanisms of Perception” and Germany’s Excellence Strategy.

43.316 AUTOPSYCH: AUTOMATED PSYCHOPHYSICS FOR INTERPRETABILITY AND DIVERSITY BENCHMARKING

Habon Issa¹, Sunny Liu¹, David Klindt¹; ¹Cold Spring Harbor Laboratory

Interpreting neural representations is a fundamental challenge for neuroscience. This is complicated by the fact that single neurons respond to multiple independent stimuli (i.e., they exhibit mixed-selective or superposed representations). In the AI mechanistic interpretability field, population coding methods including sparse autoencoders (SAEs) have successfully lifted features out of superposition, revealing the hidden feature representations of large neural networks. However, there are no rigorous quantitative metrics to measure the success of SAE architectures against existing alternatives. Human psychophysics experiments are considered the gold-standard method to benchmark population coding methods, but the low scalability of this method presents a major bottleneck to feature evaluation. While recent work automates this using large language models (LLMs), these methods are limited to textually-describable concepts and can be difficult to generalize. We introduce AutoPsych, a framework for conducting automated psychophysics experiments where our approach replaces human subjects with perceptual similarity models (e.g., DreamSim and LPIPS for vision models) that are pretrained to mimic human psychophysics judgments. Using this framework, we propose two novel, quantitative benchmarks. The first, an Odd-One-Out (O3) task, measures the interpretability of individual neurons or SAE latents by assessing the perceptual consistency of their maximally activating inputs. The second, cross-O3 (xO3), extends this to measure the diversity of a set of units, providing a proxy for how many distinct concepts they cover. Compared to existing methods, AutoPsych requires no LLM-generated textual descriptions, and is readily adaptable to diverse data types including images, text,

and biological data. This provides a scalable and generalizable solution for rigorously evaluating progress in neural network interpretability.

This work was performed with assistance from the US National Institutes of Health Grant S10OD028632-01

Face and Body Perception: Social cognition 2

**MONDAY, MAY 18, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY**

43.317 ATTRACTIVENESS AT FIRST SIGHT: INFLUENCES OF HIGHER-LEVEL COGNITIVE PROCESSING ON FACIAL ATTRACTIVENESS JUDGEMENTS

Olivia S. Cheung^{1,2,3} (olivia.cheung@uaeu.ac.ae), Ayesha Ahmed², Treedom Beiyin Zhang^{2,3}; ¹United Arab Emirates University, ²New York University Abu Dhabi, ³Center for Brain and Health, NYUAD Research Institute

At a glance of a face, observers make consistent inferences of social traits. Notably, facial attractiveness judgment is thought to be unambiguously based on visual appearance, unlike other traits (e.g., competence). However, it remains underexplored whether first impressions of facial attractiveness might be influenced by higher-level processing. To examine whether such judgments depend primarily on lower-level visual processing of physical appearance, or higher-level processing of social-emotional cues, we presented images of a total of 80 unfamiliar faces briefly (100ms), with 20 faces each for two race (Black/White) and gender (female/male) conditions. First, we tested whether faces may be rated more attractive due to visual (e.g., symmetry) or emotional cues (e.g., genuineness). As emotional expressions lead to a mild degree of facial asymmetry, chimeric faces created from combining either the left or right half of a face with their mirror images could appear to lack emotional genuineness (Kownar, 1996). Participants (N=56) from multicultural backgrounds rated original and chimeric faces. Attractiveness ratings were higher for original than chimeric faces, regardless of face race or gender ($p < 0.02$). Second, we investigated whether consistency of attractiveness ratings remains stable among observers with differential implicit social biases for faces from various social groups, as we hypothesized that attractiveness ratings should be consistent when only visual cues are utilized. Ratings for Black and White faces between observers who were either positively biased towards Black or White people, and ratings for female and male faces between observers who were either positively biased towards female or male, were correlated. Attractiveness ratings were consistently correlated for White and Black faces ($r = 0.72$, $r = 0.70$; $p = 0.43$), but were more correlated for female than male faces ($r = 0.68$, $r = 0.38$; $p = 0.03$). Together, these results suggest that higher-level social-emotional

information, such as emotional genuineness or implicit biases, is processed rapidly and is crucial for evaluating facial attractiveness.

This work was supported by a NYU Abu Dhabi faculty grant (AD174) and by NYUAD Research Institute (CG012) to O.S.C. and a NYU Abu Dhabi capstone grant to A.A.

43.318 PERCEIVED TRUSTWORTHINESS BIASES MEMORY FOR EMOTIONAL EXPRESSIONS

Chunliang Yang¹ (cy36@illinois.edu), Stefan Uddenberg¹; ¹University of Illinois Urbana-Champaign

People rapidly and automatically form trait impressions from faces, and subtle changes in facial traits influence perceived emotional valence. Indeed, the same emotional expression is judged as more positive on trustworthy-looking faces as opposed to untrustworthy-looking faces, indicating a tight coupling between trait impressions and emotion perception (Oosterhof & Todorov, 2009). Whether facial traits also distort memory for emotional expressions, however, remains unknown. Here we tested whether invariant facial morphology biases the remembered valence of a briefly viewed expression. First, we synthesized 60 hyper-realistic, emotionally neutral identities (demographically homogenous, to limit potential other-race effects in memory). These identities were designed to be especially high or low in perceived trustworthiness (30 each), using a trait impression model trained on over one million human judgments (Peterson et al., 2022). Using open source emotion editing tools, we then created a mean luminance-matched emotional continuum (happy-neutral-sad) for each identity. In the main experiment, observers (N=100) engaged in a continuous memory reproduction task, first viewing a single face displaying a given expression from the continuum, then reproducing the remembered expression (after a brief delay) by adjusting a test face using a circular slider. Our primary measure of interest was “memory bias”, operationalized as the signed difference between the reproduced and originally viewed expression as a function of the identity’s trustworthiness level (high or low). Across both conditions, on average, observers reproduced faces as happier than they originally were. However, this “happiness” bias in memory was stronger for trustworthy-looking faces than untrustworthy-looking ones. These results suggest that enduring trait-like signals in facial morphology shape not only online perception of emotion, but also the encoding and/or reconstruction of facial expressions in memory, biasing what we remember about how others felt in the moment.

43.319 FALSE MEMORY FOR FACES: EFFECTS OF EMOTION AND AGE USING THE DEESE-ROEDIGER- MCDERMOTT PARADIGM

Rui Hu¹, Isabelle Boutet¹, Hanna Chainay²; ¹University of Ottawa, ²Université Lumière Lyon 2

False memory, the recall of events that never occurred or the misremembering of details, has been extensively studied using semantic stimuli within the Deese–Roediger–McDermott (DRM)

paradigm. However, far less is known about false memory for non-semantic and perceptually meaningful stimuli such as faces. This project investigates how perceptual categorization based on social cues influences the retrieval of face representations in memory. Using a modified DRM paradigm, we created lists composed of faces sharing a perceptual category to test whether categorically related critical items (CIs) elicit stronger false recognition than unrelated distractors. Methods: Participants (target N= 34; aged 17–26) were tested in two blocks consisting of one Study and one Recognition phase. For the study phase of each block, participants were shown 25 identities from the FACES database, each morphed along two continua: emotion (happy-angry) and age (young-old). All faces belonged to one category. During recognition, participants were shown previously studied faces (old items), new faces from the opposite category (unrelated distractors), and new faces from the same category as the studied items (CIs). CIs were designed to elicit false recognition due to perceptual similarity and shared category membership. Participants judged via key press whether each face was “old” or “new.” Results: Preliminary analyses indicate that categorically related CIs elicit higher false recognition rates than unrelated distractors, as reflected by a greater percentage of “old” responses. This pattern suggests that participants were influenced by the overall category of the studied faces. CIs belonged to the same category as the studied items, they felt familiar, which increased the likelihood of false “old” responses. By extending this paradigm to socially meaningful facial cues, this project advances understanding of false recognition of faces and highlights conditions under which errors are most likely. These findings have implications for applied settings, including contexts involving eyewitness identification.

Funding: Supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (2022-03998) to IB.

43.320 AVERAGE OR UNIQUE? DISTINCTIVENESS ENCODING FACILITATES FACE RECOGNITION ACROSS VIEWING CONDITIONS

Gray Koudinova¹, Bozana Meinhardt-Injac², Isabelle Boutet¹; ¹University of Ottawa, Ontario, Canada, ²Catholic University of Applied Social Sciences, Berlin, Germany

Introduction: Within memory research, encoding strategies that encourage participants to think about how distinctive a semantic item is can facilitate said item’s recognition (e.g. Hertzog et al., 2018). To our knowledge, only two studies have shown a similar benefit for face recognition (Winograd, 1981; Carr et al., 2015). However, these studies used identical images at learning and test, which (1) may promote image-based rather than identity-based recognition and (2) does not reflect how familiar faces vary across real-world viewing conditions (Burton, 2016). Furthermore, other encoding strategies (e.g. social encoding) have been shown to improve face recognition even when the image format changes (e.g., Schwartz & Yovel, 2019). The present study therefore examined whether a distinctiveness encoding strategy also enhances recognition of faces presented in

different formats at test. Methods: 43 undergraduate students (17-35 y/o) completed the study. In the Encoding phase, participants were shown photos of faces (each shown 5 times). Each face was accompanied by rating questions about either distinctiveness, physical features, or social traits. In the Testing phase, participants were asked to discriminate learned faces from new faces. All faces were presented twice: once with a lighting change and once with an orientation change. Analyses of recognition performance at test (d') and reaction times during the rating task at encoding were performed. Results: Participants recognized faces better when they were encoded while rating distinctiveness and social traits compared to physical features. Participants took the longest to rate distinctiveness during the rating task at encoding. Conclusion: This study suggests that performing a distinctiveness rating task during encoding contributes to the formation of invariant face representations, as the recognition benefit extended to different testing image formats. As such, adopting a distinctiveness-based encoding strategy could be a practical support for face recognition in everyday interactions.

Supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (2022-03998) to IB.

43.321 APPROACH-AVOIDANCE BEHAVIOR SHAPES PERCEPTION AND MEMORY FOR FACIAL EXPRESSIONS

Arijit De¹, William Cunningham^{2,3}, Adrian Nestor¹; ¹University of Toronto, Scarborough, ²University of Toronto, ³Vector Institute

People typically approach positive facial expressions and avoid negative ones, but it is unclear whether learned approach-avoidance mappings also bias how expressions are perceived and remembered. We therefore tested whether such distortions depend on expression valence (positive vs. negative) and/or on the modality of face processing (perception vs. memory). In an online study, adults ($N = 214$) learned either a congruent rule (approach positive, avoid negative faces) or an incongruent rule (approach negative, avoid positive faces), reinforced with probabilistic monetary feedback. Face images were parametrically generated with EmoStyle (Azari et al., 2024), allowing us to vary valence across a full continuum while holding arousal constant, and preserving facial identity. After learning, we assessed memory with a continuous “valence wheel” task, in which participants tried to match the valence of previously seen faces. We then assessed perception with Self-Assessment Manikin ratings of valence and arousal for the same stimuli, providing a complementary read-out of emotional evaluation. Across conditions and expressions, emotional valence was remembered reliably above chance, with better memory for positive than negative expressions. Critically, positive faces learned under the incongruent rule were later recalled as more negative than those learned under the congruent rule while negative faces were remembered similarly across conditions. Perceptual ratings showed a complementary pattern where valence differed between conditions only for negative faces, which were perceived as less negative in the incongruent than in the congruent condition, with no reliable effect for positive faces. These findings suggest that

approach-avoidance behavior distorts emotional face representations in a modality- and valence-dependent manner, altering both how expressions are perceived and how they are remembered.

Natural Sciences and Engineering Research Council of Canada

43.322 THE NEURAL BASIS OF PERSONAL SPACE AND RESPONSES TO NEARBY SOCIAL THREATS: AN FMRI STUDY

Louis Vinke^{1,2}, Francesca De Marneffe¹, Suhani Dheer¹, Baktash Babadi^{1,2}, Lee Cohen^{1,2}, Dan Fulford^{3,4}, Roger Tootell^{2,5,6}, Daphne Holt^{1,2,6}; ¹Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA, ²Harvard Medical School, Boston, MA, USA, ³Sargent College of Health and Rehabilitation Sciences, Boston University, Boston, MA, USA, ⁴Psychological and Brain Sciences, Boston University, Boston, MA, USA, ⁵Department of Radiology, Massachusetts General Hospital, Boston, MA, USA, ⁶Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital

The regulation and representation of personal space, defined as the space near the body into which others cannot intrude without eliciting discomfort, is a basic form of non-verbal social communication. Non-human primate neurophysiological studies and human fMRI studies have identified a frontoparietal-subcortical network of brain areas that responds to intrusions into this space. However, the modulation of this network when responding to threatening versus rewarding stimuli remains unclear. Thus, here we investigated whether activation within this personal space monitoring network is sensitive to the emotional valence of the intruding stimulus. 3T fMRI data were collected for 60 healthy control individuals (mean age: 35.1 yrs; 44 females) while they viewed 3D face images which appeared to move towards or away from the subject and displayed one of three emotional expressions: neutral, angry, or happy. Significant activation to the approach vs. withdrawal conditions (the “looming” contrast) was observed within independently-defined personal space network areas (superior parietal, inferior parietal sulcus, and ventral/dorsal premotor cortices), with looming activation absent within early visual areas. When comparing responses to angry, happy and neutral face images, significant differences were observed within the ventral premotor area, with angry faces eliciting a stronger looming response compared to happy or neutral faces. The ventral premotor (vPM) area is particularly sensitive to the valence of personal space intrusions. The vPM may be recruited during preparation of defensive actions that aim to protect the face and body from incoming threats, as vPM stimulation has been shown to elicit defensive movements. Thus these findings are consistent with prior evidence for a central role of this neural system in the protection of the body from danger, and also have implications for investigations of clinical populations who maintain abnormal interpersonal distances while also having deficits in perception of non-verbal social cues (e.g., facial expressions).

43.323 AUDITORY INFORMATION IMPACTS FACE PERCEPTION IN AN OPPOSING AFTEREFFECT PARADIGM

Xueqi Ao¹ (aox@mcmaster.ca), M.D. Rutherford¹; ¹McMaster University

Introduction: Category-contingent aftereffects occur when individuals adapt to distortions in opposite directions for two socially meaningful categories. Prior work has shown that social labels such as religious identity or food preferences delivered in the auditory domain can generate opposing aftereffects. We tested whether race-related auditory cues impact adaptation ambiguous faces randomly assigned to East Asian versus Caucasian categories. **Methods:** Participants (N = 83) completed a three-phase aftereffect paradigm (pre-adaptation, adaptation, post-adaptation) and were randomly assigned to one of four conditions, varying by audio type (explicit race vs. control) and distortion direction (East Asian expanded vs. Caucasian expanded). Stimuli were human-validated AI-generated racially ambiguous faces and race-related word lists. In the explicit condition, faces were paired with race-revealing auditory information (e.g., culturally related names, foods, hobbies, countries). In the control condition, the same faces were paired with neutral, non-racial information. In the pre- and post-adaptation, participants selected their preference between faces pairs that were slightly expanded and slightly contracted, where in the adaptation phase, participants saw heavily distorted faces paired with auditory information. Results showed that ambiguous faces paired with East Asian labels produced significantly different aftereffects compared to the same faces without race information ($t(38.96) = -2.23$, $p = .031$; 95% CI [-3.52, -0.17]). Importantly, this effect was specific to the East Asian expanded condition, highlighting the asymmetric influence of racial cues. Even more compelling, significant opposing aftereffects emerged in the East Asian expanded/Caucasian contracted condition, where preferences shifted in opposite directions depending on the assigned racial label ($t(19) = -2.10$, $p = .04$; 95% CI [-3.09, -0.01]). **Conclusion:** Explicit racial auditory cues enhance the social meaning of racially ambiguous faces, enabling category-specific adaptation and reshaping perceptual representations of race. These findings advance our understanding of how social categories impact the visual perception of faces.

This research was funded by a NSERC - Natural Sciences and Engineering Research Council grant to MDR

43.324 CONTROLLED LEARNING HISTORIES WITH ARTIFICIAL RACES INDUCE AN OTHER-RACE-LIKE RECOGNITION EFFECT

Emmanuella Sanchez¹, Emily Martin¹, Fabian Soto¹; ¹Florida International University, Miami, FL

Human face perception is shaped by experience, and one of its most robust phenomena is the Other-Race Effect (ORE)—superior recognition of own-race faces relative to other-race faces. This asymmetry is typically attributed to greater experience individuating own-race faces, though experience with categorizing other-race faces may also contribute. Experimental tests that can causally isolate these learning components remain limited, in part because natural face categories confound perceptual and social factors. The present study introduces a controlled training paradigm using parametrically generated “alien” races, constructed from anthropometric principles to mirror the structural variability found across natural human groups, to test how specific types of learning give rise to ORE-like recognition patterns. Participants were randomly assigned to one of three learning histories: identity training, in which they learned to identify and name 25 faces from Race A; categorization training, in which they learned to classify faces as belonging or not belonging to Race B; or combined training, in which both learning tasks were completed with different races. After training, recognition memory was assessed for all races to determine how each learning history shaped participants’ ability to individuate faces. The alien races reliably induced an ORE-like pattern: participants in the combined-training condition showed enhanced recognition for the race they had learned to identify and reduced recognition for the race they had learned to categorize, demonstrating that distinct learning experiences can drive asymmetric recognition performance even in the absence of real-world racial categories. Identity-only and categorization-only groups showed patterns consistent with these learning contingencies, though the combined-training condition provides the clearest demonstration of opposing effects within the same observers. These results establish that controlled, type-specific experience with artificial races is sufficient to induce an ORE-like recognition asymmetry, providing a rigorous experimental framework for testing how different learning histories shape the structure of face representations.

This research was funded by NSF grant 2319234 awarded to Fabian Soto

43.325 LEARNING HISTORY MODULATES RECOGNITION BUT PRODUCES LIMITED CHANGES IN FEATURE USE IN AN ALIEN OTHER-RACE EFFECT PARADIGM

Ula McCarthy¹ (ulamccarthy24@gmail.com), Emily Martin¹, Emmanuella Sanchez¹, Fabian Soto¹; ¹Florida International University

The Other-Race Effect (ORE)—the tendency to individuate own-race faces more accurately than other-race faces—is often attributed to differences in perceptual experience that shape the dimensions observers rely on when encoding facial structure. Understanding how early experience modifies the representation of faces from different races requires stimuli with precise control over three-dimensional facial morphology and analytic tools sensitive to subtle shifts in perceived similarity. This study uses parametrically generated “alien” races, constructed from anthropometric principles to capture structural variation found in natural human groups, to test how different learning

histories influence the information observers extract when identifying a face. Participants experienced one of three training histories: identity learning, in which they learned to identify and name 25 faces from Race A; categorization learning, in which they learned to categorize faces as belonging or not belonging to Race B; or no training for Race C. Before the reverse-correlation task, all participants were familiarized with a single target identity to isolate the morphological information supporting identity judgments across conditions. Participants then completed a reverse-correlation identity task in which, across 500 trials per race, they selected which of two randomly perturbed faces more closely resembled the target. Individual classification vectors estimated each parameter's contribution to perceived identity similarity. Permutation-based cluster tests were applied to determine which parameters significantly contributed to each race-specific template and to evaluate contrasts across learning histories. Across all races and conditions, participants produced robust, highly structured templates that deviated strongly from noise, demonstrating reliable recovery of the information guiding identity judgments. However, contrasts between identity-trained, category-trained, and control templates revealed minimal feature-level differences across conditions, despite clear recognition-memory differences resembling an ORE. These findings indicate that distinct learning histories can modulate recognition performance without producing large shifts in the specific facial features observers rely on to represent identity.

This research was funded by NSF grant 2319234 awarded to Fabian Soto.

43.326 GAZE BEHAVIOURS ASSOCIATED WITH SOCIAL AND PERCEPTUAL ENCODING OF FACES

Steven Carton¹ (scart079@uottawa.ca), Arda Erbayav¹, Isabelle Boutet¹; ¹University of Ottawa

Background: When viewing a visual scene, our eyes continually scan and fixate on different points. Memory for familiar face representations is never purely perceptual, as our memory of others is always linked to social context. Previous research has observed that new faces are better learned when participants are instructed to judge the new faces with regard to social information (e.g., how honest does this person look?), as compared to perceptual information (e.g., how round is this person's face?) (Schwartz & Yovel, 2019; Winograd, 1981). However, it is unknown if differences in patterns of eye movement might explain the social encoding advantage, or whether this advantage persists over a delay. Methods: Participants (N = 36; aged 17–25) learned new faces with social, perceptual, or control (i.e., no judgement) encoding instructions. Following this, two recognition tests were performed — immediately and delayed (24–72 hours). Recognition of target faces was measured in viewpoint and lighting change conditions. Eye movements were tracked and recorded throughout encoding and recognition phases. Results: Socially encoded faces were recognized significantly better than perceptual or control encoding, for same- and different-viewpoint faces, at immediate and delayed recognition. We

observed a trend for different encoding strategies having different scan paths. Additionally, as a face became more familiar over encoding trials, fixations to the face significantly decreased. Finally, encoding and recognition eye movement patterns were different, with significantly more fixations to the eyes at recognition. Conclusion: Memory for human faces relies on the social–cognitive processes used during encoding, as demonstrated by our finding that social encoding improves face recognition both immediately and after a delay. Gaze behaviours differ as a function of both the encoding strategy used and the stage of processing, with participants paying more attention to the eyes during retrieval.

Funding: Supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (2022-03998) to Dr. Isabelle Boutet.

Face and Body Perception: Individual differences

**MONDAY, MAY 18, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY**

43.327 THE STRENGTH OF FACE PAREIDOLIA VARIES ACROSS IMAGES, INDIVIDUALS, AND COMPUTATIONAL FACE DETECTION MODELS

Saivydas Villani¹ (s.villani@soton.ac.uk), Guido Maiello¹; ¹University of Southampton

We've all had the experience of pointing out a "face" in a plug socket, cloud, or car bumper, only to be met with a puzzled look and a polite "I don't see it." This everyday disagreement raises a deeper question: does the strength of pareidolic impressions vary reliably across images and observers? And what does this reveal about the mechanisms that drive illusory face perception? To answer this, we curated a subset of the FacesInThings dataset, a pareidolia image set with Easy/Medium/Hard human annotations (from a single rater) for the difficulty of perceiving the illusory face (Hamilton et al., ECCV, 377–395). Our final test set contained 25 stimuli per difficulty level, plus 25 independent control images. We presented these stimuli to seven human observers and two computational face detection models (RetinaFace with MobileNet and ResNet-50 backbones) fine-tuned on the FacesInThings training set to identify pareidolic faces. Humans and models were tasked with detecting and localizing pareidolic faces and rated the "faciness" of the pareidolic images. FacesInThings difficulty levels systematically affected human behaviour. The probability of detecting a pareidolic face as well as the "faciness" ratings decreased from Easy to Hard ($p < .001$), while reaction times increased ($p < .001$). Face detection models also followed this pattern, correctly identifying the pareidolic faces with decreasing confidence as difficulty level increased ($p < .001$ for both models). Additionally, the strength of pareidolia (the "faciness" ratings) varied across individual

images (average between-participant agreement $\bar{r}=0.70$, $p<.001$) and participants (split-half reliability $r=.95$, $p<.001$). These data demonstrate that the strength of pareidolic impressions varies markedly and reliably across stimuli and individuals. Further, computational face detection models can also be trained to exhibit pareidolia-like responses, serving as a scalable, objective measure of pareidolic sensitivity, while also allowing us to model individual- and image-level differences in illusory face perception.

This research was funded by a South Coast DTP studentship awarded to Saivydas Villani

43.328 AGE-RELATED DIFFERENCE IN PERIPHERAL INFORMATION BENEFITS FOR FACE PERCEPTION

M. Eric Cui^{1,2}, Rojin Firooz^{2,3}, Janet Pham¹, Yifei Wang¹, Sarah Creighton¹, Björn Herrmann^{1,2}, Allison Sekuler^{1,2,4}; ¹Rotman Research Institute, Baycrest Academy for Research and Education, ²Department of Psychology, University of Toronto, ³Department of Psychology, University of Toronto Scarborough, ⁴Department of Psychology, Neuroscience & Behaviour, McMaster University

While younger adults typically fixate on the eyes for face identification, many older adults tend to fixate on the nose, which generally contains less information for identification. It remains unclear whether the age-related fixation differences reflect covert sampling mechanisms, such as differential reliance on peripheral information. The current study addresses this gap. We recruited 76 younger ($M = 26.2$ years) and 34 older adults ($M = 63.2$ years) to complete a face identification task under two gaze-contingent conditions that simulated either tunnel vision (the Window condition) or a blindspot (the Mask condition). We varied the size of the window and mask (range of diameters = $[1.5^\circ - 8.5^\circ]$) to control for available information in central and peripheral regions. Overall, performance was much higher in the Mask than the Window condition. As more information became available (smaller Mask or larger Window), accuracy improved for both age groups, but at different rates: the age gap in performance slightly reduced in the Mask condition, but moderately widened in the Window condition. Together, these findings suggest that older adults are less efficient and flexible in integrating information from the peripheral regions for face identification. Our ongoing analyses are focused on using eye-movement data to explain the mechanisms underlying the behavioral results.

43.329 INDIVIDUAL STABILITY OF SOCIAL VIEWING STRATEGIES ACROSS REPEATED CONVERSATIONS

Soazig Casteau¹ (soazig.casteau@durham.ac.uk), Alice Nevard¹, Ulrik Beierholm¹, Holger Wiese¹; ¹Durham University

Human social interaction is highly dependent on effective visual attention, yet current eye-tracking research largely relies on static stimuli or single-session analyses. This limits our understanding of whether individual social viewing patterns are consistent, persistent strategies that stay over time, or if they are dependent on short-term

context and individual adaptation. To address this, we tested whether individual viewing strategies remain invariable across multiple, naturalistic social interactions. We (1) used Hidden Markov Models (HMMs) to identify probabilistic social viewing strategies across multiple conversational sessions; and (2) determined if the adoption of a specific HMM-defined strategy is consistent over time and predictive of behavioural outcomes. Forty participants engaged in a free-form social conversation with an unfamiliar partner (to-be-learned face) during four 10-minute-long sessions spaced over a one-week period. Eye movements were recorded using mobile eye-tracking, and face recognition accuracy of the partner was measured after each session to assess learning. Consistent with previous findings, the HMM analysis revealed two probabilistic viewing strategies: Analytic and Holistic. Crucially, our results showed that each participant's dominant viewing strategy remained consistent across sessions, demonstrating stable individual differences independent of session progression. While participants improved their face recognition accuracy over time, overall rate of learning was similar between both strategy groups. Participants predominantly adopting the Holistic strategy showed higher initial face recognition accuracy, but this early difference converged quickly across sessions, and strategy type was ultimately not a reliable predictor of long-term recognition performance. Taken together, our findings establish that an individual's social viewing strategy is a highly persistent characteristic independent of time and accumulating experience. Although an initial performance difference was observed early in the task, this consistent strategy wasn't related to long-term recognition performance and the overall rate of learning, suggesting a limited functional role in the context of this specific naturalistic social task.

This research has been supported by an ESRC grant (ES/X014657/1)

43.330 DIFFERENTIAL EFFECTS OF FORM INFORMATION AVAILABILITY AND ACTION FAMILIARITY ON EYE MOVEMENT BEHAVIOR DURING BIOLOGICAL MOTION PREDICTION

Dongcheng He¹, Janet H. Hsiao¹; ¹Hong Kong University of Science and Technology

Recent research has suggested that the perception of human movement from point-light displays (PLDs) depends on deriving form from motion. We investigated how form information availability and action familiarity modulate attention behavior and performance in biological motion (BM) prediction through eye tracking. In the experiment, stimuli were 28 short PLDs of human movements depicting either familiar daily actions or unfamiliar, animal-imitating movements. Each consisted of 10 frames (IFI=0.1s) of twelve point-lights (size=0.25 deg) marking three joints on each of four limbs. Stimuli varied by viewpoint (front/side), orientation ($\pm 20^\circ$, $\pm 40^\circ$), and the presence or absence of connecting skeleton lines that provided explicit form information. After a fixation display (1s), participants viewed a stimulus and then reported the speed and direction of a target joint by drawing a matching line with the mouse. Participants ($N=12$,

age:18-27 years) showed faster response times (RTs) ($p=0.042$) and reported higher speeds ($p=0.003$) for familiar actions, and the familiarity effect in RT interacted with form information availability ($p=0.018$), with a larger familiarity effect when form information was unavailable. Using Eye Movement analysis with Hidden Markov Models (EMHMM), we found two distinct eye movement patterns through clustering: a centralized pattern focusing on the stimulus center and a distributed pattern attending to both left and right of the stimulus in addition to the center (with a significant difference; $p<0.01$). Participants showed a more distributed pattern when form information was unavailable and thus had to be inferred from motion alone ($p=0.014$). Additionally, participants had less consistent eye movement behavior for unfamiliar actions, indicated by higher HMM entropy than for familiar actions ($p=0.024$). These findings suggest that the effects of form information availability and action familiarity on BM prediction were reflected in different aspects of eye movement behavior, which may jointly influence BM prediction performance.

Many thanks to the Research Grant Council of Hong Kong for supporting this work (AoE/E-601/24-N).

43.332 A COMMON ENSEMBLE PERCEPTION FACTOR FOR OBJECTS AND FACES STRONGLY RELATED TO GENERAL INTELLIGENCE

Ting-Yun Chang¹ (ting-yun.chang@hotmail.com), Oakyoon Cha², Yanchao Bi^{3,4,5}, Isabel Gauthier⁶; ¹State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research at Beijing Normal University, China, ²Department of Psychology, Sogang University, South Korea, ³School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, China, ⁴IDG/McGovern Institute for Brain Research, Peking University, China, ⁵Institute for Artificial Intelligence, Peking University, China, ⁶Department of Psychology, Vanderbilt University, USA

Ensemble Perception (EP) is the ability to extract summary statistics from groups of objects. While pairwise correlations have suggested that EP for faces and simple features may rely on distinct mechanisms, recent multivariate work revealed a domain-general EP ability spanning low- and high-level nonfacial features, one that correlates with constructs associated with general intelligence (g). Two questions remain: Does this domain-general ability extend to facial features, and is it separable from g? We administered seven EP tasks—three facial (identity, expression, age), two low-level (orientation, aspect ratio), and two high-level nonface (robots, birds)—alongside standard measures of processing speed, reasoning, and crystallized intelligence. Confirmatory factor analyses showed that low- and high-level nonface EP tasks loaded onto a single factor, and the face and object EP factors were perfectly correlated, suggesting a common EP mechanism. Although CFA models including g produced improper solutions, likely due to strong collinearity, results indicated that EP and g share substantial variance. To examine the relationship between EP and g, we conducted regression analyses controlling for age and

gender. For object EP, performance on each task (simple or complex) was uniquely predicted by both g and the aggregate performance of tasks at the other level, indicating that object EP tasks were related above and beyond the g variance. For face EP, g uniquely predicted only the facial expression task. Each face task was significantly predicted by the aggregate across the four object tasks, but not by the other face tasks, suggesting no face-specific variance once object EP variance was accounted for. A small portion (1–7%) of EP performance was explained by an ability distinct from g. These findings support a domain-general EP ability that spans facial and nonfacial tasks and indicate that this ability, while largely explained by g, may include a specialized component specific to ensemble processing.

The ST12030-Major Project (2021ZD0204100) awarded to Yanchao Bi; the David K. Wilson Chair Research Fund at Vanderbilt University and NSF Award 2316474 awarded to Isabel Gauthier; and the NSFC Grant 32500938 and the Postdoctoral Fellowship Program of China Postdoctoral Science Foundation (GZC20240129 and 2025T180944) awarded to Ting-Yun Chang.

43.333 HYPERREALISM IN AI FACE DETECTION IS CONTEXT-DEPENDENT AND PREDICTED BY OBJECT RECOGNITION ABILITY

Isabel Gauthier¹; ¹Vanderbilt University

Hyperrealism—judging AI-generated faces as real more often than genuine photographs—has been attributed to algorithmic bias that makes White AI faces statistically prototypical. The present work suggests this account is insufficient. Using poor-quality webcam photographs as real faces and ChatGPT-4o generated faces approximating their aesthetic, we studied how object recognition ability (o) predicts accuracy in AI face categorization across two studies with 68 and 101 participants. Study 1 paired faces differing in race and gender on each trial; Study 2 matched the same faces within trials on these attributes, elevating the relatively more polished appearance of AI faces as a salient cue. Hyperrealism was high in both studies and not limited to White faces. Most observers misjudged AI faces as real, doing so significantly more in Study 2. This demonstrates hyperrealism as a product of comparison context rather than fixed image properties, because the faces were the same across studies. Most observers showed high sensitivity to differences between real and AI faces (high |d'|), but their correct category assignment depended on o. O significantly predicted performance after controlling for age, gender, and general intelligence. This extends results from Chow et al. (2025), where o predicted AI face detection with StyleGAN faces. In Study 2's more challenging context, o predicted the rate of learning correct categorizations over the last 80% of trials, without receiving feedback. High-o observers appear less dependent on single salient cues and better able to integrate multiple subtle features (for instance, looking for different kinds of imperfections to identify real faces). A high o was not necessary to discriminate AI faces from real ones, but it predicted correct inferences about category membership. These results suggest

that o could help identify who would benefit most from training to detect AI-generated content.

This work was supported by the David K. Wilson Chair Research Fund at Vanderbilt University and NSF award 2316474 to Isabel Gauthier.

43.334 EFFICIENT EYES: FACE RECOGNITION ABILITY SHAPES HOW MUCH INFORMATION IS NEEDED—AND HOW MUCH THE EYES ARE USED—FOR IDENTITY AND GENDER PROCESSING

Laurianne Côté¹, Jérémy Lamontagne¹, Mélodie Potvin-Poirier¹, Caroline Blais¹, Daniel Fiset¹; ¹Université du Québec en Outaouais

Previous studies have documented a tight link between face-recognition ability and the amount of visual information required to perform a face-recognition task (Royer et al., 2015). Better recognizers not only succeed with less, they rely more on the eye region (Royer et al., 2018), a feature strongly associated with recognition success (e.g. Butler et al., 2010). However, it remains unclear whether the best face recognizers are more efficient only when processing identity, or also in other eye-based tasks such as gender categorization (e.g. Dupuis-Roy et al., 2009). In the present study, 97 participants completed two tests of face-recognition ability (CFMT and GFMT), an identity-matching task, and a gender-categorization task. To quantify the information required for task performance, we combined the Bubbles technique, which randomly samples face regions, with QUEST (Watson & Pelli, 1983) to estimate the minimum information needed to reach a 75% accuracy threshold. Our findings replicated previous work: individuals with superior face-recognition ability required significantly less visual information to perform the identity-matching task ($r = -.58$, $p < .001$). Face-recognition scores were likewise associated with needing less visual information to perform the gender-categorization task ($r = -.26$, $p = .009$), indicating that superior face recognizers require less visual information, whether judging identity or gender. Furthermore, the amount of visual information each participant required in the two tasks was positively correlated ($r = .47$, $p < .001$), suggesting a stable, cross-task signature in the quantity of information each individual needs to process faces. Importantly, this link extended to which information was used: replicating prior work, better face-recognition abilities predicted greater reliance on the eye region in both the identity-matching ($r = .27$, $p = .008$) and gender-categorization tasks ($r = .25$, $p = .012$). Together, these findings extend previous results, showing that face-recognition ability predicts the use of diagnostic eye information across tasks where this feature is relevant.

Natural Sciences and Engineering Research Council of Canada (NSERC) and Fonds de recherche du Québec - Nature et technologies

Face and Body Perception: Bodies

MONDAY, MAY 18, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY

43.335 SOFT YET ROBUST: PEOPLE SEE RICH STRUCTURE IN SPARSE DYNAMIC POINT LIGHT CLOTHS DESPITE RADICAL CHANGES IN SIZE AND LOCATION FROM MOMENT TO MOMENT

Merve Erdogan¹, Brian Scholl¹; ¹Yale University

One of the most impressive and well-known examples of the robustness of perception is biological motion: people reliably see locomoting agents from only a handful of moving dots, in point-light walkers (PLWs). But PLWs are also robust in another, even more impressive way: people reliably see biological motion even when the dot arrangements vary dramatically in size from moment to moment. Like biological motion itself, this has been attributed to special features of "bodies" — which always maintain the same constrained skeletal structure, and engage in characteristic cyclic motion patterns (e.g. pendular motion when walking). Recently, however, many of the most foundational features of biological motion (e.g. spontaneous perception, resistance to noise, disruption by spatial scrambling, and extensions to rich secondary properties) have also been observed for a very different form of non-rigid motion: point-light "cloths" (PLCs) — as when a ribbon (or a sheet on a clothesline) is waving in the wind. Are PLCs similarly robust? To find out, we showed people point-light displays of dynamic soft materials that changed radically from moment-to-moment in both size (as in previous PLW research) and location (not previously explored). Despite these constant drastic changes, people still reliably and spontaneously perceived the motions of cloths. We demonstrate this in a series of both psychophysical experiments and phenomenological demonstrations, and we show how this also extends to secondary properties such as cloth stiffness. These results are perhaps even more impressive than with PLWs, since cloths can vary dynamically so much more than bodies (due to the lack of any skeletal structure or cyclical motion). This form of robust perception in point-light displays is not specific to bodies after all, but may rather reflect more general mechanisms of extracting structure from nonrigid motion, perhaps incorporating intuitive physical constraints.

43.336 SOCIAL INTERACTION ENCODING BY NEURONS IN THE MACAQUE VENTRAL SUPERIOR TEMPORAL SULCUS

Anna Bognar^{1,2} (anna.bognar@kuleuven.be), Ghazaleh Ghamkhari Nejad^{1,2}, Michael Stettler³, Albert Mukovskiy³, Martin Giese³, Rüdiger Vogels^{1,2}; ¹Department of Neurosciences, KU Leuven, Leuven, Belgium, ²Leuven Brain Institute, KU Leuven, Leuven, Belgium, ³Section Computational Sensomotorics, Department N3, Hertie Institute for Clinical Brain Research & Centre for Integrative Neuroscience, University Clinic Tübingen, Tübingen, Germany

Primates rely on observing social interactions to infer intentions and hierarchical relationships. Although a macaque fMRI study has implicated the superior temporal sulcus (STS) in interaction processing (Sliwa & Freiwald, 2017), the neural mechanisms

underlying the processing of social interactions remain poorly understood. To examine the encoding of natural social interactions versus altered interactions and facing versus non-facing scenarios (Papeo, 2000), we recorded units (single and multi-units) targeting the ventral mid-STS and anterior-STS body patch. Animals viewed 126 2.5 s videos depicting two monkey avatars performing fighting, following, or meeting actions either in their natural form or in versions where flipping one of the two avatars disrupted their interaction. Isolated one-body videos were included to dissociate social interaction-specific signals from responses to single body components. Videos were rendered from seven different views. Using responsive and reliable units (mid-STS: 402 & 237; anterior-STS: 377 & 241 in Monkey1 and Monkey2, respectively), we found in both monkeys and regions that average two-body responses were less than the corresponding summed single body responses, in line with studies on shape interactions showing divisive normalization. Multiple-regression analyses, with responses to the single bodies as continuous predictors, facing versus non-facing as categorical predictor and their interaction terms, showed that two-body responses were better predicted by the contralateral than ipsilateral avatar with this hemifield-specific contribution being stronger in mid-STS. In anterior-STS, facing fighting avatars produced significantly stronger averaged responses than non-facing ones, and 13–17% of units showed significant interaction terms for fighting. The stronger responses to facing versus non-facing fighting videos were also supported by other analyses that compared directly facing versus non-facing interactions. Less consistent social interaction effects were observed for the meeting and following actions and in mid-STS. These results suggest encoding of social content in a minority of ventral anterior-STS units.

This research was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement Number 856495).

43.337 A DEVELOPMENTAL SHIFT IN HUMAN BODY REPRESENTATION: AGE-RELATED DIFFERENCES IN THE BODY INVERSION EFFECT FOR HEADLESS BODIES IN INFANCY

Megumi Kobayashi¹ (mkobayashi@human.niigata-u.ac.jp); ¹Niigata University

A specialized mechanism for the visual processing of human bodies is well established. One well-known example is the body inversion effect (BIE), in which discrimination performance is poorer for inverted bodies than for upright bodies (e.g., Reed et al., 2003). Recent work, however, has shown that the BIE is markedly reduced for headless bodies (Yovel et al., 2010), suggesting that the head is critical for holistic body representation. Developmental studies have reported the BIE in infancy (e.g., Zieber et al., 2015), and neuroimaging work has demonstrated cortical responses specific to human bodies (Kobayashi et al., 2021). Yet, little is known about how infants represent bodies and the role of the head. The present study examined 5- to 8-month-

old infants' ability to discriminate postures of whole and headless bodies presented upright or inverted. We employed a familiarization/novelty-preference procedure. A pre-test confirmed no spontaneous preference for either posture. Infants were familiarized with a full-color image of a particular posture and then tested for visual preference toward a novel posture as an index of discrimination. In the whole-body condition, infants showed a significant preference for the novel posture in the upright presentation (all p s < .01) but not in the inverted presentation (all p s > .1), demonstrating the BIE. In the headless-body condition, 5- and 6-month-old infants also showed the BIE (p < .05), similar to the whole-body condition. In contrast, 7- and 8-month-old infants did not show the BIE, with no significant discrimination in either orientation (p s > .1). These findings suggest that infants younger than 6 months possess relatively independent holistic representations of bodies and faces, such that removing the head does not disrupt body processing. By 7-8 months, however, infants appear to develop a more integrated, adult-like "whole body" representation in which the head plays a critical role.

This study was supported by Grant-in-Aid for Scientific Research from JSPS (19K14492, 23K25752)

43.338 THE VISUAL PROCESSING OF SOCIAL INTERACTIONS: EVIDENCE FROM EYE-MOVEMENTS

Katie L.H. Gray¹, Mahsa Barzy^{1,2}, Nicholas Hedger¹, Eugene McSorley¹, Richard Cook³; ¹University of Reading, ²University of Kent, ³University of Leeds

When viewed from a third-person perspective, the same individuals are processed differently depending on whether they are presented in an intact (e.g. face-to-face) or broken (e.g. back-to-back) interaction. One account proposes that observing an interaction engages specialised social processing mechanisms. Another proposes that the effect is driven by attentional cueing, where individuals within the interaction draw attention toward the centre of the interaction when they face each other, but attention is dispersed across the scene when they do not. Although different behavioural experiments have interrogated this effect, relatively few have used eye-tracking. In this study, we presented pairs of facing and non-facing people and arrows, allowing us to investigate how overt attention is modulated by their arrangement and the domain specificity of any effects found. On each trial, participants ($N = 31$) were presented with pairs of people or arrows whilst we tracked their eye-movements. We defined areas of interest corresponding to the heads and bodies of the two stimuli, and the space between them. A time course analysis showed that the area between the stimuli attracted more attention when the stimuli were facing each other than when they were not, for both people and arrows. Notably, this effect emerged earlier for the arrow stimuli than for people. These findings suggest that overt attention differs between interacting and non-interacting individuals, and that a similar pattern occurs for non-social directional cues. Overall, the results indicate that attention contributes to the processing of social interactions in a domain-general manner, potentially driven by attentional cueing.

The research described was funded by the Leverhulme Trust to KLHG (RPG-2019-394, and RPG-2024-245).

43.339 MULTISENSORY EMBODIMENT MODULATES BODY-CENTERED SPATIAL ATTENTION BEYOND VISUAL CUES

Hiroaki Shigemasa¹ (shigemasa.hiroaki@kochi-tech.ac.jp), Yuichi Takao¹, Harin Hapuarachchi¹; ¹Kochi University of Technology

Visual attention is captured by body-part representations (Reed et al., 2006). However, it remains unclear whether the attentional bias reported for body parts reflects purely visual factors or also critically depends on multisensory body ownership. We investigated how embodiment—multisensory integration of visual, proprioceptive, and motor signals—modulates spatial attention using virtual extended body parts. Twenty-four right-handed participants performed a visual detection task in VR across three conditions: (1) self-body with embodied extended bodies (virtual hands controlled by foot trackers), (2) self-body with non-embodied extended bodies (visually identical hands without motor control), and (3) self-body only. Following an embodiment induction task (tracking moving targets for 60 s), participants detected target stimuli appearing near self-body or extended body locations while maintaining central fixation. Since hands and feet were occupied for avatar control, responses were recorded by orienting the head-mounted display toward target locations. Reaction times (RTs) were measured. Contrary to our hypothesis that self-body would capture attention and facilitate detection, RTs were significantly faster for locations away from self-body compared to self-body locations across all conditions ($p = .046$). No significant RT differences emerged among non-self-body locations across the three conditions, suggesting that attention was not preferentially allocated based on extended body presence. Critically, comparing self-body (strong embodiment with robust proprioceptive/motor feedback) to extended bodies (weak or no embodiment with visual input only) revealed significant RT differences despite identical visual features, indicating that embodiment strength—driven by non-visual multisensory signals—contributes to attentional prioritization beyond visual body representations alone. The faster RTs away from self-body may reflect inhibition of return from initially attended self-body positions. These findings demonstrate that while visual body representations influence spatial attention, the presence of multisensory embodiment through proprioception and motor feedback provides additional attentional effects that cannot be explained by vision alone, revealing the multisensory foundations of body-centered attention.

MEXT/JSPS KAKENHI Grant Number 23K03004

Functional Organization of Visual Pathways: Retinotopy, population receptive fields

MONDAY, MAY 18, 8:30 AM – 12:30 PM,

BANYAN BREEZEWAY

43.340 SPATIALIZED VIRTUAL SOUND OBJECTS INDUCE RETINOTOPY-LIKE ORGANIZATION IN PRIMARY VISUAL CORTEX

Alara Akcan^{1,2} (alarakcan@gmail.com), Ufuk Önen¹, Zahide Pamir^{1,2}; ¹Bilkent University, Ankara, Türkiye, ²Aysel Sabuncu Brain Research Center, Ankara, Türkiye

In this study, we tested whether the primary visual cortex (V1) shows retinotopy-like responses to auditory stimuli with rich spatial structure. If auditory input can evoke such spatial maps, it would imply that topographic organization in V1 is not uniquely visual, but instead reflects a metamodal architecture in which the fundamental organizational principles of early visual cortex extend to other sensory modalities. To address this, we first generated auditory stimuli using the Dolby Atmos system, which turns the sounds into 3D sound “objects” and thus enables them to be located at a specific spatial location, with their immersive, real-like sound qualities. Then, inspired by the principles of classical retinotopy paradigms, we created a novel auditory retinotopic mapping procedure with 2-second spatialized virtual sounds consisting of pure tones alternating between lower and higher frequencies to reflect the flickering checkerboard effect, and each sound was delivered sequentially from 19 distinct spatial positions, spaced 10 degrees apart across a 180-degree horizontal field in front of the participant, through earphones. After a brief period of blindfolding and 10 minutes of training, three participants completed this auditory-adapted paradigm in fMRI. Our results showed that V1 boundaries defined using auditory stimuli were substantially consistent with the individually defined V1 boundaries by the classical retinotopy procedure. However, in control conditions where auditory stimuli did not contain any spatial information, drawing V1 boundaries using auditory stimuli was either not possible or the patterns consistent with the visually defined boundaries had weaker correlations. Previously, a retinotopic-like mapping using auditory stimuli was shown in blind echolocators, only using the echolocation sounds, indicating the formation of these maps through extensive training. Our findings suggest that auditory stimuli carrying rich spatial information evoke retinotopy-like organization in V1 even in neurotypical individuals with minimal training and short-term and reversible deprivation.

This project is funded by the Scientific and Technological Research Council of Türkiye (TUBITAK) through the 3501 Program.

43.341 TOWARDS EFFICIENT RETINOTOPIC MAPPING BEYOND THE SCREEN'S EDGE

Robert Satzger^{1,2,3}, Tomas Knapen^{4,5,6}, Martin N. Hebart^{1,2,3}; ¹Department of Medicine, Justus-Liebig University Giessen, Giessen, Germany, ²Vision and Computational Cognition Group, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, ³Center for Mind, Brain and Behavior (CMBB), University of Marburg,

Giessen and Darmstadt, Germany, ⁴Spinoza Centre for Neuroimaging, Amsterdam, The Netherlands, ⁵Department of Computational Cognitive Neuroscience and Neuroimaging, Netherlands Institute for Neuroscience, Amsterdam, The Netherlands, ⁶Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Retinotopic mapping is essential for determining the cortical topography of visual areas and delineating their boundaries. Despite its general usefulness, retinotopic mapping often suffers from two challenges. First, retinotopic mapping is typically time-intensive, but scan time is limited and expensive. Consequently, researchers would benefit from a systematic analysis of how much scan time is sufficient for stable retinotopic maps. Second, head-coil and projector geometry often place hard constraints on the field of vision inside the scanner, leaving most of peripheral visual cortex unmapped. Here, we address complementary solutions to both problems. For challenge 1, we quantify how map quality scales with scanning time to determine when performance saturates. For challenge 2, we introduce a corner-fixation paradigm that effectively expands the mappable visual field: by shifting gaze toward each corner of the screen, the stimulus reaches regions otherwise outside the displayable area. Combined, this almost quadruples the mappable visual field. We analyzed two 7T fMRI datasets, one consisting of 10 participants from the HCP retinotopy dataset (Benson et al., 2018), the other consisting of seven participants from the denseRetinotopy dataset (Zerbe et al., 2024). For challenge 1, our results surprisingly showed that a single run of ~3 min was sufficient to produce qualitatively clear polar angle and eccentricity maps at 7T that already stabilized in early visual areas, with only minimal further improvement for longer acquisitions. For challenge 2, we found that corner fixation expanded the mappable eccentricity from 10° to 18° vertically and 17° to 32° horizontally, while yielding qualitatively comparable maps to centrally presented stimuli. Together, these findings demonstrate that it is possible to strongly improve the efficiency of retinotopic mapping at ultrahigh field strengths by trading shorter acquisition duration with increased visual field coverage.

43.342 A DIRECT COMPARISON OF POPULATION RECEPTIVE FIELD MAPPING RESULTS ACQUIRED WITH 2D AND 3D ECHO-PLANAR-IMAGING FMRI ACQUISITION

Christian Windischberger¹
(christian.windischberger@meduniwien.ac.at), Dominik Zuschlag¹,
Rüdiger Stirnberg², David Linhardt¹; ¹Highfield-MR, CMPBME,
Medical University of Vienna, Austria, ²MR Physics, DZNE, Bonn,
Germany

Population-receptive field (pRF) mapping is an approach based on functional magnetic resonance imaging (fMRI) where visual stimuli are presented in the MR scanner while BOLD-weighted images are acquired. In the analysis of pRF, activation time courses in visual areas are associated with the stimulus to obtain information about receptive visual field areas processed in a given part of the visual cortex. While

2D echo-planar imaging (EPI) is still the workhorse in current fMRI brain mapping studies, EPI with three-dimensional k-space encoding (3D-EPI) has the potential to provide BOLD-weighted images with increased spatial resolution and signal-to-noise ratio (SNR). Here we compare pRF results acquired with 2D-EPI (CMRR; Moeller et al., 2010) and 3D-EPI (Stirnberg et al., 2021) protocols to evaluate potential benefits of employing 3D acquisitions. We recorded 2D and 3D data in eight young, healthy participants on an ultra-high field 7 Tesla MR scanner (Siemens Magnetom 7T) with a spatial resolution of 1.2mm isotropic (TE/TR=25/1000ms). In order to test performance at even higher spatial resolutions as required for laminar imaging, we also acquired 0.7mm isotropic resolution data in the same participants (TE/TR=25/2000ms). Minimal preprocessing was performed using fMRIPrep v.25.0.0 (<https://fmripred.org>). pRF analyses were performed using prfprepare v7.0.0 (<https://github.com/dlinhardt/prfprepare>) and GEM-pRF (Mittal et al., 2025). 3D-EPI acquisition showed medium to strong effects of improved model-fitting quantified by variance-explained values than 2D-EPI over all early visual areas (V1: Cohen's d=0.56, V2: d=0.44, V3: d=0.47). Wilcoxon signed-rank testing showed highly significant increases in explained variance values of 3D over 2D EPI results while no changes in pRF parameters were apparent. Ultrahigh-resolution runs (0.7mm) were not affected by obvious imaging artefacts and showed the expected retinotopic maps across the early visual cortex (V1-3). Our results clearly demonstrate the advantages of 3D-EPI acquisition protocols in pRF studies with millimeter or even sub-millimeter resolution.

This research was supported by the FWF (10.55776/P35583; 10.55776/PAT8722623; 10.55776/ESP1445525)

43.343 STIMULUS CONFIGURATION AND TEMPORAL SAMPLING JOINTLY MODULATE PRF REPRODUCIBILITY

David Linhardt¹ (david.linhardt@meduniwien.ac.at), Siddharth Mittal¹,
Michael Woletz¹, Samuel Schwarzkopf², Christian Windischberger¹;
¹Medical University of Vienna, ²University of Auckland

Population receptive field (pRF) mapping is widely used to quantify visual field representations, yet the degree to which reproducibility depends on specific stimulus properties remains unclear. Here we systematically tested the impact of stimulus sequence design and temporal sampling on pRF test-retest reliability. We re-analyzed the CHN retinotopy dataset (N=12; two sessions, three runs per session; Chang et al., 2025; openneuro.org/datasets/ds004698). pRF parameter estimations were obtained based on the 2D Gaussian model in GEM-pRF (Mittal et al., 2025), using two modelling pipelines. In the standard pipeline, the stimulus was downsampled to the TR (1.2 s) before HRF convolution. In the high-resolution pipeline, HRF convolution was performed at the stimulus-native temporal resolution, and only the final model time course was downsampled. Each run used an otherwise identical, smoothly moving bar stimulus that differed only in the pseudo-randomized starting angles for each crossing. Order of the three-bar presentations was identical across sessions and

participants. Reliability was assessed via within-subject correlations between session-matched runs. High-resolution modelling produced a statistically reliable improvement in eccentricity and size estimates ($p < 0.001$). The absolute effect was small ($\Delta r \sim 0.01$), given the already high baseline correlations. In contrast, the choice of starting angle had a much larger impact. Despite matched duration, number of peaks and stimulus energy, median reliability varied markedly across stimulus configurations, by almost an order of magnitude more than the effect of high temporal resolution modelling. Together, these results show that while high-resolution modelling yields a measurable benefit, stimulus sequence design is a stronger factor for pRF reliability. Careful control of stimulus conditions offers a simple way to improve reproducibility and is straightforward to implement in studies aiming for clinical use.

Research was supported by the Austrian Science Fund (FWF) [10.55776/P35583, 10.55776/PAT8722623]

43.344 POPULATION RECEPTIVE FIELD PARAMETERS ARE BIASED BY STIMULUS FEATURES

Garikoitz Lerma-Usabiaga^{1,2} (garikoitz@gmail.com), Miguel Martinez-Zaldivar¹, Yongning Lei¹; ¹BCBL, Basque Center on Cognition, Brain and Language, San Sebastián, Gipuzkoa, Spain, ²IKERBASQUE, Basque Foundation for Science, Bilbao, Bizkaia, Spain

Population receptive field (pRF) modeling characterizes the aggregate spatial tuning of neuronal populations within a voxel. However, these estimates theoretically depend on the specific neural subpopulations recruited by the texture and features of the stimulus carrier pattern. Prompted by preliminary observations that word stimuli yield more foveal eccentricity estimates than standard checkerboards, this study systematically investigates how differing stimulus carrier patterns, flickering frequency, and pattern size modulate recovered pRF parameters in healthy human visual cortex. We employed a high-precision, dense-sampling fMRI protocol, scanning 10 participants over 10 sessions each. Retinotopic mapping was performed using 2 runs of standard 8-bar apertures filled with one of three carrier patterns: flickering checkerboards, words, or false fonts. To rigorously test the stability of carrier-dependent effects, we varied the internal element size and refresh frequency of the patterns across sessions. We quantified systematic deviations in recovered pRF parameters and assessed the inter-subject replicability of these shifts. Results demonstrate high consistency across sessions, confirming measurement reliability. Recovered pRF parameters were modulated by stimulus properties. Beyond the eccentricity shifts driven by carrier content, variations in element size and refresh frequency altered pRF estimates. The magnitude of these shifts was small in the early visual cortex but substantially larger in more anterior regions such as VO-1 and hV4, reaching more than 1 standard deviations outside of the fovea. Notably, we observed substantial inter-subject variability, with different participants showing distinct patterns of carrier-dependent modulation. We conclude that pRF parameters are not fixed anatomical constants but dynamic measures dependent on the

specific neural subpopulations recruited by stimulus features. By quantifying how texture, size, and frequency drive these variations, our findings provide a first step toward resolving discrepancies in prior literature and establish essential constraints for designing future visual mapping experiments.

This research is supported by the Basque Government through the BERC 2022-2025 program and Funded by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation CEX2020-001010/AEI/10.13039/501100011033 and through project PID2024-161512NB-I00 funded by Spanish Ministry of Science and Innovation

43.345 POPULATION RECEPTIVE FIELD (PRF) MAPPING: IS GRID FIT ALL YOU NEED?

Siddharth Mittal¹ (siddharth.mittal@meduniwien.ac.at), Michael Woletz¹, David Linhardt¹, Christian Windischberger¹; ¹Center for Medical Physics and Biomedical Engineering, Medical University of Vienna

Population receptive field (pRF) mapping quantifies retinotopic organisation by estimating receptive-field location and size across the visual cortex. Most approaches use a two-stage fitting process: gridfit to obtain an initial solution, followed by refinefit to optimise pRF parameters. Despite its widespread use, the actual impact of refinefit and whether the improvement justifies the computation time has not been quantified systematically. We evaluated refinefit adjustments using two software implementations: mrVista (Dumoulin & Wandell, 2008) and GEM-pRF (Mittal et al., 2025). Both use similar modelling, but GEM-pRF applies a reformulated mathematics tailored for GPU computation and improved computational performance. Analyses were performed on the NYU retinotopy dataset (44 subjects; Himmelberg et al., 2021), focusing on V1. The stimulus radius was 12.4° , and both analyses used a 24.8° search space. mrVista used a triangular grid with 0.25° spatial spacing and hybrid pRF-size spacing, whereas GEM-pRF used a Cartesian grid with 0.33° spatial and 0.3° size spacing. For each voxel, gridfit and refinefit estimates of pRF centre (x, y) and size (σ) were extracted, including only voxels with $>10\%$ variance explained. In mrVista, 46.18% of voxels and in GEM-pRF, 11.97% voxels, showed no change between gridfit and refinefit. Among voxels that did change, 75.53% in GEM-pRF and 72.09% in mrVista stayed closer to their original gridfit position than to any neighbouring grid point. Overall, 78.46% (GEM-pRF) and 84.98% (mrVista) either remained on a grid point or shifted by less than the grid spacing. These adjustments led to modest improvements in explained variance (R^2), with mean gains of 0.0623% in GEM-pRF and 0.1220% in mrVista. These findings suggest that improving the accuracy of pRF mapping results using refine-fitting might not be necessary if the grid is dense and meets the accuracy requirements of the given application.

This research was funded in whole or in part by the Austrian Science Fund (FWF) [<https://www.doi.org/10.55776/P35583>] and [<https://www.doi.org/10.55776/PAT8722623>]

43.346 NORMALIZATION BASED POPULATION RECEPTIVE FIELD MODEL CAPTURES FMRI RESPONSE CHANGES DRIVEN BY SHIFTS IN NEURAL INPUT IN HUMAN EARLY VISUAL CORTEX

Jeongyeol Ahn¹, Lilyanna Gross¹, Elisha Merriam¹; ¹National Institute of Health

Normalization has been proposed as a core computation in early visual cortex and helps explain key response properties such as surround suppression. Recent work suggests that population receptive field measurements reflect both excitatory and suppressive components, and some models have begun to capture these in fMRI data. These approaches show that inhibitory effects can be detected, but they rely on additive suppression and do not offer a generative account based on divisive normalization. A population receptive field framework that explicitly incorporates normalization and explains response changes driven by shifts in neural input has not yet been established. Here we propose a population receptive field model that predicts voxel responses through a generative process based on divisive normalization. We assume that each voxel contains a population of neurons that produces excitatory responses to visual stimulation and another population with similar receptive field structure that contributes suppressive drive through normalization. We simulated the resulting BOLD responses under this framework and found that the model predicted response profiles in two situations. The first involved increases in baseline activity induced by external inputs such as brain stimulation. We compared these predictions with findings from a recent fMRI study that measured population receptive fields after brain stimulation. The model captured increases in baseline activity and surround suppression. The second involved increases in stimulus size that elevated the pooled input to neighboring neurons. To test our model, we presented moving bar apertures of three widths to human participants while acquiring fMRI time series. The observed responses showed systematic changes in suppression across bar widths, and the model successfully captured these patterns. These findings show that a population receptive field model grounded in divisive normalization provides a coherent account of fMRI response changes driven by shifts in neural input in human early visual cortex.

This research was supported by the Intramural Research Program of the NIMH (ZIA MH002966) to E.P.M.

MONDAY MORNING POSTERS IN PAVILION

Spatial Vision: Binocular vision

MONDAY, MAY 18, 8:30 AM – 12:30 PM, PAVILION

43.401 PREDICTING BINOCULAR VISUAL ACUITY WITH ANALYTICAL SUMMATION MODELS

Wei Hau Lew¹, David Borja¹, Derek Nankivil¹; ¹Alcon Research, LLC

Binocular summation models traditionally predict cyclopean performance in contrast discrimination and detection based on monocular inputs. A key question is whether these models can extend to visual acuity (VA). Clinically, measurements of binocular VA curves across accommodative demand are used to evaluate presbyopic corrections (intraocular lenses, multifocal contacts). These curves reveal how binocular summation changes across vergence planes and the impact of interocular differences. We adapted two binocular summation models (Legge & Gu 1989—LG and Ding & Sperling 2006—DS) to predict binocular VA curves using retrospective data from seven published clinical studies (collectively, $n = 320$ subjects). The LG model employs quadratic summation of monocular contrast inputs, while the DS model incorporates weight for gain control between the two eyes' contrast. To improve the LG model, a scaling factor and a cross-term were added to allow penalization for intraocular differences, thereby giving both models two degrees of freedom. Parameters were optimized under constraints to minimize the root mean square error (RMSE) of the pooled data. Leave-one-out cross-validation was performed to assess model generalizability. Mean RMSE for LG was 0.030 [range: 0.020-0.044] while DS model was 0.019 [range: 0.010-0.035] logMAR. All clinical data exhibited summation, and both models predicted summation consistently. Maximum absolute error was 0.083 logMAR for LG (< 0.05 in 51/56 predictions) and 0.065 for DS (< 0.05 in 54/56 predictions). While both models typically achieved accuracy within half a line of VA, the DS model more closely matched clinical data and behaved more reasonably outside the observed data range. In summary, both models reliably predicted binocular VA curves meeting clinical standards for accuracy. DS model may serve as a useful tool in developing and evaluating presbyopic correction strategies. However, results are based on retrospective data; further validation and calibration across diverse populations are needed to confirm generalizability.

43.402 RECURRENT PROCESSES SYNCHRONIZE LOCAL COMPETITIONS INTO COHERENT PERCEPTION IN BINOCULAR RIVALRY

Junjie Yan¹, Chencan Qian¹, Peng Zhang¹; ¹Institute of Biophysics, Chinese Academy of Sciences

Conscious perception spontaneously alternates between the two eyes' images during binocular rivalry. While the neural mechanisms driving the temporal dynamics of binocular rivalry have been extensively studied, how local competitions are synchronized into a spatially coherent perceptual representation remains elusive. In a series of psychophysical experiments, we found that spatial synchronization of

local rivalry dynamics is strongly dependent on visual awareness, spatial attention, and eye-based attention, but not sensitive to low-level factors such as stimulus position or cortical distance, thereby implicating a critical role of top-down feedback in synchronizing local competitions. To further investigate the underlying computational mechanisms, we built a hierarchical predictive-coding model and verified that high-level predictive activity is required to synchronize local rivalry across space, but is not necessary for generating local competitions themselves. While high-level switches typically precede those at lower levels, perturbation experiments showed that increasing the synchronization among low-level nodes exerts a stronger bias on subsequent high-level switches, forming a recurrent loop to drive the rivalry dynamics. Together, these results support a recurrent mechanism by which local and global competitions are coordinated to produce a spatially coherent perceptual representation.

43.403 RAPID ASSESSMENT OF STEREO THRESHOLDS USING AN INTERACTIVE RANDOM DOT STEREOGRAM IN VIRTUAL REALITY

Triya Belani¹ (tb4266@rit.edu), Chengyi Ma¹, Gabriel J. Diaz¹; ¹Rochester Institute of Technology

Stereovision provides precise depth perception and supports accurate judgments during visually guided actions. Approximately 7% of the population younger than 60 years is stereoblind [Chopin et al., 2019], yet despite its essential role in daily activities, stereovision is often under-assessed in clinical practice, largely due to limitations of current methods and the presence of monocular cues [Chopin et al., 2019; O'Connor et al., 2018]. To address this issue, we developed a method to render virtual reality (VR) environments using a real-time random-dot stereogram that isolates binocular disparity information in interactive VR tasks. Here, we demonstrate how this tool can be used in a stereo-threshold assessment task. Fine stereoacuity thresholds were measured in VR using a three-disc “Randot” test. A 2-up 1-down adaptive staircase procedure was used to obtain an individual rough stereoacuity estimate and to select the disparity range. The staircase started with an initial disparity of 150 arcseconds and a step size of 50 arcseconds, which was halved after each incorrect response. The procedure ended after four incorrect responses. Thresholds were measured for eight adults (ages 18–37 years) with self-reported normal binocular vision and no corrective lenses. Participants completed 2–3 blocks of 100 trials each, with 10 fixed disparities around the threshold estimate obtained from the staircase procedure. We estimated disparity thresholds by fitting a Weibull psychometric function (guess rate = 1/3, lapse rate = 0.05) and defined threshold as the disparity yielding 64% correct performance. Each block required approximately 5 minutes. The mean within-participant standard deviation of thresholds was 3.15 arcseconds, suggesting that the tool can measure fine stereoacuity with improved isolation of disparity and reduced monocular cues. In addition to this data, we present preliminary data on whether these measures predict performance in a visually guided interception task.

43.404 STEREOSCOPIC VISION FACILITATES THE TRACKING OF MULTIPLE OBJECTS, BUT HAS LITTLE IMPACT ON THE TRACKING OF A SINGLE OBJECT

Remy Allard¹ (remy.allard@umontreal.ca), Sam Zakko¹, Gabriel Ladaa¹, Maryam Rezaei¹, Jocelyn Faubert¹; ¹Universite de Montreal

Predators, including humans, have front-facing eyes enabling binocular vision over a large visual field. Why did evolution sacrifice a wider visual field for more binocular vision in the periphery? Stereoscopic peripheral vision must provide a considerable advantage for predators to track preys. The current study investigated if binocular disparity is more useful when tracking multiple objects or a single object, and if binocular disparity is helpful to track objects that are sometimes occluded. In the single-object tracking condition, participants were asked to track one of three balls rotating around a point at the eccentricity of 17 degrees of visual angle. In the multiple-object tracking condition, three balls were presented in each of the four quadrants and the participant had to simultaneously track one ball per quadrant. The contribution of binocular disparity was evaluated by measuring the maximum speed threshold at which objects could be tracked when the disparity cues were available relative to when they were not. The amplitude of the rotation was 6 cm in 3D space along the meridian and depth axis, and was either 0, 2, 4 or 6 along the eccentricity axis. When the amplitude was 0, the balls rotated along a disk that was aligned with the observer and were sometimes occluded behind another ball. Under occlusion conditions (eccentricity amplitude=0), the contribution of binocular disparity was considerably high when tracking 4 balls (~2x), but weak when tracking only one ball (~1.2x). When the balls were never occluded, disparity cues did not provide any advantage. In sum, binocular disparity was considerably useful only for tracking multiple objects that are sometimes occluded. These results suggest that predators have front-facing eyes to facilitate the tracking of multiple objects that are sometimes occluded.

This research was supported by NSERC grants to RA and JF.

43.405 STEREOACUITY AND INTEROCULAR CONTRAST BALANCE: VISUAL FIELD LOCATION MATTERS

Preeti Verghese¹ (preeti@ski.org), Ângela Gomes-Tomaz², Noelia Alcalde¹, Jian Ding², Adrien Chopin¹, Dennis Levi²; ¹Smith Kettlewell Eye Research Institute, ²University of California, Berkeley

An imbalance in contrast sensitivity between the eyes is thought to contribute to poor stereoacuity in amblyopia (Vedamurthy et al., 2015; Webber et al., 2020). However the precise locus (foveal or peripheral) that mediates stereoacuity was not determined. To investigate this, we made local measurements at the fovea and at multiple locations across the central 20° of the visual field, in controls and in persons with amblyopia. We used random-dot stimuli that scaled with eccentricity to measure both stereoacuity and contrast imbalance. Observers (14 controls; 5 with amblyopia: 3 strabismic, 1 anisometropic and 1 mixed) adjusted the contrast of the dominant-eye stimulus to match the perceived contrast of the non-dominant eye

stimulus, at the fovea and at the locus of best stereopsis. Results showed no systematic relationship between stereoacuity and the contrast ratio at perceptual balance for controls at the fovea ($r = -0.40$, $p = .16$), or at the locus of best stereopsis ($r = -0.15$, $p = .62$), or for the amblyopic group at their best locus ($r = 0.20$, $p = .78$). Individuals with amblyopia had no measurable stereopsis at the fovea. We also used the Ding et al. (2013) model to estimate each eye's contribution to binocular combination after contrast normalization, but this asymmetry ratio also showed no systematic relationship to stereoacuity at the fovea (control: $r = 0.40$, $p = .16$), or at the best locus (control: $r = 0.15$, $p = .62$, amblyopia: $r = -0.20$, $p = .78$). However, in the stereoblind fovea of individuals with amblyopia, asymmetry ratios were significantly larger (GLM $p = 0.01$, Cohen's $d = 1.13$, large effect size). Our results together with previous findings suggest that while contrast imbalance may contribute to impaired stereopsis in the central retina in amblyopia, this is not necessarily true at the (peripheral) locus of best stereopsis.

NIH R01 EY034370

43.406 NEURAL RESPONSES TO INTEROCULAR CORRELATION

Bruno Richard¹ (bruno.richard@rutgers.edu), Robert Hess², Sing Ip Lee³, Daniela Marinova³, Daniel Baker³; ¹Rutgers University - Newark, ²McGill University, ³University of York

The visual system encodes binocular signals via summing and differencing channels, yet the relative weighting of these pathways remains unclear. Here, we use Steady-State Visually Evoked Potentials (SSVEPs) to investigate the neural responses associated with summing and differencing mechanisms. In Experiment 1, we measured responses to binocular noise stimuli in 15 observers. Across the eyes, stimuli could oscillate (at 3Hz) from no correlation to perfect interocular correlation, from no correlation to perfect interocular anticorrelation (i.e., opposite interocular contrast), or no interocular correlation (i.e., the control condition). Responses to correlated stimuli showed a peak at the fundamental frequency (3Hz); however, we found no evidence of measurable responses to anticorrelation. In Experiment 2, we increased the sample size to 48 observers and measured SSVEPs in response to interocular correlation and anticorrelation while manipulating the degree of interocular correlation. The correlation between the left and right eye images could be ± 0 , 0.12, 0.24, 0.48, or 0.96. We found SSVEP responses at the fundamental frequency in both interocular correlation and anticorrelation conditions. Interocular correlation generated SSVEPs with a correlation of 0.48 or 0.96, while anticorrelation SSVEPs were only measurable at a correlation of -0.96. Signal-to-noise ratios were 1.15 times larger for the correlation condition than the anticorrelation condition. To determine the influence of individual differences on SSVEPs to interocular correlation, we collected psychophysical measures of binocular function and correlated them with neural responses. Two measures of binocularity were used: stereoacuity and binocular balance, which refers to the ratio of dominance duration in

each eye during rivalry. We found only small associations between SSVEP magnitudes and psychophysical measures of binocularity (stereoacuity and binocular balance). We conclude that while neural responses to both correlation and anticorrelation can be recorded via SSVEPs, the latter requires high signal coherence and a large sample size.

43.407 WHAT DETERMINES SWITCH VERSUS RETURN TRANSITIONS DURING BINOCULAR RIVALRY?

Oakyoon Cha¹; ¹Department of Psychology, Sogang University

Under binocular rivalry, two incompatible images compete for awareness, producing perceptual alternations. Rivalrous perception consists of periods of exclusive dominance, where only one image is consciously visible, and mixed dominance, where a dynamic patchwork of the two images is experienced. Mixed dominance can lead either to a switch, in which perception progresses to the other image, or to a return, in which perception briefly becomes patchwork-like but subsequently reverts to the previous image. To investigate the factors that determine switch versus return transitions, I modeled the durations of the two mixed-dominance types using separate Linear Ballistic Accumulator (LBA) models. Drift rates did not differ between models, suggesting that the state at exclusive-dominance breakdown does not bias which transition occurs. In contrast, both the decision boundary and variability of starting point were larger for returns than for switches. These results support the interpretation that the "distance to travel" in perceptual space governs mixed-dominance dynamics. Return events may reflect incomplete progress toward the alternative percept. This account predicts that the boundary for returns should be less than twice that for switches, because failed progress would revert to the original percept. Consistent with this prediction, this pattern was observed for all 10 participants, suggesting that switches and returns may be captured within a single model incorporating a failure-to-switch component. Altogether, these findings provide a mechanistic account of mixed-dominance trajectories in binocular rivalry that can be captured within the LBA framework.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT) (No. RS-2023-00211668) and by the Sogang University Research Grant of 2025 (202510012.01).

43.408 TRANSIENTLY INCREASING SUPPRESSED STIMULUS STRENGTH SHORTENS RIVALRY DURATIONS

Taylor J. Knickel^{1,2}, Stephen A. Engel^{1,2}; ¹University of Minnesota, Minneapolis, MN, United States., ²Center for Applied and Translational Sensory Science, University of Minnesota, Minneapolis, MN, United States.

Purpose: In binocular rivalry, two incompatible images are presented one to each eye, and perception alternates between the two. The primary hypothesis behind switches is adaptation, involving the weakening of the dominant stimulus. An alternative theory proposes

that information from the suppressed eye accumulates, and switches occur when a threshold is reached. Here, we test for accumulation by transiently increasing the contrast of the suppressed stimulus to see whether this added information shortens the time to the subsequent switch (“dominance duration”), even if a switch was not immediately produced. Methods: Six observers viewed mid-contrast (0.15) orthogonal grating patterns presented one to each eye and reported their perception. Rivalry was measured in 80 second runs during which a one second Gaussian “pulse” was added to the suppressed stimulus. Five different pulse contrast levels (0.05, 0.1, 0.2, 0.4, 0.8) were used in different runs and could occur at two different onsets (zero or one second after a switch). 250 runs were collected over 10 sessions. Results: Dominance durations followed the gamma distributions typical of rivalry. For analysis, we only considered dominance periods that ended after the end of the pulse, effectively removing “breakthroughs” where a visible pulse caused an immediate switch. Overall, as the contrast of the pulse increased, mean dominance durations decreased monotonically from 3.77 seconds (no-pulse) to 2.504 seconds at 0.8 contrast. These effects were statistically significant within all observers for the highest contrast levels, (K-S Test, $p < 0.001$). Even the longest twenty percent of dominance durations, when switches occurred long after the pulse end, were reliably shortened. The effect of the pulse was relatively consistent between the two pulse onsets. Conclusion: The simplest account of these results is that information from the suppressed eye persists and can influence rivalry as it accumulates.

5T32EY025187-09

Eye Movements: Natural, complex tasks

MONDAY, MAY 18, 8:30 AM – 12:30 PM,
PAVILION

43.409 GAZE BEHAVIOR DURING PERFORMANCE OF CONTROLLED TASKS ON NATURAL SCENES

Amy Poole¹ (poole163@umn.edu), Eline R Kupers², Adrian Wong¹, Clayton Curtis³, Kendrick Kay¹; ¹University of Minnesota, ²York University, ³New York University

Several studies have shown that gaze patterns during natural scene viewing are different across individuals (Castelhano & Henderson, 2008), directed towards salient objects (Anderson et al., 2015), and task-dependent (Yarbus, 1967). To better characterize these effects, we conducted an eye tracking experiment in which we independently manipulated the natural scene viewed as well as the task performed. Participants ($n=11$) engaged in 6 different tasks on 30 scenes that were balanced with respect to content and spatial layout (a total of 180 trials). Each scene was shown for 2 seconds. The tasks were to (1) detect a contrast change of the entire scene, (2) categorize the scene as indoor or outdoor, (3) classify the dominant object, (4) determine

the location of the dominant object, (5) decide how one would act in the scene, and (6) hold the scene in working memory (and after an 8-s delay, decide whether content was added or removed). Eye tracking data were recorded with an Eyelink 1000, and a density map indicating fixated locations was created for each trial. We find that, after collapsing scenes and tasks, individuals' density maps differ significantly from one another. Moreover, when collapsing participants and tasks, density maps also differ significantly across scenes. Finally, collapsing participants and scenes, we see that most tasks have distinct density maps. In particular, the working memory task provoked the most widespread gaze, while the contrast change task provoked the least widespread gaze. We conclude that all three factors—individual, scene, and task—affect where people look. Future analyses will evaluate how well existing models of eye movements can account for these data.

NIH R01 EY034118 to KK and CC

43.410 ARCMINUTE-SCALE SIMULATED CENTRAL SCOTOMAS INDUCE SYSTEMATIC CHANGES IN PERFORMANCE AND FINE OCULOMOTOR BEHAVIOR DURING READING

Yue Guzhang¹, Krishnaveni Nagarajan¹, Yesol Yu¹, Martina Poletti¹; ¹University of Rochester

Simulated scotomas are often used to study how central vision loss affects reading. However, due to technical limitations in eyetracking, most studies employ relatively large text sizes and scotomas occluding an area much larger than the foveola. Under these conditions, saccadic behavior is substantially altered. Yet it remains unclear whether much subtler disruptions to central vision also produce measurable changes in eye movements during reading, changes that could serve as early biomarkers for certain forms of retinal degeneration. Here, we used a high-resolution eyetracker paired with a gaze-contingent display that enables precise localization of the line of sight. We simulated a small scotoma centered on the preferred locus of fixation, rendered with a Gaussian transparency window ($SD = 0.2^\circ$) that blended with the background. Participants ($N = 8$ emmetropes) read brief paragraphs (0.2° x-height). After each paragraph, they answered true/false comprehension questions and completed a word-recognition task indicating whether four target words had appeared in the passage. Participants showed an 8% drop in comprehension accuracy ($P < 0.01$) and a 9% drop in word-recognition accuracy ($P < 0.01$) in the scotoma condition compared to the no-scotoma condition. The small central scotoma also altered fixational eye movements; the duration of intersaccadic periods increased by an average of 42 ms ($P < 0.01$), and ocular drifts became significantly less curved and covered a larger spatial span relative to baseline. Additionally, in the scotoma condition, both saccade and microsaccade rates decreased (by 10% and 12.3%, respectively, $P < 0.01$), and the rate of regressive microsaccades increased.

This work was funded by NIH Grant EY029788 to MP and NIH grant EY001319 to the Center for Visual Science.

43.411 INFLUENCE OF SCENE CONTEXT ON EYE MOVEMENTS DURING GAZE PERCEPTION

Srijita Karmakar¹ (srijita@ucsb.edu), Ana Dai Liu¹, Miguel P. Eckstein¹; ¹University of California, Santa Barbara

Introduction: The perception of someone's gaze direction is influenced by the gazer's eye and head orientation, face eccentricity, and the presence and location of gaze goals (Han & Eckstein, 2023; Loomis et al, 2008; Todorović, 2009). However, in the real world, humans infer others' gaze in the context of their impending actions. The extent to which gaze judgments and eye movements are influenced by scene context remains poorly understood. **Methods:** Here, we develop a new image stimulus type that preserves head and eye gaze information while introducing an object that alters the implied gazer's intended actions (Gaze Winograd Images, GWI). Specifically, we manipulated the presence of an object in the gazer's hand that was contextually related to one of the other objects in the vicinity of the gazed-at object. We collected perceptual judgements and eye-movement data across an online behavioral study (N = 180) and an eye-tracking study (N = 20) for 52 GWIs (26 pairs), where observers reported which object the gazer was looking at. We hypothesized that the presence of an object in the gazer's hand would bias observers towards the related object and away from the gazed-at object. **Results:** In the "object-in-hand-present" condition, the non-gazed contextually-related object was perceived as the gazed object ($p < .001$) as well as fixated more frequently ($p < .001$) than the gazed-at object. The mean proportion of fixations (per trial per subject) on the contextually-related object ($M = 0.17$, $SD = 0.033$) was significantly higher than on the gazed-at object ($M = 0.11$, $SD = 0.023$) ($p < 0.001$). In a control "object-in-hand-absent" condition, observers fixated more frequently on the gazed object and selected it as being gazed at ($p < 0.05$). **Conclusion:** Together, our findings demonstrate that human judgements of gaze and related eye movements reflect an inferential process of integrating the gazer's head direction with contextually relevant scene information related to the gazer's future actions.

43.412 DRIVER GAZE BEHAVIOUR WHEN MONITORING FOR ROAD HAZARDS: EFFECTS OF VISUAL DISTRACTION

Ginnie Wee¹ (ginnie.wee@mail.utoronto.ca), Jiali Song¹, Benjamin Wolfe¹; ¹University of Toronto Mississauga

Safe driving requires detecting hazards in dynamic road scenes. Dividing attention between the road and a secondary task increases collision risk, and eye tracking can reveal how monitoring strategies change under distraction. During visual search, divided attention increases the incidence of looked-but-failed-to-see (LBFTS) errors, in which observers miss targets despite having looked directly at them. This study examines whether these errors occur during dynamic road hazard monitoring (Experiment 1), and whether eyes-off-road (EOR)

warnings reduce distraction's impacts (Experiment 2). In Experiment 1, twenty licensed drivers performed two tasks – a hazard detection task in which drivers reported the locations of hazards in 320 dashcam videos (2 to 8s), and a secondary visual task requiring continuous off-road monitoring. Tasks were performed separately under focused attention, or simultaneously under divided attention. Hazard detection accuracy was lower under divided attention compared to focused attention regardless of whether drivers fixated on the hazard ($t(19) < -3.81$, $p < 0.001$). Critically, visual distraction increased both LBFTS and didn't-look-and-didn't-see (DLDS) errors, in which drivers miss hazards because they did not look at them. Under distraction, approximately two-thirds of missed hazards were DLDS errors. As most misses were DLDS errors, Experiment 2 investigated whether real-time gaze monitoring with EOR warnings reduced miss rates. Eight drivers completed the divided attention condition from Experiment 1 in an EOR warning condition and a warning-absent condition. In the warning condition, a visual alert was given when drivers looked away from the road video for a set duration (1 or 1.5s, in separate conditions). EOR events rarely elicited warnings in the 1.5s condition, suggesting that this threshold is less likely to be helpful, but frequent warnings in the 1s condition may reduce LBFTS and DLDS errors. Together, these results suggest that distraction increases both kinds of errors and point to interventions that may reduce these errors.

This work was funded by Transport Canada Enhanced Road Safety Transfer Payment Program.

43.413 ACTIVE FIXATIONAL CONTROL IN A REAL-WORLD GROOMING TASK

Ruitao Lin¹, Jie Z. Wang¹, Michele A. Cox¹, Paul Jolly¹, Yuanhao H. Li¹, Soma Mizobuchi¹, T. Scott Murdison², Alina Neverodskaya¹, Zhetuo Zhao¹, Michele Rucci¹; ¹University of Rochester, ²Reality Labs

During natural visual exploration, head and eye movements cooperate to control the motion of the stimulus on the retina. Previous research with head immobilization has shown that observers regulate their fixational drifts according to task demands: retinal image motion is attenuated when fine spatial discrimination is required. Here we asked whether similar tuning of retinal motion occurs under fully natural, head-free conditions during a dynamic task where demands for fine spatial vision and visuomotor coordination change over time. Observers (N = 7) performed a complex visual search and motor-interaction task inspired by primate social grooming. They searched for small, colorful "fleas" (2.5-mm radius spheres) hidden in a box filled with brown crinkled paper simulating fur. Subjects freely moved their head and hands, using tweezers to manipulate the "fur," locate the fleas, and deposit them into a container. Precise measurements of head, eye, and hand movements were continuously recorded using a custom apparatus combining a high-precision motion-capture system with a specifically designed magnetic-induction eye tracker. Subjects wore scleral coils in both eyes and a tight-fitting helmet with markers for head tracking. Tweezer position and state (open/closed) were tracked with additional markers. Retinal image motion was

reconstructed from head and eye measurements using a two-nodal-point eye model. Retinal motion traces showed the characteristic alternation between saccades and drifts. Critically, intersaccadic drift exhibited systematic modulation during the task: retinal speed declined within individual fixations and was markedly attenuated in the seconds preceding the grasping of a flea. These changes increased the relative power of higher spatial frequencies both within each fixation and across the trial. The results parallel previous findings under head immobilization and demonstrate continuous, task-dependent tuning of retinal motion during natural behavior, revealing a dynamic oculomotor strategy that adaptively reduces retinal motion to support fine spatial processing in complex, naturalistic tasks.

Research supported by Reality Labs, NIH EY18363 and P30 EY001319

43.414 GAZE BEHAVIOR WHILE POURING A CUP OF COFFEE

Niteesh Midlagajni¹, Carola Stork¹, Jolande Fookien¹, Constantin Rothkopf¹; ¹Centre for Cognitive Science, Technical University of Darmstadt

Many everyday activities, such as pouring a glass of water, require continuous visuomotor control. Interaction between gaze and motor control during such naturalistic tasks remains poorly understood. Here, we designed a novel experimental setup, combining mobile eye tracking, object tracking, and a custom-made digital scale to investigate the visuomotor control of pouring. Twenty participants poured water from three different jugs into three different cups under two conditions: 'self-paced', where they poured naturally, and 'fast', where they were instructed to pour as quickly as possible. In both conditions, participants received no explicit instructions regarding visuomotor behavior, target fill-levels, or time constraints. Our results reveal that participants had preferred fill-levels and that gaze behavior during the task was highly consistent. Participants looked at the jug before grasping it, then rapidly shifted their gaze to the cup and held it there until pouring began. During pouring, gaze remained confined to the cup region, hovering either on the wall of the cup, the liquid surface, or the rising fill-level. Once pouring was complete, gaze became task-irrelevant, and participants returned the jug without visual guidance. To determine when gaze is most critical, we quantified the variability in gaze position as a function of fill-level. Gaze density systematically decreased as participants approached their preferred fill-level, reflecting increased visual monitoring towards the end. In the 'fast' condition, gaze was even more concentrated throughout the task, suggesting a stronger reliance on continuous visual feedback to pour quickly. Finally, we observed individual differences in visuomotor strategies: some participants tracked the rising liquid via smooth pursuits, whereas others monitored only near the end. Notably, these differences did not affect task performance. Overall, this work demonstrates that structured visuomotor strategies emerge even in simple, naturalistic tasks, underscoring the importance

of studying everyday tasks to better understand the role of natural vision.

This research was supported by the European Research Council (ERC; Consolidator Award 'ACTOR'- project number ERC-CoG-101045783).

43.415 NOVEL OCULOMOTOR SEQUENCE LEARNING PARADIGM: IMPLICATIONS FOR AUDIO-VISUAL CUED LEARNING

Michael Petrovski¹ (michael.petrovski00@gmail.com), Susu Beheiry², Danna Robayo³, Udichi Das⁴, Royze Simon⁵, Joseph Desouza⁶; ¹York University

Motor sequence learning involves the coordination of both oculomotor and manual motor systems through the practiced repetition of a fixed sequence of actions, resulting in automatized execution of movement through habit learning. This study aims to develop a new visual-motor-based learning paradigm with music, potentially promoting neuroplasticity and creating new interventional tools as rehabilitation using dance in people with Parkinson's disease shows behavioural and putative neural changes. Eye movements of 10 participants (8 female, 2 male) were tracked using the Eyelink 1000 Plus system during a 68-second eye-dance sequence. The experiment consisted of a learning phase and a performance phase, with music incorporated into both the phases to aid memorization of 4 spatial locations. After each performance, the participant was shown a visual reinforcer and asked for their thoughts on how well they executed the dance. Participants were scored based on timing of the steps as well for exact (1.00), good (0.66), ok (0.33) or missed (0) steps. Data was analyzed using R4.3.1, MATLAB, and Experiment Builder: Data Viewer software. Repeated measures ANOVAs revealed significant main effects of session on both performance accuracy ($F(4, 36) = 6.99$, $p < 0.001$, $\eta^2G = 0.26$), and timing precision ($F(4, 36) = 11.67$, $p < 0.001$, $\eta^2G = 0.25$), suggesting a defined learning curve and consolidation of performance pattern and increase in precision with repeated practice. These findings suggest that visual-motor-based interventions have the potential to enhance motor and neurocognitive functions in neurodegenerative diseases such as Parkinson's Disorder (PD). The results provide a foundation for developing targeted therapies that integrate learning paradigms to improve functional outcomes, warranting further exploration of their long-term efficacy in clinical settings.

43.416 AI-POWERED FEATURE EXTRACTION FROM NATURALISTIC EGOCENTRIC RECORDING AND EYE-TRACKING

Yanbin Xu¹ (bruceyuanbin@gmail.com), Iba Baig^{2,1}, Kevin Li¹, Seiji Cattelain³, Hayato Ono⁴, Sho Tsuji^{3,4}, Ming Bo Cai^{1,4}; ¹University of Miami, ²Northeastern University, ³École normale supérieure, ⁴The University of Tokyo

Children are excellent “self-directed” learners. They learn tasks efficiently by actively sampling information within a complex environment. Modeling the dynamics of children’s naturalistic information-sampling behaviors, such as eye gaze and hand movements, is critical for elucidating how internal states and endogenous rewards guide active learning. Head-mounted eye tracking combined with third-person camera recordings allows developmental psychologists to simultaneously capture children’s naturalistic interactions with their environment and others, along with their fixation targets. Despite data collection advancement, extracting interpretable features from large volumes of video and audio data for computational modeling still relies heavily on time-consuming manual annotation. Therefore, we developed a set of AI-based analysis tools to automate the process of psychologically interpretable feature extraction from multi-modal recordings: (1) a pipeline that temporally aligns multi-device video recordings using audio synchronization with frame-level accuracy; (2) a graphical user interface for semi-automatic segmentation of objects in ego-centric videos from head-mounted eye-trackers; (3) a pipeline for automatically extracting time courses of the participants’ body postures and hand actions from third-person videos utilizing a large vision-language model. The tools are modular and can be applied to diverse data acquisition setups. We validate our pipeline using a newly collected dataset of child–parent interactions in a naturalistic memory task. This dataset includes egocentric video with eye tracking from both child and parent perspectives, and two third-person video recordings from different angles. The result shows that our pipeline’s inference of participants’ poses is near-perfect for adults and has an accuracy ~ 80% for children, while the accuracy for hand action inference is lower, potentially due to occlusion. In conclusion, our tools support moment-by-moment annotation of attended objects, body poses, and actions. These features allow analyzing the dynamics of decision-making of eye-gaze shift and body motion during natural activity of children, including interaction with caregivers.

Eye Movements: Clinical

MONDAY, MAY 18, 8:30 AM – 12:30 PM,

PAVILION

43.417 BINOCULAR RIVALRY AND OPTOKINETIC NYSTAGMUS AS A SCREENING TOOL FOR AMBLYOPIA

Adrien Chopin¹ (adrien.chopin@gmail.com), Preeti Verghese¹, Dennis Levi²; ¹Smith-Kettlewell Eye Research Institute, San Francisco, ²Herbert Wertheim School of Optometry and Vision Science, University of California, Berkeley

Background: Amblyopia is the leading cause of monocular vision loss, yet current screening methods lack the scalability required for mass screening in preschool children. Binocular rivalry is a perceptual alternation that occurs when incompatible stimuli are presented to the two eyes. We have recently demonstrated that static binocular rivalry

dynamics distinguish amblyopic patients from controls. However, this method relies on subjective reports, which are often unreliable in preschool children. Here, we propose a scalable diagnostic tool leveraging motion rivalry dynamics and Optokinetic Nystagmus (OKN) as an objective biomarker of amblyopia. Methods: Adult participants (7 amblyopes, 7 controls) viewed vertical gratings drifting in opposite directions, which elicited motion binocular rivalry and OKN that follows the dominant percept. Participants reported subjective percepts via keypresses while we simultaneously extracted objective percepts from eye-tracked slow-phase OKN velocities. The same participants reported percepts during static rivalry to replicate previous findings. Feasibility for pediatric populations was assessed using a motion rivalry task adapted for young children. Results: First, we replicated the perfect separation of patients and controls using subjective reports in static rivalry, as we previously reported. Second, we confirmed that motion rivalry could achieve the same result. We then compared OKN-derived rivalry metrics, such as alternation rates or eye dominance, to those derived from manual reports, and found high correlations. Crucially, we achieved perfect patient-control separation when based solely on OKN-derived metrics, and using only a single minute of data (F1-score: 100%). Furthermore, a 6-year-old child yielded a high proportion of usable eye-tracking data (77%) in the adapted task, confirming feasibility for pediatric use. Conclusion: Motion rivalry dynamics provide a highly sensitive signature of amblyopia. By utilizing OKN as a proxy for perception, this method bypasses the need for subjective responses, paving the way for a rapid and automated screening tool suitable for preschool children.

43.418 OCULOMOTOR AND PERCEPTUAL ADAPTATION TO A SIMULATED SUB-FOVEAL SCOTOMA

Krishnaveni Nagarajan¹, Martina Poletti²; ¹Department of Brain and Cognitive Sciences, University of Rochester, ²Department of Brain and Cognitive Sciences, Neuroscience, Center for Visual Science, University of Rochester

Foveola, central 1° of the retina, contains the highest cone density and supports high-acuity vision. Our previous work showed that under diffraction limited viewing, subjects establish an alternative preferred locus of fixation even when visual occlusions are limited to a region of just 0.2 deg in diameter around the gaze center. However, it remains unclear whether a similar restructuring of fixation behavior occurs under natural optics and how rapidly such adaptation unfolds. Six emmetropic observers performed a 4AFC orientation-discrimination task with tumbling-E optotypes (1.25 arcmin stroke-width; 20/25 equivalent acuity). Each trial was preceded by a self-paced fixation period on a blank display. A simulated central scotoma, implemented as a flat-top Gaussian mask ($\sigma \sim 2.0$ arcmin) with a fully opaque center (~ 3.8 arcmin diameter), rendered gaze-contingently at the observer’s preferred locus of fixation using a high-precision digital dual-Purkinje image eye tracker. The task was first completed under normal viewing and then repeated across three sessions in the presence of the simulated scotoma. Overall, visual discrimination performance was

reduced by the scotoma (Mean $\pm$ SD: $87.61 \pm 7.89\%$ vs. $69.78 \pm 13.86\%$, $p = 0.031$), but improved systematically over a short time, consistent with short-term functional adaptation ($p = 0.011$). Comparisons of trial-wise fixation centroids revealed significant, idiosyncratic shifts in gaze position for all participants ($p < 0.05$), with 68% gaze-position contour centroids displaced by ~ 4 arcmin between baseline and scotoma condition. These shifts emerged already during the initial self-paced fixation period, confirming the recruitment of an alternate retinal locus of fixation, right outside scotoma's opaque zone. These findings indicate that even subtle disruptions of central vision, as small as 0.2° are sufficient to alter both high-acuity performance and fine oculomotor behavior under normal viewing conditions, and that observers can begin to reorganize fixation patterns over just a few experimental sessions.

This work was funded by NIH Grant EY029788 to MP and NIH grant EY001319 to the Center for Visual Science

43.419 EXPLORING EYE-HAND COORDINATION WITH CENTRAL FIELD LOSS IN VIRTUAL REALITY

Jade Guénot¹, Preeti Verghese¹; ¹Smith-Kettlewell Eye Research Institute

Eye-hand coordination relies on both accurate depth perception and gaze control, which are impaired in central field loss (CFL). To better understand how CFL affects eye and hand control during dynamic 3D tasks, we developed a realistic VR experiment using the HTC Vive Pro Eye headset and the PTVR toolbox. Participants (4 with CFL, 9 controls) tracked a butterfly moving unpredictably at an average speed of 0.5m/s, with random changes in speed and direction to prevent anticipatory strategies. Three conditions were tested: eye-tracking only (following the butterfly with gaze), net-tracking (keeping the butterfly centered in a virtual net rendered with the hand controller), and catching (capturing it with the net as quickly as possible). Gaze (eye + head) and hand movements were recorded continuously. Participants completed these conditions monocularly and binocularly, and controls additionally performed them with simulated scotomas. Controls caught the butterfly significantly faster with binocular compared to monocular viewing, with simulated scotomas producing additional slowing. Patients were overall slower, with disproportionately longer catching times in monocular viewing. When misses occurred, they often resulted from underestimating the butterfly's distance. Gaze precision analyses revealed distinct eye-hand coordination patterns. Controls consistently showed slightly better angular precision in the eye-tracking condition than in the net-tracking condition ($1\text{--}2^\circ$ difference across conditions). In contrast, patients showed improved gaze precision during net-tracking binocularly, reducing angular error by 2° on average. Pursuit latency revealed the strongest difference between both groups: patients exhibited significantly reduced latency during net-tracking compared to eye-tracking (85 ms vs 121 ms), whereas controls showed comparable latencies (78 ms vs 70 ms). These preliminary findings suggest that in CFL, hand movements may help guide the eyes,

enhancing gaze precision and reducing pursuit delays. VR offers a powerful framework to quantify these compensatory eye-hand strategies in realistic environments, with potential implications for rehabilitation approaches.

NIH funding R01 EY27390

43.420 CAN OPTICAL COHERENCE TOMOGRAPHY BE USED TO QUANTIFY FIXATION STABILITY?

Rijul Saurabh Soans¹ (rijul.soans@berkeley.edu), Susana T.L. Chung¹; ¹University of California, Berkeley

Fixation stability assessment is important for understanding oculomotor function in many eye disorders. It can be quantified using eyetrackers, microperimeters and scanning laser ophthalmoscopes. These devices are not always available in typical clinical settings, whereas optical coherence tomography (OCT) is more accessible. Can the OCT be used to quantify fixation stability given that it does not collect a lot of "samples" in a single scan? We developed a computer vision framework to test the feasibility of quantifying fixation stability from OCT volume scans. The framework reconstructs scan paths by cross-correlating en-face projections from individual B-scans, with the infrared confocal scanning laser ophthalmoscopy reference image. Planned scan positions were extracted from OCT metadata to calculate deviations from intended acquisition paths. En-face profiles were generated by projecting intensity values from the outer retinal layers of each B-scan. The framework employs multi-orientation Gabor filtering for vessel enhancement, followed by template matching using normalized cross-correlation. It uses adaptive warping to account for compression artifacts caused by eye movements during OCT acquisition. Detected positions were refined using Hampel outlier detection and spline interpolation. To validate the framework, we compared fixation stability using this framework with that derived from a 10-second fixation task using a microperimeter from the right eye of 30 observers with normal vision. Fixation stability was quantified as the 68% isoline area (ISOA) of gaze distributions in both modalities. In general, ISOA_{OCT} was higher than ISOA_{fixation}, by an average of 0.21 deg^2 (statistically different from 0, $p < 0.001$), but there was a significant correlation between the two values (Pearson $\rho = 0.55$, $p = 0.002$). The limits of agreement based on Bland-Altman analysis were: upper= 0.57 deg^2 ; lower= -0.14 deg^2 . These results demonstrate that it is feasible to quantify fixation stability from routine OCT scans without additional hardware or exam time, but OCT may overestimate the absolute fixation stability values.

43.421 FIXATIONAL INSTABILITY IN OPHTHALMIC AND NEUROLOGICAL DISEASE: A COMPREHENSIVE REVIEW

Zainab Faisal², Edrich Silva², Angelina Joseph², Nancy O. Mahfouz², Sophia Castro¹, Ayesha Mulla², Simrat Renu², Robert G. Alexander²; ¹CUNY School of Medicine, ²Department of Psychology & Counseling, New York Institute of Technology, USA

Fixational eye movements (microsaccades, drifts, and tremors) are essential for everyday visual function. Clinical disorders result in detectable abnormalities in eye movement patterns, though these impairments are often not consciously perceived by the patient. This review examines the research on disease-related fixational instability across ophthalmological and neurological conditions. Methods: We conducted a literature review of peer-reviewed articles examining fixational eye movements in both neurologic and ophthalmological diseases. Articles were selected through extensive keyword searches and searches of references cited in retrieved publications. Studies reporting quantitative measures of microsaccades, drift, tremor, or fixation stability were included. Results: Findings across the literature review indicate that ophthalmic and neurological diseases produce measurable disruptions in fixational control. Ophthalmic conditions such as Behçet uveitis, cataracts, amblyopia, macular disease, and glaucoma each show specific changes in fixational behavior, including increased drift, altered microsaccade timing, or wider fixation areas. Some conditions also differ based on disease progression, such as the reduced stability seen in Stargardt disease compared with Best vitelliform macular dystrophy. Neurological disorders show similar disruptions. For example, increased square-wave jerks have been observed in Amyotrophic Lateral Sclerosis, and difficulties maintaining stable fixation have been reported after mild Traumatic Brain Injury. Conclusion: Fixational instability is a useful way to understand how ophthalmic and neurological diseases affect visual function. Measuring changes in microsaccades, drift, and fixation stability can help identify functional problems that are not always visible through standard tests. Continued research may support the use of these measures in clinical settings to better track disease and evaluate treatment.

This work was supported by the National Institute of Health (Award R16GM159810 to RGA), which provided support for student training that contributed to this project. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

43.422 ALTERED EYE MOVEMENTS IN ADULTS WITH GLAUCOMA WHILE WALKING THROUGH DIFFERENT ENVIRONMENTAL FEATURES

Michelle Harter¹ (mjk160@pitt.edu), J. Patrick Mayo¹, Galen Holland¹, Rakié Cham¹; ¹University of Pittsburgh

People with glaucoma experience progressive visual field loss which leads to significant challenges in maintaining mobility, particularly in complex environments. People with glaucoma may be able to partially compensate for their visual field loss by using a broader range of eye movements to bring important environmental features into their remaining visual field. This work sought to identify differences in eye movements between adults with and without glaucoma while walking through environmental features commonly encountered in daily life. Seven people with glaucoma (69±4 years; 6F; visual field mean deviation better/worse eye: -7.9±6/-10.3±5dB) and seven normal

vision controls (63±8 years; 4F) wore eye-tracking glasses (Tobii Pro Glasses 3) while 1. walking straight 6-meter segments, 2. completing 180-degree turns and 3. stepping over obstacles. For each task, eye movement events were identified and measures of fixations (number, average duration) and saccades (average amplitude, scanpath length) were compared between groups using Student's t-tests. Glaucoma participants had shorter average fixation duration during straight walking ($p=0.03$), more fixations in obstacle traversals ($p=0.05$), and smaller saccade amplitudes during turns ($p<0.01$). For all tasks, glaucoma participants had greater scanpath length than normal vision controls ($p\leq0.05$). The decreased fixation duration, increased fixation number, and increased scanpath length in glaucoma participants indicate that they move their eyes more while walking compared to normal vision controls. However, these increased eye movements are not necessarily indicative of an efficient strategy for gathering environmental information. In fact, the reduced saccade amplitude during turns suggests that their eye movements may be concentrated in a smaller area, rather than indicating a broader range of visual scanning. Further analyses of the spatial patterns and sequences of eye movements in people with glaucoma will be critical for developing training paradigms to promote efficient visual scanning that can allow for safe and independent mobility.

Department of Ophthalmology, University of Pittsburgh Hillman Foundation Funds

Action: Reaching

**MONDAY, MAY 18, 8:30 AM – 12:30 PM,
PAVILION**

43.423 DISSOCIATION BETWEEN ENDPOINT JUDGEMENT ACCURACY AND METACOGNITIVE ABILITY ACROSS BRIEF WINDOWS OF VISUAL FEEDBACK IN UPPER-LIMB REACHING

Gabriela Oancea¹, Luc Tremblay¹; ¹University of Toronto

Producing effective goal-directed behaviours requires judging the quality of our actions. In some tasks, natural feedback is insufficient to rate our actions, so we rely on an internal estimate of success. Such metacognitive judgements are critical for skill acquisition because the alignment between confidence and performance can help or hinder skill acquisition. Visual information plays a critical role in supporting metacognitive judgements about voluntary movements (Culot et al., 2025). Indeed, metacognitive ability can vary for judgements made on the basis of information from different stages of an upper-limb reach (i.e., start, middle, and end) (Oancea et al., 2025). Because work on the visual control of movement suggests that windows of visual feedback prior to peak limb velocity are particularly relevant to limb-target regulation mechanisms (Tremblay et al., 2017), this study focused on metacognitive judgements based on visual cues presented

during upper-limb reaches. Participants ($n = 22$) performed goal-directed reaches and received a 20-ms window of vision when real-time limb velocity exceeded one of five selected criteria, which corresponded to 2% to 48% of movement time. After each trial, participants judged their endpoint bias (under- vs. overshoot) and rated their confidence in their judgement. Overall, participants showed above-chance metacognitive ability, as they were more confident on correct than incorrect trials. More importantly, although one would expect metacognitive ability to increase as the movement unfolded, there was a non-linear relationship between metacognitive ability and movement proportion. Additionally, participants were not always able to use the information from the endpoint judgement to accurately rate their confidence. We conclude that a brief window of vision is sufficient for accurate scaling of confidence ratings with endpoint judgement accuracy, but only at specific “optimal” windows that may be related to online visuomotor control mechanisms. Additionally, confidence is not always a direct read-out of endpoint judgement accuracy.

43.424 DISTINCT SENSORIMOTOR INTEGRATION IN REACHING TOWARD RIGID VS. NON-RIGID MOTION TARGETS

Markus Lappe¹, Riccardo Brandolani², Krischan Koerfer¹, Patrizia Fattori², Rossella Breveglieri²; ¹University of Muenster, ²University of Bologna

Reaching toward a moving object requires a complex interplay between different components of the visuomotor system. A novel class of non-rigid motion stimuli, for which certain visuomotor processes such as smooth pursuit are known to fail, offers a tool to study these interactions. Here we investigated how the reaching system integrates visuomotor information in targeting such a non-rigid motion stimulus. Participants fixated on a point while a moving stimulus (a rigid disk or a non-rigid vortex) traveled horizontally from left to right. Each trial began with a button press, followed by stimulus onset. After a variable delay, an auditory cue triggered participants to release the button and reach toward the right edge of the stimulus. In Experiment 1, the stimulus maintained constant velocity (5°/s or 10°/s). In Experiment 2, the stimulus disappeared at reaching onset, leaving only the fixation point on a blank background and requiring participants to reach its remembered position in open-loop fashion. In Experiment 3, the stimulus either continued at the same speed, accelerated, or stopped at reaching onset. All trials ended upon target touch, and sessions were organized into randomized blocks of conditions within each experiment. Both reach planning and execution were consistently impaired when targeting the vortex compared to rigid motion. Reaction times were unaffected by the velocity of the vortex, contrary to what happened when planning a reach for the rigid disk. Memory-guided reaching toward the vortex showed greater undershoots compared to the rigid disk. Visually-guided reaching movements towards the vortex were less accurate than towards the rigid disk, with movement dynamics reflecting impaired positional and velocity encoding during the feedback-driven phase. These findings suggest that the reaching

system does not accurately process this non-rigid motion, similar to the deficiency of the oculomotor system in smooth pursuit.

EU MSCA SE 101086206 – PLACES; Space It Up by ASI and MUR, n. 2024-5-E.0 - CUP n. I53D2400006000; MuWALK”, 2022BK2NPS_001 - CUP: J53D23010900006 by EU-NextGenerationEU (PNRR)

43.425 DISTINCT SPATIOTEMPORAL DYNAMICS OF SALIENCE-BASED SUPPRESSION IN REACHING MOVEMENTS

Shinhae Ahn¹ (a.shinhae@wustl.edu), Richard A. Abrams¹; ¹Washington University in St. Louis

For successful goal-directed actions, it is crucial to prioritize relevant information while filtering out salient-but-irrelevant distractors to guide accurate movements toward a target. Previous studies have shown that when perceptual salience is manipulated by color contrast between a distractor and other items, highly salient distractors can be proactively suppressed more strongly than less salient ones under top-down control, as evidenced by reduced oculomotor capture. However, little is known about how the perceptual salience of distractors modulates the spatiotemporal dynamics of attentional suppression during reaching. The present study investigated the mechanisms underlying the strong suppression of highly salient distractors and examined how these mechanisms differ from those involved in suppressing less salient distractors by analyzing reaching movement trajectories. Participants made rapid mouse-reaching movements to a target in the presence of distractors in two experiments. Distractor salience was manipulated by: (1) varying the color contrast between the distractor and other items (highly salient vs. less salient colors), and (2) varying color-fill status, where search items appeared as fully colored shapes or as colored outlines (filled vs. unfilled shapes). Across both experiments, stronger distractor salience led to greater distractor-presence benefits, with suppression of high color-contrast distractors benefiting movement onset latencies and suppression of filled shapes benefiting movement durations. Interestingly, reaching movement trajectories revealed repulsion away from highly salient distractors, whereas less salient distractors elicited reduced repulsion or even attraction toward the distractors. Overall, the findings show that both the type and magnitude of distractor salience distinctively modulate suppression mechanisms, exhibiting specific spatiotemporal dynamics during reaching movements.

43.426 EFFICIENT ESTIMATION OF REACH ENDPOINT VARIABILITY USING A WISHART-PROCESS MODEL

Zinong Li^{1,2,3} (zl2242@nyu.edu), Fangfang Hong⁴, Alex H. Williams^{5,6}, Robert Volcic^{1,7,8}, Michael S. Landy^{2,3,5}; ¹New York University Abu Dhabi, ²New York University, ³Department of Psychology, New York University, ⁴Department of Psychology, University of Pennsylvania, ⁵Center for Neural Science, New York University, ⁶Center for Computational Neuroscience, Flatiron Institute, ⁷Center for Artificial

Visually guided reaching shows systematic differences in endpoint bias and variability across the workspace. We introduce an efficient approach based on a Wishart-Process model to measure such endpoint distributions. It provides estimates of endpoint distributions at tested and untested locations in the workspace. Nine participants performed unseen-hand, center-out reaches (3 amplitudes $\times$ 24 directions $\times$ 20 trials, 2 hours/participant) using a stylus on a tablet. Participants were encouraged to take 600 ms to complete the reach (ready-set-go procedure). Trials with durations deviating by more than 30% were excluded and re-run. Endpoint distributions were modeled as bivariate Gaussians, with both the bias (the difference between the mean and the target location) and covariance matrix treated as free parameters. These parameters were assumed to vary smoothly across target locations. Each was estimated by maximum likelihood using 10 Zernike-polynomial basis functions, with this number selected by cross-validation to avoid overfitting. Reach-distance bias exhibited a regression to the mean target distance. Reach covariance showed Weber-like growth with increasing target distance and idiosyncratic patterns of bias and covariance across reach space and across participants. To evaluate the potential for increased trial efficiency, we simulated the experiment while varying the number of trials per target location. As a measure of fit accuracy, we used a target-distance-normalized Bures-Wasserstein distance between the ground-truth and fitted covariance matrices. Results showed fit accuracy nearly leveled off at 15 trials/location with half the number of target locations, indicating that 540 trials and less than 1 hour of data collection would be sufficient. Our approach provides an efficient way to profile individual motor variability across a 2D workspace from a relatively modest number of trials. Results can be interpolated to untested target locations. The model can be extended to higher-dimensional problems: 3D reaches, varying duration, final grasp pose, etc.

NIH EY08266 (to MSL), NYUAD Center for Brain and Health, funded by Tamkeen under NYU Abu Dhabi Research Institute award CG012 (to MSL and RV), NYUAD Center for Artificial Intelligence and Robotics, funded by Tamkeen under NYU Abu Dhabi Research Institute award CG010 (to RV).

43.427 ELECTROPHYSIOLOGICAL MEASURES OF RESPONSE INHIBITION DURING “AUTOMATIC” ONLINE REACH CORRECTIONS.

Christopher L. Striemer^{1,2} (striemerc@macewan.ca), Avery S. Hudson¹, Cameron D. Hassall¹; ¹MacEwan University, Edmonton, Alberta, Canada, ²Neuroscience and Mental Health Institute, University of Alberta

Previous research indicates that the visuomotor system contains an “automatic pilot” that can make rapid online corrections to reaches in response to sudden changes in target position. These automatic corrections can occur even when participants are instructed not to

correct (i.e., to inhibit) their movements. Furthermore, recent work suggests the ability to inhibit online reach corrections is related to individual differences in executive function. One important question that remains unanswered is whether one’s intention to correct their reach to a target jump requires the inhibition of the response to the initial target location, the engagement of a second response to the new target location, or a combination of the two. We examined this question in the current study by recording EEG while healthy adults ($n=34$) completed a double-step reaching task on a touch screen where they were instructed to either “correct” their movement to the target jump, or to “ignore” the target jump, and reach to the initial target location. To index response inhibition, we analyzed the anterior N200 (N2) ERP component, and frontal midline theta (4-8 Hz), both of which have been linked with response inhibition and cognitive control in previous studies. Consistent with previous research, participants made significantly more online corrections when they were told to “correct” their reach to the target jump, compared to when they were told to “ignore” the target jump. Interestingly, both the anterior N2 and frontal midline theta were significantly larger in response to target jumps in the “correct” condition, compared to the “ignore” condition. These results are consistent with the notion that, in order to initiate an online correction to a target jump, the movement to the initial target location must first be inhibited before the correction to the new target location can be initiated.

This work was supported through Natural Sciences and Engineering Research Council of Canada Discovery Grants awarded to C.L.S. and C.D.H.

43.428 FROM PERCEPTION TO ACTION: STEREOMOTION CHANNELS PREDICT REACHING DYNAMICS

Pin Yang¹, Geoffrey Bingham¹; ¹Indiana University Blomington

Human motion-in-depth (MID) perception is supported by two stereomotion mechanisms, Change in Disparity Over Time (CDOT) and Inter-Ocular Velocity Differences (IOVD), but whether they serve different functional roles in visually-guided-action remains unclear. We examined whether individuals who rely more on CDOT versus IOVD exhibit distinct visuomotor control strategies during reaching. 200 participants first completed relative-distance discrimination tasks, allowing us to classify them as CDOT-dominant, IOVD-dominant, or Combined observers. Then, 72 participants (24 per group) performed dynamic reaching tasks using random-dot targets and hand avatars with constant size and dot-density, removing monocular information. Dotlife (1/2/8/16 frames, and infinite) was also manipulated to vary the temporal coherence of the MID signal. We found clear dissociations in visuomotor control as a function of stereomotion-channel dominance. Specifically, group effect on movement-time ($F(2, 69) = 11.949$, $p < 0.001$), reaching-accuracy ($F(2, 69) = 10.583$, $p < 0.001$), peak velocity ($F(2, 69) = 8.152$, $p < 0.001$) are all significant. CDOT-dominant observers showed slower velocities and longer movement times, consistent with CDOT’s role as a slow-motion processing

channel. IOVD-dominant observers reached faster but with reduced mapping accuracy, reflecting a velocity-based strategy tuned to rapid yet less spatially precise targeting. Combined observers exhibited a hybrid profile, matching the IOVD group in reaching speed but the CDOT group in spatial accuracy. Dotlife manipulations highlighted these dissociations: Mapping accuracy in the CDOT group remained unaffected across dotlife conditions ($F(4,92) = 1.94$, $p = 0.11$); the IOVD group showed a significant improvement in accuracy from 1-frame to 2-frame dotlife, with no additional gains at longer dotlives ($F(4,92) = 6.4$, $p < 0.001$). Together, these findings show that stereomotion-channel specialization influences both MID perception and visuomotor strategy used to visually guide action, underscoring the need to account for online guidance and individual differences in visuomotor models.

43.429 VARIABILITY OF VISUALLY GUIDED REACHES DEPENDS ON REACH DISTANCE AND DURATION

Qingjie Sun¹, Zinong Li², Michael S. Landy^{1,3}; ¹Department of Psychology, New York University, ²New York University Abu Dhabi, ³Center for Neural Science, New York University

Accurate goal-directed reaching requires humans to coordinate movement speed, distance, and duration while maintaining endpoint precision. We have found that movement plans are near-optimal for reaches under risk with incentives and penalties for reach outcome and a short duration limit (e.g., Trommershäuser et al., *Spatial Vision*, 2003). But, to determine whether optimality holds for reach planning in general requires a model of reach uncertainty across target locations and reach durations. Here, we measured endpoint variability as a function of movement duration and distance. Twelve right-handed participants performed unseen-hand reaches on a digitizing tablet. At trial onset, participants remained at the start point for 0.5 s until a green target circle appeared and began shrinking at a constant rate. Participants were instructed to reach the target before it disappeared. Endpoint feedback and a score that decreased linearly with endpoint error were displayed after each trial. The experiment consisted of three blocks defined by different target-disappearance durations (0.6, 0.79, 1.04 s). Each block included three target distances (overall range: 57–172 mm), scaled to keep movement speeds relatively consistent. Target direction was chosen uniformly ($\pm 15^\circ$) relative to leftward or rightward. Reach uncertainty (SD) was fit reasonably well as a linear function of reach distance and duration, with moderate positive slopes for each. Fitts' Law predicts the movement time a human will use as a function of reach distance and target size. One hypothesis is that Fitts' Law derives from the amount of uncertainty as a function of reach distance and duration, i.e., from an attempt to maintain a fixed percentage of target hits. Our data are inconsistent with that hypothesis.

NIH EY08266, NYUAD Center for Brain and Health, funded by Tamkeen under NYU Abu Dhabi Research Institute award CG012

43.430 VISUAL LANDMARKS MODULATE POSTERIOR PARIETAL ACTIVITY DURING A MEMORY GUIDED REACH TASK

Brando Sheldrick¹, Veronica Nacher¹, Jennifer Lin¹, Hongying Wang¹, Saihong Sun¹, Xiaogang Yan¹, Douglas Crawford¹; ¹York University

Neurons along the intraparietal sulcus (IPs) are known to encode visual targets in egocentric (e.g., eye-centered) coordinates, but the influence of allocentric (e.g. landmark-centered) cues is less studied. The purpose of the current study was to investigate how local field potential (LFP) activity along the mid-posterior IPs is modulated by stable or shifting visual landmarks before and during reaches to remembered visual targets. To investigate this question, we recorded LFPs from putative lateral, medial, and ventral IPs sites using 32-channel Plexon probes in two female Rhesus monkeys. Simultaneous array recordings in posterior ventrolateral prefrontal cortex (pvLPFC) were reported elsewhere (Lin et al. SFN 2025). A landmark (four identical dots positioned at the vertices of a virtual square) was displayed at one of fifteen locations within reach on a touch screen. A visual target then appeared, either within or outside of the landmark square, followed by a visual mask. After the mask disappeared, the landmark either reappeared at the same location (stable-landmark condition) or shifted by 8 degrees in one of eight possible directions (landmark-shift condition). Gaze and head position were allowed to move freely, but animals were rewarded for reaching within 4.7 cm of the target. Control trials employed the same procedures without the landmarks. For comparison, pvLPFC LFPs showed low-frequency modulations during the memory, planning and execution phases, with additional low-frequency synchronization after reaches in the landmark conditions (Lin et al., SFN 2025). Preliminary analysis of parietal LFP data in the first monkey suggests higher-frequency synchronization in the alpha- and beta-bands prior to target onset that fades during the hand movement in all visual conditions. Additionally, there was again low-frequency (delta and theta) synchronization after reach/before reward in the landmark task conditions compared to no-landmark controls. The latter might suggest visual feedback of reach errors relative to the landmark.

The Connected Minds Program, funded in part by the Canada First Research Excellence Fund.

Action: Grasping, affordances

MONDAY, MAY 18, 8:30 AM – 12:30 PM,
PAVILION

43.431 APRAXIA DISRUPTS THE SPATIOTEMPORAL STRUCTURE OF GRASPING, DESPITE PRESERVED APERTURE SCALING

Felicia Tassone¹ (feliciat@yorku.ca), Elisabeth Rounis^{2,3,4}, Fatemeh Geranmayeh², Matthew Banger², Alison McGregor², Erez Freud¹;

Limb apraxia is a neurobehavioural consequence of stroke, characterized by impaired skilled action that cannot be attributed to basic motor impairments. It has traditionally been thought to selectively disrupt higher-level action representations while preserving basic visuomotor control. However, this view is supported largely by peak-derived metrics, such as maximum grip aperture (MGA), which may overlook potential deficits in the fine-grained spatiotemporal structure of movements. To address this gap, we assessed 17 patients with clinically-confirmed apraxia and 19 age-matched controls on an Efron-block width-scaling task (four objects, 5mm increments) involving reach-to-grasp movements (Experiment 1). We also examined 25 younger and 25 older healthy adults to determine whether any observed deficits were stroke-specific rather than age-related (Experiment 2). Consistent with previous findings, apraxic patients successfully scaled their MGA with object size, similar to controls. However, patients exhibited significantly slower movements and prolonged time to MGA. To assess the spatiotemporal structure of the movements, Dynamic Time Warping (DTW) was used to measure the similarity between trajectory pairs across object sizes. The observed similarity matrix was then correlated with an ideal object size-based model. Although both patients and controls exhibited strong correspondence with the size model, apraxic patients showed significantly weaker correspondence compared to controls, even when using their ipsilesional hands. Bayesian analysis of Experiment 2 supported the null hypothesis, revealing that young and older adults showed equivalent object size-based trajectory organization patterns. Furthermore, both healthy age groups demonstrated stronger correspondence with the size model than apraxic patients' ipsilesional hands. These findings uncover novel deficits in apraxic reach-to-grasp behaviour. While object-related aperture scaling remains intact, the spatiotemporal organization of trajectories is disrupted: a deficit that is distinct from healthy aging, and an impairment previously under-characterized in apraxia.

43.432 ASSESSING THE ROLE OF BINOCULAR VISION IN EYE-HAND COORDINATION DURING SEQUENCING ACTIONS

Xiaoxin Chen¹, Benjamin Thompson^{2,3,4}, Daphne Maurer⁵, David Shore^{5,6}, Xiaoqing Gao⁷, Agnes Wong⁸, Haotian Lin⁹, Ewa Niechwiej-Szwedo¹; ¹University of Waterloo, ²School of Optometry and Vision Science, University of Waterloo, Canada, ³Centre for Eye and Vision Research, Hong Kong, ⁴University of Auckland, Liggins Institute, Auckland, New Zealand, ⁵Department of Psychology, Neuroscience & Behaviour, McMaster University, Canada, ⁶The Multisensory Mind Inc., Ontario, Canada, ⁷Center for Psychological Sciences, Zhejiang University, China, ⁸Department of Ophthalmology and Vision Sciences, The Hospital for Sick Children, and the University of Toronto, Canada, ⁹Zhongshan Ophthalmic Center, Sun Yat-Sen University, China

Binocular viewing supports efficient visuomotor task performance. Monocular viewing causes slower movement and less accuracy. The present study investigates gaze behaviour during monocular viewing during a bead threading task, which consisted of reaching, grasping, transporting, and threading a small bead onto a vertically mounted stationary needle. Participants (n=33, 19-40 [average 29.5] years old; 19 females, normal vision) completed this task under binocular, left eye and right eye viewing conditions, 15 trials each, in random orders. Monocular conditions were averaged together. An Optotrak motion capture system recorded finger movements and an EyeLink 2 eyetracker recorded gaze position. Measures of interest included fixation duration and the latency difference between eye and hand movements. Binocular advantage was quantified as: (monocular viewing - binocular viewing)/(binocular viewing). During binocular viewing, the hand movement initiated later than the eyes when reaching (120ms) and simultaneously when returning (1ms). During monocular viewing condition, the hand initiated significantly later (reaching: 144ms, p=0.006; returning: 22ms, p=0.047). Despite slightly longer planning, actions were still less efficient under monocular viewing, with a 6% binocular advantage in reach duration, 7% in peak velocity, 115% in grasp duration, and 64% in thread duration. The eyes needed to fixate longer on the bead during grasping (binocular 214ms vs monocular 414ms; p<0.001), and on the needle during threading (binocular 374ms vs monocular 588ms, p<0.001). These results indicated that the slightly longer planning during monocular viewing was insufficient to support efficient task completion. The prolonged fixations suggest inefficient encoding of object features to guide task performance due to monocular viewing. Binocular input is critical for the efficient performance of manipulation tasks. Together, these data indicate that gaze behaviour during the bead threading task is sensitive to changes to viewing conditions and may be suitable for investigating gaze behaviour and visuomotor performance in patient populations with binocular deficits.

This study was funded by a Canadian Institutes of Health Research (CIHR) grant.

43.433 INTRINSIC OBJECT FEATURE INTERFERENCE REVEALS HOLISTIC VISUOMOTOR PROCESSING IN AUTISM

Zoha Ahmad¹, Ilana Tulbovich¹, Batsheva Hadad², Erez Freud¹; ¹York University, ²University of Haifa

A central prediction of the Two Visual Systems framework is that perception integrates multiple object features, whereas visually guided actions rely on analytic extraction of task-relevant dimensions. Our previous work showed that this functional dissociation is reduced in autism, with autistic grasping influenced by extrinsic contextual information such as spatial illusions and stimulus-history statistics. Here, we examined whether this reduced specialization extends to intrinsic object structure. Autistic (n=28) and non-autistic (n=28) adults perceptually estimated and grasped objects varying orthogonally in their width (relevant) and height (irrelevant). In the perceptual task,

both groups showed a clear Height–Width illusion, demonstrating that the irrelevant height dimension biased size judgments. However, only autistic participants showed this interference in grasping, whereas non-autistic participants scaled their grip solely to width, preserving the expected analytic dissociation between perception and action. Full-trajectory analysis using Dynamic Time Warping (DTW) revealed two additional signatures of atypical visuomotor computation in autism: (1) non-autistic trajectories clustered by the relevant width dimension, whereas autistic trajectories clustered by object identity, indicating a holistic encoding of shape; and (2) autistic participants exhibited reduced within-condition consistency, reflecting diminished visuomotor efficiency. This pattern is consistent with broader evidence for reduced cortical specialization in autism, suggesting diminished functional differentiation between perceptual and visuomotor systems. Together, these findings extend prior contextual effects to intrinsic object structure and demonstrate that autistic visuomotor control is driven by holistic, perception-like representations rather than analytic, dimension-based processing.

43.434 CO-ACTION ACCELERATES MOVEMENTS AND TRIGGERS TASK-DEPENDENT SYNCHRONIZATION

Mario Costantino¹, Erez Freud¹; ¹Department of Psychology, York University

Daily activities, such as eating with a peer or walking in a crowd, inherently unfold within a social context. One mechanism thought to promote synchrony between individuals is the formation of shared task representations, in which people implicitly encode elements of a partner's actions alongside their own. While previous research has extensively examined deliberate coordination driven by competition or collaboration, far less is known about the spontaneous, uninstructed synchronization that emerges when actors pursue independent goals side-by-side. Here, we examined how social context modulates visuomotor control and whether co-action promotes alignment even in the absence of an explicit shared goal. Pairs of young adults completed a reach-to-grasp task under three conditions: acting individually, acting concurrently with identical objects (congruent), and with different-sized objects (incongruent). In paired conditions, partners sat face-to-face and could observe each other's movements. Motion-tracking technology captured hand kinematics. To assess the fine-grained spatiotemporal structure of the movement trajectories and quantify interpersonal synchrony, we applied Dynamic Time Warping (DTW). DTW preserved temporal structure across partners, allowing us to quantify trajectory similarity independent of speed differences, overcoming limitations of time-averaged metrics. Results revealed two effects of co-action. First, movement time and time-to-maximum grip aperture were faster during co-action than during individual performance, regardless of object size. Second, co-action altered the structure of participants' movements. During the paired condition, lateral hand trajectories became more similar than in the individual condition, indicating that co-action induces spontaneous alignment even when partners pursue independent goals. Additionally,

synchrony in grip-aperture trajectories emerged when partners grasped identical objects. Grip-aperture profiles converged more in the congruent than in the individual condition, reflecting enhanced alignment when task parameters were matched. These findings indicate that co-action generates a baseline tendency toward synchronous movement, and that shared task parameters amplify this coordination beyond the effects of co-presence alone.

43.435 HAPTIC CONTRIBUTIONS TO VISUAL-MEMORY-GUIDED GRASPING

Logan McIntosh^{1,3} (Lmcintosh@nyu.edu), Robert Volcic^{1,2,3}; ¹New York University Abu Dhabi, ²Center for Artificial Intelligence and Robotics, New York University Abu Dhabi, ³Center for Brain and Health, New York University Abu Dhabi

There is a large body of research examining how reaching and grasping movements to visually-previewed locations change under memory conditions. However, not all movements are made purely under visual memory, such as when unscrewing the lid of a jar. In this case, the hand holding the jar might provide haptic information that would supplement the initial visual preview. Additionally, the studies on visual-memory-guided movements also used discrete delay timings for movement initiation, typically 0 and 2 seconds (and/or 5 seconds), so it is unclear how performance changes between these delays. Here, participants completed movements to grasp objects of different sizes, presented in different locations, with their right hand. These were completed in three conditions; haptic-only, vision-only, and visuo-haptic. In the vision-only condition, participants were provided a 1 second viewing time before vision was again obscured. In the visuo-haptic condition, the viewing time began once the object was grasped by the left hand. Grasping movements were initiated by a tone, which was presented at a random delay of up to 2 seconds after obscuration, increasing the extent to which the movement depended on memory-based information. In the haptic-only condition, participants' vision was obscured for the entire block and they first moved their left hand to grasp the object for 1 second before the starting tone. Maximum grip aperture, peak velocity, and duration were recorded as measures of performance. We replicated previous results, with better performance under visuo-haptic reaching than vision-only, which was again better than for haptic-only grasping. Performance decreased with longer delays between obscuration and movement initiation, with a slower decline in the visuo-haptic condition than in vision-only.

We acknowledge the support of the NYUAD Center for Artificial Intelligence and Robotics and the NYUAD Center for Brain and Health, funded by Tamkeen under the NYUAD Research Institute Awards CG010 and CG012.

43.436 OBJECT REPRESENTATIONS IN THE HUMAN PARIETAL AND OCCIPITO-TEMPORAL CORTICES IN A GRASPING TASK

Amirali Soltani Tehrani¹ (aasolte@udel.edu), Kristin Woodard², Emalie McMahon², Leah Ettensohn², Natalia Pallis-Hassani², John

Ingeholm², Leslie Ungerleider², Chris Baker², Maryam Vaziri-Pashkam^{1,2}; ¹Movement and Visual Perception lab, Department of Psychological and Brain Sciences, University of Delaware, ²Laboratory of Brain and Cognition, National Institute of Mental Health

The human visual system supports a broad range of behaviors toward objects, from recognizing them to performing actions on them. Object representations have been found in both the human occipitotemporal and parietal cortices, yet we have a limited understanding of how these representations support such diverse behaviors. In this fMRI study, we focused on two behaviors—grasp movements and object similarity judgments—and investigated how the representations in parietal and occipitotemporal cortex relate to each behavior. In behavioral experiments, we used a set of 64 3D-printed objects and measured two behaviors: an online odd-one-out task in which participants selected the most distinct item from sets of three object pictures, and hand-kinematic recordings as participants grasped and held the objects. In the fMRI experiment, participants viewed and grasped the same 3D-printed objects. The objects were placed on a table, and participants viewed them through a mirror for two seconds; eight seconds later they were instructed to grasp and hold the object, allowing us to separate motor-related responses from visual responses. Two regions of interest—the lateral occipital cortex (LOC) and the anterior intraparietal area (AIP)—were examined, and a representational similarity analysis (RSA) was performed to evaluate how their response patterns related to the two behaviors. During the visual phase, LOC responses were more strongly related to the odd-one-out behavior, whereas AIP did not show this similarity pattern. During the grasp phase, AIP patterns related to both the grasp and the odd-one-out behaviors, while LOC did not show this similarity pattern. These findings highlight the distinct roles of object-selective regions across the visual cortex, with LOC encoding semantic qualities independent of action and AIP expressing both semantic and grasp-based features, suggesting that parietal regions simultaneously encode visual and motor representations to support coordinated actions toward objects.

43.437 STEREOPSIS, BUT NOT MOTION PARALLAX, MODULATES THE SPATIOTEMPORAL STRUCTURE OF GRASP MOVEMENTS

Nathaniel Goldstein¹ (nathany@my.yorku.ca), Laurie M. Wilcox¹, Erez Freud¹; ¹York University, Toronto, Canada

Humans rely on visual depth cues in our day-to-day lives, however the extent to which specific sources of depth information are involved in reach-to-grasp actions is not clear. There is evidence that binocular depth cues (vergence, stereopsis) improve grasping efficiency; however, this conclusion is typically based on comparisons of discrete timepoints from grasping sequences (e.g., the maximum grip aperture). While valuable, these measurements overlook potentially rich information from the full reach-to-grasp trajectory. To better understand the role of depth cues in grasping, we compare the spatiotemporal structure of grip aperture trajectories under two viewing

modalities: binocular and monocular, with and without motion parallax. In two experiments (n=30 each) participants grasped 3D-printed discs and moved them to a nearer or further location. In Experiment 1 motion parallax was available via natural head movements. In Experiment 2 the head was stabilized on a chinrest, thereby limiting parallax. Consistent with previous studies, in both experiments we found that kinematic attributes like movement time and wrist velocity were modulated by viewing modality. To compare the trajectories across conditions, we used Dynamic Time Warping (DTW) to compute dissimilarities between all grip aperture trajectories and built participant-level representational dissimilarity matrices (RDM). We then conducted a representational similarity analysis (RSA) to compare empirical RDMs to a theoretical model encoding only viewing modality. Our results showed a significant correlation between participant-level dissimilarity matrices and the viewing modality model, confirming that stereopsis shapes grasping dynamics. Also, binocular trajectories were more consistent than monocular trajectories. In both cases the availability of motion parallax did not impact these result patterns. Our work confirms that under these natural grasp conditions, motion parallax does not compensate for the lack of stereopsis. Further, our analysis provides new evidence that stereoscopic information modulates not just specific timepoints, but the overall structure of grasping aperture sequences.

Natural Sciences and Engineering Research Council of Canada (NSERC); CFREF Vision: Science to Applications (VISTA); CFREF Connected Minds (CM)

43.438 EFFECTS OF ILLUMINATION ON AFFORDANCE JUDGMENTS WITH SIMULATED PERIPHERAL VISUAL FIELD LOSS IN VIRTUAL REALITY

Maisha Tahsin Orthy¹, Hunter Finney¹, Jeanine Stefanucci¹, Bobby Bodenheimer², Sarah Creem-Regehr¹; ¹University of Utah, ²Vanderbilt University

People with low vision, those with uncorrectable vision loss who rely on residual vision, represent a rapidly growing global population. Peripheral visual field loss (PVL) is caused by conditions such as retinitis pigmentosa and glaucoma, which can impair spatial layout perception, obstacle detection, and locomotor guidance. Peripheral vision is essential for action-related tasks as well as affordance perception. We found previously that simulated PVL led observers to underestimate their perceived capabilities to step over gaps relative to no vision loss. However, low vision is also sensitive to contrast-related deficits, especially under low illumination, which may compromise the detection of boundaries or terrain features necessary for safe movement. The present research investigated whether environmental illumination modulates gap-affordance judgments under simulated PVL. Twelve participants with normal vision judged their gap-stepping affordances in virtual reality (VR) while wearing an eye-tracked head-mounted display that simulated PVL (tunnel vision), resulting in a field of view of about 26 degrees. Participants viewed an urban scene with gaps on the ground ranging from 0.55 to 1.6 m presented randomly

under blocks of trials in bright and low illumination. For each trial (24 total in each block), participants made a binary yes/no judgment, rated confidence on the judgment, and their response time was recorded. Gap width judgments were scaled to individual leg length and modeled using binomial logistic regression. Contrary to our hypothesis, lighting conditions did not significantly affect affordance judgments; participants showed similar transition thresholds and confidence ratings in bright and low light. However, response times were slower in the low-light condition, suggesting increased perceptual or decisional uncertainty that did not translate into altered action boundaries. These results indicate that simulated PVL may reduce overall visual efficiency under dim light without affecting perceived affordances. Future work will test clinical low-vision populations and explore additional environmental variables.

43.439 INVERSE DYNAMICS INDUCED VECTOR FIELDS EXPLAIN OBJECT AFFORDANCES

Aalap D. Shah¹ (aalap.shah@yale.edu), Ilker Yildirim¹; ¹Yale University

Gibson's proposal conceptualized affordances as 'action possibilities' jointly determined by the agent's capabilities and the structure of the physical environment. For instance, the affordance of 'pick-ability' applies to a human and an apple, but not to an ant and an apple. Despite the large body of empirical work accumulated in the half-century since its introduction and the several competing conceptual frameworks attempting to organize these findings, a computational specification of affordances remains elusive. This knowledge gap poses a central challenge for vision science and related fields: Are affordances computable? Here, we present the first computational account of affordance, formalizing it as goal-conditioned, dynamics-based vector fields over objects. Given an object, the physical environment, and a desired object trajectory, we derive a vector field over the object via 'inverse dynamics'. Subject to the constraints imposed by the agent (biomechanics of opposable digits and non-slipping contacts), the object (shape and mass), and the environment (surface configurations), this yields a distribution over contact points between the object, agent, and the environment, and their associated force vectors for reproducing the desired trajectory. We empirically evaluated our account via an experiment in which participants viewed an object moving along a trajectory across an environment of rigid surfaces and subsequently indicated two grasp locations on the object — one for the thumb and one for the index finger — to reproduce the depicted motion. Notably, each object-environment pairing was associated with multiple distinct trajectories. The results were clear and striking: participants systematically and flexibly adjusted contact points in response to changes in the objects, trajectories, and the configuration of surfaces; and their performance aligned with the predictions of our inverse-dynamics model. These results demonstrate that affordances can be quantitatively determined, opening a path to synergize affordances within the computational frameworks of visual cognition and physical interaction.

National Science Foundation (under CAREER Award No. BCS-2441520)

Motion: Illusions

MONDAY, MAY 18, 8:30 AM – 12:30 PM,
PAVILION

43.440 TASK-IRRELEVANT MOTION STRENGTHENS MOTION SILENCING OF COLOR CHANGES

Hyerim Kim¹, Oakyoon Cha¹; ¹Department of Psychology, Sogang University

Motion silencing is a visual illusion in which feature changes in small elements become less noticeable when they move together as a group. Previous studies on motion silencing typically presented task-relevant elements whose features changed, and participants reported those changes. The current study asked whether adding task-irrelevant, moving elements can strengthen the motion-silencing effect on the task-relevant elements. This question is partially motivated by Wu and colleagues (2024), who showed that motion embedded in the background can alter the strength of motion silencing. Participants viewed two arrays to the left and right of fixation. Each array contained small visual elements that moved either upward or downward at individually sampled speeds. The average motion speed was matched across the two arrays. On each trial, participants judged which array contained elements whose colors alternated between yellow and blue. In the other, comparison array, elements had static colors randomly sampled along the yellow–blue continuum. We indexed the strength of motion silencing by reaction time (RT), as longer RT indicates greater difficulty in detecting color changes. Thin horizontal bars were added to both arrays in two experimental conditions. In the Bars-In condition, the bars were attached to and moved in synchrony with the task-relevant target elements, whereas in the Bars-Out condition, the bars moved independently of the targets. In the No-Bars condition, the arrays contained no additional elements. The motion-silencing effect was stronger in both the Bars-Out and Bars-In conditions than in the No-Bars condition, with no difference between the Bars-Out and Bars-In conditions. These results suggest that additional motion signals within the display enhance the motion-silencing effect, regardless of whether they originate from the elements undergoing feature changes.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT) (No. RS-2023-00211668).

43.441 PERCEPTUAL GROUPING REDUCES PERCEIVED SPEED BUT DOES NOT CHANGE UNCERTAINTY

Alejandro G. Garcia^{1,2} (aggarcia@yorku.ca), Chi T. K. Dao^{1,2}, Pascal Mamassian³, Peter J. Kohler^{1,2}; ¹Department of Psychology, York University, ²Centre for Vision Research, York University, ³CNRS & École Normale Supérieure

When four dot pairs rotate about the vertices of an implied square, they are initially perceived as four independently rotating pairs: the “local percept”. Soon, however, observers begin to perceptually group the dots into two overlapping illusory squares translating along a circular trajectory: the “global percept” (Anstis, 2003). Previous research has demonstrated that the global percept appears to move slower than the local percept (Kohler et al., 2009). This so-called “global slowdown” cannot be explained by differences in perceived size, emergent rotation, or number of perceived moving elements between the two percepts (Kohler et al., 2014). Here, we test the possibility that the effect arises from systematic differences in sensory uncertainty of speed estimates between the two percepts. If the global percept is associated with higher uncertainty, a Bayesian prior toward slow motion (Weiss et al., 2002) would lead to a reduction in perceived speed. As in previous work, we used a 2AFC task where participants compared the speed of two configurations presented on either side of fixation. The reference configuration always moved at a fixed speed, while the test configuration took on a range of different predetermined speeds across trials. We measured the slowdown effect by comparing a global-biased reference configuration to a local-biased test configuration and computing the Point of Subjective Equality. To assess uncertainty, we measured the slopes of the psychometric functions obtained from control conditions in which both reference and test configurations were either local-biased or global-biased. We found no significant difference in the slopes between the two percepts, indicating comparable levels of uncertainty. We replicated the global slowdown effect but found no systematic relationship between size of the effect and local-global differences in uncertainty. This suggests that a Bayesian prior for slow motion is not responsible for the global slowdown effect.

This research was undertaken thanks in part to funding from the Vision: Science to Applications, supported by the Canada First Research Excellence Fund. Additional support was provided by a Natural Sciences and Engineering Research Council of Canada Discovery Grant awarded to PJK.

43.442 FOVEAL LOCALIZATION BIAS EXPLAINS THE FLASH-GRAB EFFECT'S DIRECTIONAL ASYMMETRY, BUT NOT THE FLASH-LAG'S

Jye Marchant¹ (jye.marchant@sydney.edu.au), Jiahua Hui, Alex O. Holcombe¹; ¹The University of Sydney

Evidence from individual differences suggests that different motion-position illusions are mediated by distinct mechanisms (Cottier et al., 2023). To probe these apparent processing differences, we have been investigating directional asymmetries. The flash-lag effect, where a flashed object is perceived to lag behind a moving one despite spatiotemporal alignment, is larger for motion toward the fovea compared to away (Kanai et al., 2004; Mateeff & Hohnsbein, 1988; Shi & Nijhawan, 2008). Here we investigated the flash-grab effect (FGE), in which the perceived position of a flash coinciding with the reversal of a moving background is displaced in the direction of post-flash

motion. Our study used post-flash motion either toward or away from the fovea. Toward-fovea post-flash motion produced a robust displacement toward fixation, consistent with the classic FGE. Unexpectedly, away-from-fovea post-flash motion also resulted in a toward-fixation shift (a negative FGE), although the effect was small. As a control for general flash mislocalization bias, we also included a condition with no background motion; here too, flashes were mislocalized toward fixation, and after subtracting this static bias from the motion conditions the difference between toward- and away-from-fovea motion in the FGE was eliminated. This suggests a dissociation with the flash-lag effect, as Shi & Nijhawan (2008) found that flash mislocalisation cannot fully account for the flash-lag effect's directional asymmetries. This difference may reflect processes that contribute to the flash-lag effect but not the FGE. Unlike the flash-lag effect, the FGE does not require temporally binding the flash to a specific position in the trajectory of the moving object, so the FGE may be better for isolating the effect of motion on position perception.

43.443 INDIVIDUAL DIFFERENCES IN THE PERCEPTION OF A MOTION-BASED BISTABLE STIMULUS BASED ON MULTIPLE AXES OF ROTATION

Brooke Lim¹, Anna Kosovicheva^{2,3}, Keisuke Fukuda^{2,3}; ¹McGill University, ²University of Toronto Mississauga, ³University of Toronto

Bistable perception occurs when an ambiguous stimulus triggers spontaneous alternations in perception, and these switch rates have been shown to vary across individuals. Moreover, differences in switch rates are correlated across certain classes of stimuli, indicating shared mechanisms (Cao et al., 2018). However, for bistable stimuli based on rotational motion, little is known about how individual differences in switch rates change with alterations in rotational axes. Here, participants viewed a bistable structure-from-motion sphere that rotated around either the horizontal or vertical axis on each trial. One dot, shown in a different colour, served as a perceptual landmark to examine whether perceptual switches aligned with specific spatial locations. Participants continuously indicated their perceived rotation direction via keypress while EEG was recorded. Behaviourally, participants' perceptual switches were more frequent when the landmark approached the edges of the sphere than its center for both directions ($p < 0.001$). Interestingly, horizontal-axis rotation induced more frequent switches than vertical-axis rotation ($t(24) = -3.47$, $p < 0.01$). Additionally, those who experienced more switches in one direction also did so in the other ($r = 0.57$, $p < 0.01$), indicating a common mechanism of bistable perception across two rotation axes. Electrophysiologically, we successfully decoded the landmark's physical position ($p < 0.001$) and the perceived direction of rotation ($p < 0.05$) above chance in both directions. Furthermore, decoding performance for landmark location was positively correlated across rotation axes, such that individuals who showed higher decoding accuracy in one direction also showed higher decoding accuracy in the other direction ($r = 0.87$, $p < 0.001$). Similar results were found for decoding of perceived rotation ($r = 0.52$, $p < 0.01$), thus providing

converging support for the common mechanism of bistable perception across two rotation axes. Overall, these findings suggest that the stability of bistable stimuli is consistent across axes.

43.444 VVIQ PREDICTS THE VIVIDNESS BUT NOT THE CONTROL OF ILLUSORY APPARENT MOTION.

Julia McClellan¹ (juliamcc@bu.edu), Allison Allen², Nicolas Davidenko²; ¹Boston University, ²University of California, Santa Cruz

In Illusory Apparent Motion (IAM; Davidenko et al., 2017), randomly changing pixel arrays, refreshing at around 1.5Hz, give rise to illusory percepts of coherent apparent motion that can be perceptually controlled, for example by deliberately changing one's perception of the direction of motion (Allen et al., 2022). However, mechanisms underlying the marked individual differences in the ability to perceive and control IAM have yet to be explored. Previous work has shown that perception can be biased in favor of imagined stimuli, especially in individuals with more vivid visual imagery (Keough and Pearson, 2018; Dijkstra et al., 2021). The current study examined whether observers with more vivid visual imagery would be better at perceiving and controlling IAM. In an IAM task, participants (n=41) were presented with 8-second trials of IAM stimuli and asked to perceptually control the motion to move congruently with a previously shown vertical or horizontal arrow. After each trial, participants reported whether they successfully controlled the motion and provided a 1-5 rating of the motion's vividness. We assessed three measures of participants' mental imagery vividness by measuring participants' (a) pupil size as they imagined bright or dark shapes, (b) performance in a mental rotation task, and (c) scores on an extended version of the VVIQ (Marks, 1973) that included 8 additional questions related to motion imagery (mVVIQ). We found that VVIQ and mVVIQ strongly predicted participants' vividness of IAM ($r = 0.55$, $p < .001$ for VVIQ; $r = 0.58$, $p < .001$ for mVVIQ) but did not predict participants' ability to control IAM ($r = 0.05$, $p > 0.5$ for VVIQ; $r = 0.05$, $p > 0.5$ for m-VVIQ). Our results suggest that individual differences in self-reported vividness of visual imagery can account for the vividness of IAM, but not the ability to control it.

43.445 RADIAL ROTATION INDUCED RIPPLES: A NEW MOTION AFTEREFFECT

Gideon Caplovitz¹; ¹University of Nevada Reno

Radial gratings can induce all sorts of perceptual effects like Mackay's Chrysanthemum Rays or if you move your eyes: The Sunburst Illusion. If you rotate radial gratings on a computer monitor things get really weird with effects ranging from shape distortions, wagon-wheel type motion reversals and very strong spiraling motion after effects when they stop. I will demonstrate that when they stop rotating radial gratings also produce a novel dynamic illusion that manifests as a brief sequence of circular rings that rapidly expand from the near center out into the periphery, like the ripples produced by throwing a rock in a pond (but faster!) Even though the inducing stimulus is achromatic, this "discharge" is commonly reported as being yellowish but can

appear in different hues to different observers. Like Laviant's Enigma or the Mackay Rays, the effect is disembodied from the inducing stimulus. Intriguingly, once experienced, some observers report seeing stationary chromatic rings superimposed on the radial grating even when it is not rotating, although we have not tested how long this effect lasts. In a series of experiments we document effects of radial and temporal frequency on the presence, hue and persistence of these ripples.

Decision Making: Actions, metacognition

**MONDAY, MAY 18, 8:30 AM – 12:30 PM,
PAVILION**

43.446 ATTENTIONAL OFFLOADING DURING A MULTI-DIMENSIONAL DISCRIMINATION TASK

Jenna Donet¹, Joe Opdenaker¹, Philip Marshall¹, Miranda Scolari¹; ¹Texas Tech University

People often rely on external aids to support cognitive performance, a process known as cognitive offloading. Most research on offloading examines memory-based tasks in self-paced environments; less is known about offloading in rapid, attention-demanding tasks. We conducted an orientation and contrast discrimination task (n =24) where change magnitudes between serially presented gratings varied across four levels of difficulty. The experimental group could offload by requesting an optional cue before each trial, which indicated the relevant dimension. There was no additional cost associated with selecting to offload. The control group always received the dimensional cue without request. As expected, response accuracy significantly decreased across levels of difficulty for both groups. Furthermore, participants took longer to respond on the more difficult blocks, and this effect was largely driven by the experimental group. Within this group, participants tended to electively offload more frequently on the most difficult trials (50%) compared to the easiest trials (34%). A binomial generalized linear model revealed that task difficulty was a significant predictor of the decision to offload on a trial-by-trial basis; while the decision to offload also predicted accuracy at the trial level. These results indicate that participants can make rapid decisions to offload attentional demands, and that they may consider task difficulty to do so. However, the rates of offloading suggest that on the easiest trials, participants may have found attending to both dimensions less effortful than deciding to offload.

43.447 DO I KNOW THAT MY ATTENTION NEVER STAYS STILL? METACOGNITIVE MONITORING OF SUBSECOND ATTENTIONAL DYNAMICS

Cheongil Kim¹ (kimcheongil@gmail.com), Sang Chul Chong¹; ¹Yonsei University

Despite the general impression of a stable visual world, our visual performance fluctuates continuously—even within a second—due to dynamic shifts of attention. Specifically, attention varies transiently following exogenous cues (transient attention) and rhythmically in synchrony with intrinsic neural oscillations (rhythmic attention). Here, we asked whether people can monitor these subsecond attentional dynamics and what mechanisms support such metacognitive monitoring. Using an exogenous cueing paradigm combined with confidence judgments, we examined whether confidence tracks fluctuations in orientation discrimination (Experiment 1) and target detection performance (Experiment 2) across cue–target intervals. To this end, we decomposed fluctuations in performance and confidence into subcomponents reflecting transient and rhythmic attention and examined performance–confidence correlations within each component. In Experiment 1, we observed positive correlations between orientation discrimination performance and confidence for both transient and rhythmic attention, indicating accurate metacognitive monitoring. In Experiment 2, we aimed to further clarify the mechanisms underlying metacognitive monitoring using a target detection task (Kim & Chong, 2025), focusing on whether observers monitor perceptual evidence (e.g., the quality of target perception) or visual system states (e.g., attentional states). Confidence in detection tasks can be informed by perceptual evidence when judging target presence (e.g., high confidence when the target is clearly perceived), but not when judging target absence, where little perceptual evidence is available. Therefore, if confidence is based solely on perceptual evidence, it should correlate with performance fluctuations for target-present responses but not for target-absent responses. Consistent with this prediction, we found positive correlations between detection performance and confidence for target-present responses but little correlation for target-absent responses. These findings suggest that metacognitive monitoring of attentional dynamics relies primarily on perceptual evidence rather than direct readouts of internal states.

This work was supported by the Yonsei University Research Fund (Post Doc. Researcher Supporting Program) of 2025 (project no.: 2025-12-0208) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2022-NR070542).

43.448 CONFIDENCE IN PERCEPTUAL ESTIMATES REVEALS THAT SENSORY NOISE IS DOUBLY STOCHASTIC

Avinash Ranjan¹ (akr2979@eid.utexas.edu), Jongmin Moon¹, Robbe L. T. Goris¹; ¹The University of Texas at Austin

Perceptual interpretations of the state of the environment are subject to noise and therefore inherently uncertain. Humans know this and typically report more confidence in certain than in uncertain interpretations. The computational processes that underlie this form of metacognition have mostly been studied using binary decision-making tasks in which subjects assign sensory stimuli to one of two categories. Yet many real-world tasks involve continuous estimation rather than binary decision-making. We wondered whether observers possess

similar introspective insight in the reliability of continuous perceptual estimates. To investigate this, we conducted a behavioural experiment in which subjects were presented with a visual stimulus and then provided both a continuous estimate of stimulus orientation and their confidence in this estimate (using a binary response scale). We manipulated task difficulty by varying stimulus contrast, dispersion, and duration. As expected, mean estimation error depended on each of these factors. Interestingly, ‘confident’ estimates were systematically associated with lower estimation error than ‘not confident’ estimates. This was true across and within experimental conditions, revealing a robust link between perceptual confidence and estimation quality. We then asked whether standard models of decision confidence can capture these results. Surprisingly, the answer is no. When sensory noise is modelled as arising from a single stochastic process, the downstream confidence computation cannot reveal variations in estimation quality within the same experimental conditions. To capture such introspective insight, it is necessary to assume that (1) sensory measurements are corrupted by a doubly stochastic noise process, and (2) subjects have some knowledge of the second layer of variability. Our findings thus both enrich our understanding of the noise that limits the quality of perception, and of the metacognitive mechanisms that enable us to accomplish goals in spite of the resulting uncertainty.

UO1NS136338

43.449 AN ADAPTIVE PARADIGM COMBINING CONTINUOUS ORIENTATION REPORTS AND CONFIDENCE RATINGS FOR ASSESSING VISUAL FUNCTION AND METACOGNITION

Jan Skerswetat¹ (jskerswe@hs.uci.edu), Rushil Kumar¹, Andrew Browne¹, Cherlyn J Ng¹; ¹University of California, Irvine

Visual metacognitive confidence, the ability to evaluate one's own performance, is underutilized in clinical vision assessments. To address this, the Angular Indication Measurement Metacognition (AIM-MetaCognition) paradigm was developed to jointly assess visual functions and metacognitive confidence. As proof-of-concept, AIM-MetaCognition assessed visual acuity (VA) in twenty healthy adults. Participants reported gap orientations of Landolt-Cs (4×4 grid, randomized) and rated confidence per cell. Stimulus sizes were log-scaled and adaptively adjusted (threshold $\pm 2\sigma$). Hence, VA and confidence were estimated on continuous scales. Reported gap-orientation errors and confidence ratings were fitted separately to psychometric functions to estimate VA derived from confidence and performance. Metacognitive calibration was assessed using linear regression. A Confidence-Signal-Detection (CSD) model estimated the confidence scaling factor (k), which quantifies how sensory evidence translated to confidence. AIM-MetaCognition was validated in several experiments assessing its repeatability, and effects of binocular versus monocular viewing, simulated vision loss via Gaussian blur, task instructions, predictive versus retrospective confidence judgements. Additional experiments adaptively estimated

VA on confidence alone and confidence-weighted performance. Cross-visual function generalizability was also demonstrated with AIM-Metacognition. Bland-Altman analyses revealed no significant test-retest bias for psychometric or metacognitive parameters. Confidence ratings were bimodally distributed with comparable proportions of low- and high- judgements. RM-ANOVA showed significant main effects of viewing condition and measurement type on VA($p < .001$): binocular VA outperformed monocular VA, and Confidence-VA was worse than Performance-VA($p < .01$). Binocular viewing produced greater CSD-k miscalibration than monocular ($p = .03$). Neither task instruction(Confidence versus Certainty) nor the method of determining VA (based on performance, confidence, or both) affected VA($p > .05$). AIM-Metacognition offers a reliable, self-administered, and adaptive method for assessing visual functions and confidence. By capturing metacognition, it provides new insights for evaluating vision for clinical and basic vision science.

J.S. co-inventor of AIM (Angular Indication Measurement) method, owned by Northeastern University, Boston. J.S. is a shareholder and co-founder of PerZeption Inc., which has an exclusive licensing agreement for AIM with Northeastern University

43.450 HIERARCHICAL PROCESSING OF CHOICE AND CONFIDENCE IN HUMAN PREFRONTAL CORTEX

Yiran Ge¹, Siyuan Cheng¹, Yixue Zhao², Nihong Chen²; ¹Tsinghua University, ²South China Normal University

Decision confidence, a subjective estimate of the probability of being correct, enables individuals to evaluate their decisions and guide flexible behavior. Despite its pivotal role in metacognition, its neural basis remains unclear. To investigate the cortical hierarchy underlying confidence formation, we conducted an fMRI study using a perceptual decision task, in which subjects judged the motion direction and then provided confidence ratings. We first searched for brain regions that tracked subjects' confidence. Whole-brain analyses revealed confidence-related signals in dorsal anterior cingulate cortex (dACC) and rostral prefrontal cortex (rPFC), both showing greater activation under low-confidence condition. Importantly, we uncovered distinct roles of dACC and rPFC in choice and confidence generation. dACC exhibited an interaction between task difficulty and response correctness, indicating its role of error monitoring during decision making. In contrast, confidence-related activity in rPFC predicted individual differences in behavioral metacognitive sensitivity, suggesting a role in monitoring the quality of subjective confidence. Finally, connectivity analyses demonstrate that dACC, parietal cortex and occipital visual areas, transferred decision-related information to rPFC to support confidence computation. In sum, our findings highlight a role for rPFC in integrating decisional signals from sensory and decision-processing stages to construct subjective confidence judgments.

This work was supported by STI2030-Major Projects (2021ZD0203600), National Natural Science Foundation of China

(32571227), and the Research Center for Brain Cognition and Human Development, Guangdong, China (2024B0303390003).

43.451 CONFIDENCE UPDATING OF VISUAL INFORMATION GATHERING IN DIFFICULT AND COSTLY SITUATIONS

Laura De Laere^{1,2}, Ruth J. van Holst², Pascal Mamassian¹; ¹École Normale Supérieure Paris - PSL, ²Amsterdam UMC - University of Amsterdam

Metacognition, or the ability to monitor and regulate one's cognitive processes, is often expressed through decision confidence. In perceptual decision-making, confidence reflects subjective certainty and can guide behavior by shaping decision boundaries. However, it is unclear whether confidence is monitored continuously during information gathering, and how sampling costs influence this process. We hypothesized that confidence would flexibly track task demands, decreasing under greater difficulty or cost, and that confidence would be more sensitive to new evidence when sampling was costly. To test these hypotheses, we adapted a visual information-gathering task in which participants ($n=25$) revealed as many hidden cards of a deck as they wished before reporting the dominant color. Using a 2x2 design, we manipulated decision difficulty by varying color proportions (60% or 75% dominant color) and sampling cost (1 or 2 points per card revealed). After each revealed card, participants had a 30% chance of having to provide a temporary decision and confidence rating, allowing us to track dynamic updates in confidence. Participants earned rewards for correct decisions, and a greater reward if fewer cards were revealed. As expected, participants showed lower accuracy in the difficult condition and sampled more when trials were difficult or sampling cost was low. Confidence decreased under high cost and for difficult trials. Accuracy had the largest impact on confidence, with incorrect decisions decreasing confidence substantially. A linear mixed-effects model predicting confidence based on evidence strength, cost, difficulty, and accuracy revealed that confidence increased with stronger evidence. Importantly, participants' confidence ratings were more strongly influenced by new evidence when further sampling was discouraged by higher cost. Together, these results support our hypotheses and highlight that participants strategically calibrate their confidence. They reduced it when the task is harder or costlier, and weighted evidence more heavily when further sampling is discouraged.

MSCA Doctoral Network "CODE" (grant number: 101119647)

43.452 PERCEPTUAL CONFIDENCE SEEMS BLIND TO PRIOR INFORMATION

Akash Raj Gunniya Prakash¹ (akashrajgp@utexas.edu), Jongmin Moon¹, Robbe L. T. Goris¹; ¹The University of Texas at Austin

Perceptual interpretations of the state of the environment are not flawless. Humans know this and typically indicate to be more confident in accurate than in inaccurate interpretations. This introspective

knowledge resembles metacognition of theoretically ideal decision-makers who leverage the normative framework of Bayesian inference to compute the probability that an uncertain inference is correct. Key to this strategy is that two distinct types of knowledge are used: one pertains to the noisy nature of sensory measurements, the other to structural regularities in the environment (i.e., prior knowledge). Whether human decision confidence is also based on both types of knowledge is a topic of debate. To shed light on this question, we conducted a perceptual confidence experiment using a task designed to distinguish between different confidence computation strategies. Each trial, subjects judged whether a noisy visual stimulus was rotated clockwise or counterclockwise relative to 45 degrees and reported their confidence in this decision (high vs low). We manipulated both sensory uncertainty (3 levels of orientation dispersion) and stimulus strength (10 different orientations). Critically, we ran the experiments in blocks of 120 trials and used this temporal structure to manipulate environmental regularities. Within a block, stimulus orientation was either sampled from a uniform distribution or from a bi-delta distribution. Subjects were cued about the current stimulus distribution and knew the task statistics well. Under these circumstances, prior knowledge has no influence on the perceptual decisions of a Bayesian decision-maker but drastically impacts the pattern of their confidence reports. Human perceptual decisions closely resembled the Bayesian prediction, but this was not true of their confidence reports. Specifically, perceptual confidence seemed blind to prior information. Together, these results suggest that perceptual confidence results from a computation that leverages knowledge about the noisy nature of perception, but ignores informative regularities in the environment.

NIH R01 EY032999

43.453 PROBABILITY VERSUS EVIDENCE: COMPARING CONFIDENCE MODELS IN MULTI-ALTERNATIVE PERCEPTUAL DECISION MAKING

Kai Xue¹, Medha Shekhar, Dobromir Rahnev; ¹Georgia Institute of Technology, ²Université Libre de Bruxelles, ³Georgia Institute of Technology

Confidence models in perceptual decision making fall into two main categories: probability- vs. evidence-based models. Probability-based models derive confidence by computing posterior probabilities via Bayes' rule, whereas evidence-based models derive confidence directly from raw internal evidence without applying Bayes' rule. Despite their fundamental difference, clearly distinguishing between these two categories has remained challenging. A key difficulty is that most prior research has relied on simple two-choice tasks, which often provide weak constraints and therefore cannot clearly dissociate the different model classes. Here we show how confidence models can be distinguished using rich multi-alternative perceptual decision-making tasks. We conducted three experiments where subjects indicated which of 3 to 5 colored dot clouds was the most numerous. Critically, we included up to 12 different numerosity combinations, creating a complex dataset that can more efficiently adjudicate between

confidence models. The choice data was very well-explained by a parsimonious 2-parameter model of the internal evidence distributions. Building on this model of the internal evidence, we compared three prominent evidence-based models (Positive Evidence, Top-2 Difference, and SoftMax) and three prominent probability-based models (Bayesian Confidence Hypothesis, Probability-Based Top-2 Difference, and Entropy). We found that the evidence-based Top-2 Difference model captured all qualitative features of the data well and performed better than all other models across all three experiments (smallest summed AIC differences = 61, 303, 2110 for each of experiments). In addition, all probability-based models failed to capture a critical pattern replicated across all three experiments: showing larger numbers of dots while preserving the ratios across different alternatives produce higher confidence despite matched accuracy. Instead, all probability-based models predicted that both accuracy and confidence would remain unchanged when increasing dot numerosity. Our results challenge probability-based confidence computation framework in perceptual metacognition, providing evidence that confidence may operate in the evidence space rather than the probability space.

This work was supported by the National Institute of Health (award: R01MH119189) and the Office of Naval Research (award: N00014-20-1-2622).

Motion: Optic flow, in-depth, biological, higher-order

**MONDAY, MAY 18, 8:30 AM – 12:30 PM,
PAVILION**

43.454 INFANTS' SENSITIVITY TO SIZE AND SPATIOTEMPORAL CONTENT OF A DYNAMIC 1/F TEXTURE ACROSS THEIR PERIPHERAL VISUAL FIELD

Aishwarya Ravi¹, Stephanie A Biehn¹, Fuseina Safianu¹, T Rowan Candy¹; ¹Indiana University

Purpose Infants experience motion in retinal images resulting from self-motion, eye movements, and motion of objects in their visual periphery. Adult studies have shown scaling of sensitivity to motion from the fovea to the periphery. Our goal was to determine this scaling in infants for dynamic 1/f textures. Methods A spatially 1/f noise texture was presented on a rear-projection screen (PROPixx projector, VPixx) at a series of diameters (3, 5, 8, 12 deg) and speeds (2, 8, 16 deg/s) on a mean luminance background. Eye movements were recorded using an Eyelink 1000 Plus (SR Research, 1000 Hz). A 9-point calibration routine was performed before presenting the stimulus at the center of the screen. It moved in a series of six interleaved linear smooth ramps and steps to two of three eccentricities (5, 12, 20 deg) within a single run. A total of 6 steps (2 diameters x 3 speeds) each to 5 and 12 deg eccentricity, and 12 steps (4 diameters x 3 speeds) to 20

deg eccentricity were presented across four runs (24 steps). The experimenter controlled the stimulus presentation. Results Adults and infants (12-20 weeks) provided stimulus-driven visual tracking to linear ramps and saccadic responses to steps at all three eccentricities (5, 12, 20 deg). After applying threshold criteria to define a response to a peripheral target, proportions were highest at 3 deg diameter & 8 deg/s speed at 5 deg eccentricity, 5 deg & 32 deg/s at 12 deg eccentricity, and 12 deg & 32 deg/s at 20 deg eccentricity for infants. In contrast, adult proportions were >80% across all stimulus combinations. Conclusion This preliminary study describes the scaling in the developing visual system's sensitivity to motion, and the requirements for an equally sensitive clinical test with eccentricity for naturalistic spatial image statistics.

NEI R01 EY014460

43.455 OPTIC FLOW SPEED DISCRIMINATION IS ENHANCED DURING WALKING

Cassidy Brown¹, Star Niu¹, Frank Durgin¹; ¹Swarthmore College

Durgin and Gigone (2007), employed a head-mounted-display (HMD) on mobile observers to show that the perceived speed of optic flow was reduced during walking (relative to the same optic flow presented while standing). Moreover, perceptual discrimination of optic flow speeds near (and above) walking speed was enhanced while walking. These findings are consistent with sensory-prediction models of efficient coding, but were based on small numbers of participants (often 10 per condition). The present study sought to replicate some of these prior observations using more rigorous methods: a larger set of participants, more trials, and a wider field-of-view HMD. In an initial set of 22 participants, we observed strong mean reductions in perceived visual speed while walking such that 2.0 m/s during walking was matched to 1.55 m/s [95% CI: 1.45, 1.65], while standing. Although there was fairly high variance in this value, the effect size was large ($D = 1.96$). The mean difference in Weber fractions between discriminations made while standing ($W = 0.073$) and those made while walking ($W = 0.055$) at the mean walking speed (1.39 m/s) were in the predicted direction ($p < .05$), but had a smaller effect size ($D = 0.40$) in this fairly small sample. However, an unexpected effect of testing order suggested that the difference only emerges in the second block tested (presumably because of familiarity with the task), where Weber fractions for the Walking condition ($W = 0.043$) were significantly less than those in the second-block Standing condition ($W = 0.074$), $t(20) = 2.87$, $p = .009$ ($D = 1.04$). This replicates the prior observation that perceptual precision is enhanced. This occurs in naturalistic conditions where sensory information concerning expected flow can be derived from motor and vestibular signals.

Swarthmore College

43.456 PREDICTING GRAVITATIONAL SELF-MOTION: FALLING DOWN VERSUS FALLING UP

Bjoern Joerges¹, Marlene Wessels², Heiko Hecht², Laurence R. Harris¹; ¹Center for Vision Research, York University, ²Johannes Gutenberg-Universität Mainz

It is well known that humans do not fully use the acceleration of a target to estimate the time-to-contact of a disappearing object. Rather, they rely on the object's speed at the point of disappearance to extrapolate its motion, which leads to an overestimation of time-to-contact when the target is under the influence of a positive acceleration. It has further been observed that this overestimation is attenuated or even fully absent when the object moves under the influence of Earth gravity, indicating that humans use a strong Earth gravity prior to make such judgements. In this project, we investigated whether this is also the case for self-motion. We immersed 8 participants in a 3D virtual environment consisting of tall buildings and a road and then visually simulated them as either falling down (towards the ground) or "falling up" (towards the simulated ceiling). To avoid confounds, an identical stretch of road was used both as the ceiling and the ground. Participants started out from rest and accelerated at +9.8 m/s² either upwards or downwards in randomly interleaved trials. After experiencing such optic flow for 0.8 to 1.4s, the display blacked out and participants were asked to indicate by button press when they either hit the ceiling with their head (-1g condition) or the ground with their feet (1g condition). Participants pressed the button significantly earlier when falling down than when falling up, indicating that they relied more on acceleration-based cues when predicting their own downwards motion than for upwards motion. This shows that – just like when judging the motion of objects – humans rely on a strong Earth gravity prior when assessing their self-motion.

43.457 BEYOND EXPANSION AND SPIRALS: COMMON MOTIFS OF OPTIC FLOW DERIVED FROM NATURAL VISUAL EXPERIENCE

Giselle R. Urquijo¹, Mark D. Lescroart¹; ¹University of Nevada, Reno

As we move through our environment, visual patterns of motion constantly project onto our retinas. Recently, studies investigating patterns of retinal flow during natural locomotion revealed that, as we walk, we experience patterns of spiral expansion. However, they have dominantly focused on structured walking tasks in one setting. In order to characterize the diversity of patterns of motion that occur in day-to-day life, we analyzed the Visual Experience Dataset which provides hundreds of hours of egocentric video with gaze tracking as humans perform naturalistic behavior. To subdivide the data into tasks, we binned the dataset's labels into six broad categories of activities: locomotion, manual tasks, passive, screens, social, and sports. We estimated gaze-centered optic flow during fixation epochs using Recurrent All-Pairs Field Transforms (RAFT). From these samples of optic flow we removed full-field slip by factoring out horizontal and vertical vectors. We then performed principal component analysis (PCA) on them. Consistent with past work, we find that the first two principal components (PCs) across all tasks are patterns of expansion and spiral motion. We then computed PCA on the optic flow samples

from each task separately. We found that locomotion and sports, which are dominated by self-motion, have lower dimensional patterns of optic flow compared to tasks where there is little to no self-motion i.e. screens, social, passive, and manual tasks. Within locomotion, we find that a pattern of vertical divergence, i.e. leftward and rightward motion away from the vertical meridian, appears within the first three PCs. This suggests that people tend to fixate on vertical boundaries in an environment as they approach them. This is a previously unreported motif in visual motion which may be manifest in tuning of neurons and/or in the organization of the brain.

Supported by NSF CAREER #2340895 to M.D.L.

43.458 THE ACCURACY OF 3D HEADING PERCEPTION

Oliver Layton¹ (oliver.layton@colby.edu), Daniel Yu¹, Alexander Lyon¹, Maddie Puzon¹, Sam Cohen¹; ¹Colby College

Moving observers experience structured patterns of retinal motion known as optic flow. Humans can accurately judge their linear direction of self-motion (heading) from optic flow, often within 1° for nearly straight-forward movement (Warren et al., 1988). However, error increases for more eccentric headings, reaching 10° or more (Sun et al., 2020). Studies that characterize the accuracy of heading perception over the widest range have focused on the horizontal plane and report systematic biases toward peripheral (left/right) headings (Crane, 2012; Cuturi & MacNeilage, 2013). In the present study, we quantified 3D heading perception across the full range of azimuths and elevations. On each trial, observers viewed optic flow generated by simulated linear self-motion through a 3D environment in a VR headset. After the motion ceased, observers indicated their perceived heading by physically pointing to align a virtual laser, yoked to a handheld motion-tracked controller, with their perceived direction of movement. Consistent with existing work, forward headings were judged more accurately than backward headings. Contrary to horizontal-plane studies, however, we found central rather than peripheral biases: responses were shifted toward the fore-aft axis (0°/180°) in azimuth and toward the equatorial plane (0°) in elevation. To reconcile these divergent results, we conducted a control experiment in which observers judged horizontal headings either by pointing or by adjusting a dial, the method used by Crane (2012) and Cuturi & MacNeilage (2013). Strikingly, we replicated both effects: pointing judgments yielded center bias, whereas dial adjustments yielded peripheral bias. These findings provide a more comprehensive characterization of 3D heading perception and suggest that previously reported directional biases may reflect not only perceptual mechanisms, but also task-dependent response constraints introduced by different psychophysical paradigms.

The Haynesville Project

43.459 A FUNCTIONAL GRADIENT OF OBJECT MOTION TO SELF MOTION BASED SELECTIVITY ACROSS THE PROPOSED HUMAN HOMOLOG OF FST

Kaylie J. Capurro¹, Arnab Biswas¹, Giselle R. Urquijo¹, Yu Zhao¹, Mark D. Lescroart¹; ¹University of Nevada, Reno

The fundus of the superior temporal sulcus (FST) has been established in macaques to be selective for 3D motion. Here, we investigate the region in and around the recently-proposed human homolog of FST for selectivity to different kinds of 3D motion: self motion and object motion. We localize hMT and FST with established functional contrasts. To investigate tuning for object and self motion, we created rendered animations of geometric objects tumbling in space and animations suggestive of self motion down a hallway. We used the optic flow of these renders to define random-dot kinematogram (RDK) versions of the animations. Subjects viewed these RDKs in 3D anaglyph form while in an MRI scanner. To localize areas selective for object motion, we contrasted the object RDKs with an expanding dot condition and a linear full-field dot motion condition. To localize areas selective for self motion, we contrasted the hallway RDKs with the linear full-field dot motion condition. Contrasts were projected onto flatmaps and ROIs for hMT and FST were delineated for each participant. We find that the object motion contrast activates regions ventral to and overlapping with the posterior portion of FST. We find that the self motion contrast activates regions anterior to and partially overlapping with FST. Our findings suggest that there is a gradual shift in tuning from object motion to scene or self motion, going from posterior to anterior across the proposed location of human FST. Since we typically interact with objects at a closer distance than we do scenes, we propose that the representation of 3D motion in this region may be organized according to the distances at which we encounter different types of motion.

Supported by NSF CAREER #2340895 to M.D.L.

43.460 VARIATION IN MOTION-IN-DEPTH DIRECTION DISCRIMINATION ACROSS THE VISUAL FIELD WITH ISOLATED BINOCULAR CUES

Jovan Kemp^{1,2} (jovan_kemp@brown.edu), Bas Rokers^{1,2}; ¹Psychology, New York University Abu Dhabi, ²Center for Brain and Health, New York University Abu Dhabi

Accurate motion in depth perception is critical for interaction with the physical world. However, ~50% of individuals are unable to exploit binocular motion cues in part of their visual field - a type of apperceptive agnosia (Barendregt et al., 2014, Barendregt et al., 2016). Motion in depth stimuli contain two binocular cues. An interocular velocity difference (IOVD) cue, defined by differences in velocity of the two retinal projections, and a changing disparity (CD) cue based on differences in binocular disparity over time. Prior work has studied the contribution of these cues, but not under complete cue isolation conditions. Here, we leveraged a recently developed isolation technique, which exploits the periodic nature of gratings (Sheliga et al., 2016; Whitner et al., 2021). Participants judged the direction of gratings moving in depth (toward/away) as defined by either the full cue, or the IOVD/CD cues in isolation. Gratings were presented in the

central 7.5 deg of the visual field at each of 32 locations (8 polar angles x 4 eccentricities). Sensitivity varied significantly both across and within observers. For all observers performance dropped with eccentricity across conditions ($p = 0.002$), as expected based on a decrease in retinal resolution. In addition, performance varied idiosyncratically across eccentricity-matched polar angle locations not easily captured by group level statistics. Finally, sensitivity in the full cue condition was best predicted by performance in the IOVD cue ($p = 0.022$), but not the CD cue ($p = 0.1$) condition, consistent with previous findings using cue-reduced stimuli. In all, isolating contributions of the two binocular cues to motion-in-depth are consistent with observed between- and within-participant variation in prior work. Our design provides an improved framework for identifying motion processing deficiencies in the human visual system, and may further our understanding of the neural basis of visual agnosias.

NYUAD Center for Brain and Health, funded by Tamkeen under NYU Abu Dhabi Research Institute grant CG012

43.461 INTERPRETING SOCIAL ACTIONS: THE ROLE OF JOINT ACTION IN BIOLOGICAL MOTION PERCEPTION

Uriel Gonzalez¹ (uriel.gonzalezbr@rutgers.edu), Jacob Feldman¹; ¹Rutgers, The State University of New Jersey

Understanding others' actions and intentions is central to human perception, cognition, and social interaction. Classic research on biological motion shows that people can infer complex actions from reduced visual information (Johansson, 1973). Most studies have examined single agents; however, many social actions gain meaning only through interactions between two people (e.g., Ikeda, Destler, & Feldman, 2025). This study examines how people interpret biological-motion displays comprising two individuals, focusing on whether, and to what degree, such interpretations depend on their joint action. Observers viewed either both agents together or each one in isolation. Each subject ($N = 145$) viewed 24 short videos depicting social actions, presented as point-light displays (PLDs) or stick figures. The PLD and stick-figure stimuli were generated from naturalistic videos using OpenPose to extract joint positions. Each action appeared in one of three versions—full interaction, only agent A, or only agent B—counterbalanced so that no participant viewed the same action twice. After each video, participants provided free-text descriptions of the perceived action. A separate participant group viewed the full interactions, and their responses were used to derive reference labels. We used an AI-based text analysis procedure to compute semantic similarity among participants' responses, quantifying (a) agreement among observers and (b) similarity between their descriptions and the reference labels. A hierarchical Bayesian model estimated posterior distributions across representation type, interaction condition, action function, and agent role. Observers showed higher agreement when viewing both agents interacting, demonstrating that social context enhances action understanding. Agreement among participants was more variable than similarity to the reference labels, which remained stable across conditions. When viewing only one agent, observers

showed lower agreement and sometimes converged on a different action, indicating that the absence of a partner changes perceived meaning. Together, these findings shed light on how humans interpret moving and interacting human forms.

43.462 CAN A HOLE HAVE A GOAL? EFFECTS OF FIGURE/GROUND CUES ON THE DETECTION AND EVASION OF CHASING IN A 'KEEPAWAY' TASK

Sarah Lassise¹, Marina Pace², Benjamin van Buren³; ¹The New School, ²The New School, ³The New School

Imagine you are in a crowd trying to detect the movements of an assassin. Which entities would your visual system most readily interpret as chasing you? Intuitively, you might be biased to see chasing among figural regions—objects—rather than ground regions (spaces between and within objects and surfaces). However an alternative possibility is that you readily see chasing in any bounded visual region that moves appropriately, independently of whether you perceive it as figure or ground. Subjects wore 3D anaglyph glasses and played a 'keepaway' game, using the mouse to move a small disc around a rectangle presented in front of a textured background. The rectangle was populated by moving Objects (circular cutouts of background texture presented in front of the rectangle) and Holes (circular apertures which revealed parts of the background as they moved). On each trial, the subject tried to detect and evade one of the shapes, which chased their cursor with one of four degrees of chasing subtlety—with heading updates 0° toward their cursor, or drawn from a $\pm 60/120/180^\circ$ range centered on their cursor. Chasing detection and evasion were greater for Objects than Holes across subtlety levels, but for both types of entities, the rate of detecting and evading the shape as a function of chasing subtlety followed a strikingly-similar U-shaped curve: Subjects easily detected and evaded obvious 0° chasing shapes, fell prey to $60/120^\circ$ shapes approaching obliquely like assassins, and easily evaded 'incompetent' 180° shapes chasing too indirectly to approach. These results show that chasing detection favors figural objects, but that holes can still be perceived as chasing to some extent. The similar effects of chasing subtlety for Objects and Holes are consistent with a chasing detector that operates over the movements of bounded regions, which is only partially sensitive to figure-ground distinctions.

43.463 WHAT LOOKS LIKE HORIZONTAL MOTION TODAY MAY LOOK LIKE VERTICAL MOTION TOMORROW: WITHIN-OBSERVER DIFFERENCES IN VISUAL APPARENT MOTION.

Nicolaas Prins¹ (prins.nicolaas@gmail.com); ¹University of Mississippi

Apparent motion refers to the phenomenon in which motion is perceived between two or more stationary visual tokens that are presented in successive frames. When frames contain multiple tokens, the visual system must solve the motion correspondence problem: Which token in one frame corresponds to which token in the alternate frame? Generally, the visual system prefers to assign correspondence

between tokens that are separated by short distances (the nearest neighbor principle). However, when motion tokens positioned on diagonally opposed corners of a rectangle are alternated with tokens on the other two corners, horizontal and vertical motion are equally likely to be perceived when the vertical sides of the rectangle are longer than the horizontal sides. Substantial individual differences exist in the magnitude of this bias. Here I investigate whether the magnitude of this bias also varies within observers across different testing occasions. Each of eight observers tested in five identical sessions that were conducted on separate days. In each session, six independent adaptive trial sequences were conducted to relate the probability of perceiving horizontal motion to the ratio of vertical to horizontal distances between potential motion matches and to determine the equivalence ratio: The aspect ratio at which horizontal and vertical motion are equally likely to be perceived. While equivalence ratios were highly consistent within individual sessions, they varied substantially not only between observers but also between the sessions of individual observers. This within-observer, between-session variance was comparable in magnitude to the between-observer variance, exceeding the within-session variance of equivalence ratios by approximately half an order of magnitude. The mechanism behind the observed within-observer variance across sessions is speculative at present, but its existence carries important implications for the design, analysis, and interpretation of research into visual apparent motion.

43.464 MULTIPLE OBJECT TRACKING UNDER MESOPIC ILLUMINATION: THE ROLE OF OBJECT CONTRAST AND SPACING

Quan Lei¹, Taylor Shupsky¹; ¹Wichita State University

Humans can simultaneously track a limited number of moving objects, a capability known as multiple object tracking (MOT). Whereas many visual functions are compromised under low lighting conditions, it is yet unknown how MOT performance is affected by reduced illumination. This study was aimed to measure MOT performance under mesopic illumination and examine the role of object contrast and spacing. In Experiment 1, participants tracked 1-4 targets out of 10 objects (dots of 1° diameter with a Weber contrast of 0.1) moving along random trajectories (speed: 6°/s) within a 30°×30° field for 6s, after which they were probed on one of the objects (target or distractor). They performed the task under either photopic (~100 cd/m²) or mesopic illumination (~1 cd/m²) implemented by wearing goggles fitted with neutral density filters. Both target number and illumination level had an effect on tracking performance (percent correct of target identification). There was also a significant interaction between target number and illumination level, such that tracking performance was similar between photopic and mesopic illuminations when tracking only one target but performance was worse under mesopic illumination than photopic illumination when tracking more than one target. In Experiment 2, participants performed the same MOT task but with increased contrast of the objects (Weber contrast of 1); in Experiment

3, object contrast being low at 0.1, a minimal spacing between objects (which equals eccentricity*0.4 following Bouma's law) was dynamically maintained throughout the motion phase. In both experiments, there was a main effect of target number on tracking performance, but more importantly, there was neither an effect of illumination nor an interaction between target number and illumination. The results suggest that illumination level is a limiting factor in MOT and the performance decrement under mesopic illumination might reflect a data-limited process due to reduced contrast sensitivity and acuity.

43.465 MULTI-LAYERED HIERARCHICAL MOTION: ANCESTORS MATTER, DESCENDANTS DO NOT

Ziwei He¹, Yen-Ju Chen², Yung-Hao Yang¹, Shin'ya Nishida¹; ¹Graduate School of Informatics, Kyoto University, Kyoto, Japan, ²Department of Psychology, Soochow University, Taipei, Taiwan

Humans perceive motion in a highly flexible and hierarchical manner. For example, when observing a walking person, we simultaneously perceive the global translation of the body and the local oscillatory movements of the limbs. Classical work shows that the visual system can infer multiple latent motion components and their hierarchical organization. More recent work has shown that observers can dynamically assign motion to infer reference objects from dot motion. However, the perceptual limits of hierarchical motion decomposition remain unclear. Here, we investigated these limits using a vector matching paradigm. The stimulus consisted of four hierarchical levels: a static background, a large circle, a middle circle, and a small circle. Each layer was marked by a distinct color and composed of randomly distributed dots moving at a constant speed and direction. The motion of each deeper layer was defined as the vector sum of the motions of all preceding layers, thereby forming a nested structure. Before stimulus onset, participants were informed of the reference and target layers. The stimulus was presented for 1 s and repeated three times, after which participants reproduced the perceived relative motion of the target by adjusting the direction and speed of matching disks. Performance was evaluated by comparing the reproduced motion vectors with both the ground-truth absolute motion and the intended relative motion (target minus reference) across different hierarchical configurations. Accuracy was highest when the reference and target formed a direct hierarchical connection ({reference{target}}). Performance declined for both standard induced-motion configurations ({reference{irrelevant{target}}}) and double induced-motion configurations ({reference{irrelevant1{irrelevant2{target}}}}). In contrast, task-irrelevant child ({reference{target{irrelevant}}}) and grandchild ({reference{target{irrelevant1{irrelevant2}}}}) layers had minimal influence on performance. These findings demonstrate that, although humans can decompose hierarchical relative motion, their precision is systematically constrained by hierarchical distance, with dominant contributions from higher-level reference objects and little contamination from lower-level, task-irrelevant motion.

This study is supported by JST SPRING JPMJSP2110, Kakenhi JP24H00721

Perceptual Training, Learning and Plasticity: Statistical learning

**MONDAY, MAY 18, 8:30 AM – 12:30 PM,
PAVILION**

43.466 PATIENTS WITH EARLY VISUAL CORTEX DAMAGE EXHIBIT STATISTICAL LEARNING WITHIN THEIR BLIND FIELDS

Matt Cavanaugh¹, Berkeley Fahrenthold¹, Krystal Huxlin¹; ¹University of Rochester

A limitation of rehabilitation for stroke-induced cortical blindness (CB) is the burdensome nature of training interventions, which require hours of focused effort over weeks of daily training - a difficult task when adjusting to life post-stroke. Statistical learning, in which patterns of information are learned automatically, offers a potentially easier method of improving blind field performance without the concerted effort of traditional training. Here, nine CB participants and six visually-intact controls performed a single session of an 8AFC motion discrimination statistical learning task. In CB patients, motion discrimination was pre-trained to recovery at the blind field location targeted. In the statistical learning task, motion directions of a high-contrast stimulus were grouped into triplets, each repeated 21 times (high incidence) or 7 times (low incidence), with participants unaware of these groupings. CB participants performed above chance (8AFC chance=12.5% correct) on the statistical learning task, averaging 25.0+/-13.3% correct ($p=0.02$), while controls averaged 94.8+/-4.8%, with no difference seen for high- versus low-incidence trials in either cohort. To assess learning, triplets were presented in a two-interval forced-choice task, with participants indicating the more familiar triplet. Intervals contained high- and low-incidence triplets, and five "Null" triplets never shown during the learning phase. Two Null sets followed the design rules of regular triplets (Lawful), while three broke the rules (Unlawful). CBs averaged 48.4+/-5.7% correct on the familiarity task, while controls averaged 61.1+/-10.0% ($p=0.021$). However, significant differences emerged between Null cases, with controls ($p=0.006$) and CBs ($p=0.02$) performing better when distinguishing presented triplets from Unlawful Nulls versus Lawful Nulls. Thus, in a single session, CBs learned statistical rules within their blind field, despite poorer discrimination performance than controls. These results provide the first evidence that statistical learning is possible within CB fields, motivating rule-based learning as a tool to accelerate patient training, making rehabilitation less burdensome.

43.467 OBSESSIVE-COMPULSIVE TENDENCIES WEAKEN THE ANTAGONISTIC INTERPLAY BETWEEN STATISTICAL LEARNING AND COGNITIVE FLEXIBILITY

Bianka Brezóczi^{1,2,3}, Teodóra Vékony⁴, Bence Csaba Farkas^{5,6}, Flóra Hann^{1,2,3}, Dezső Németh^{4,7,8}; ¹Brain, Memory and Language Research Group, Institute of Cognitive Neuroscience and Psychology, HUN-

REN Research Centre for Natural Sciences, ²Doctoral School of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary, ³Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary, ⁴Gran Canaria Cognitive Research Center, Department of Education and Psychology, Atlántico Medio University, Las Palmas de Gran Canaria, Spain, ⁵UVSQ, Inserm, Centre de Recherche en Epidémiologie et Santé des Populations, Université Paris-Saclay, Versailles, France, ⁶LNC2, Département d'études Cognitives, École Normale Supérieure, INSERM, PSL Research University, Paris, France, ⁷BML-NAP Research Group, Institute of Psychology, Eötvös Loránd University and Institute of Cognitive Neuroscience and Psychology, HUN-REN Research Centre for Natural Sciences, Budapest, Hungary, ⁸Centre de Recherche en Neurosciences de Lyon CRNL U1028 UMR5292, INSERM, CRNS, Université Claude Bernard Lyon 1, 69500 Bron, France

Theoretical models of Obsessive-Compulsive Disorder (OCD) propose that its symptoms stem from an imbalance between habitual and goal-directed processes, marked by a shift toward habitual responding and reduced goal-directed control. Although prior research has largely focused on reward-based learning, the contribution of reward-independent learning mechanisms to this imbalance remains insufficiently understood. To address this gap, the present study investigated the antagonistic interplay between statistical learning (SL), a reward-free form of learning, and cognitive flexibility, a key dorsolateral prefrontal cortex function supporting goal-directed behavior, in a non-clinical population. Adopting a dimensional approach to obsessive-compulsive (OC) tendencies allowed us to move beyond categorical diagnoses and more precisely capture subtle, continuous variations in symptom expression, thereby clarifying how specific neurocognitive processes relate to OC tendencies. A total of 404 non-clinical participants completed the Alternating Serial Reaction Time (ASRT) task to assess SL and a Card Sorting Task to measure cognitive flexibility. Our findings revealed an antagonistic relationship between SL and cognitive flexibility. Crucially, this inverse association weakened as OC tendencies increased, suggesting that even at non-clinical levels, OC tendencies may attenuate the typical competitive balance between habitual and goal-directed processes. Importantly, SL performance itself remained intact, and the observed effects can not be attributed to reward sensitivity, as learning occurred without external reinforcement. Overall, these results underscore the value of studying OC tendencies dimensionally and highlight the importance of examining interactions among neurocognitive systems, rather than isolated functions, to advance our understanding of OCD.

43.468 EYE-MOVEMENTS REVEAL LEARNING INACCESSIBLE TO CONSCIOUS AWARENESS

Irina Lavrova¹, Miranda Scolar²; ¹Texas Tech University, ²Texas Tech University

Implicit learning refers to the abstraction of structural regularities without conscious awareness, which widely studied using artificial

grammar learning (AGL) paradigms. Eye movements can offer additional insights into the mechanisms of implicit learning, serving as involuntary indicators of acquired knowledge (Silva et al., 2017), and therefore revealing underlying learning processes. Across two studies, we tracked adults' eye movements while they completed AGL tasks to test whether gaze patterns would reveal sensitivity to grammatical violations beyond what standard grammaticality judgments capture. In Study 1, participants (N = 65) were exposed to words generated by grammars of varying complexity while making same/different judgments that did not require explicit memorization. In a subsequent surprise grammaticality judgment test, participants showed increased dwell time ($\chi^2 = 4.57$, $p = .032$) and more regressions ($\chi^2 = 24.89$, $p < .001$) to letters that violated the grammar. They also fixated longer on simpler first-order violations compared to more complex second-order violations. Interestingly, participants' accuracy judgements were not a significant predictor of eye movements. Study 2 tested whether this eye-movement signature would transfer to a new surface vocabulary. After exposure, participants (N = 44) judged strings following the same underlying grammar structure but composed of entirely novel letters. Participants again fixated longer on rule-violating elements ($\chi^2 = 143.48$, $p < .001$). Furthermore, dwell time did not differ for violations created by repeated versus unique characters, indicating that participants did not rely solely on repetition patterns when processing a new symbol set. Instead, the data suggest that they had abstracted the underlying grammatical structure in a way that generalized beyond the specific letters or surface features of the stimuli. Taken together, these findings demonstrate that eye movements serve as indicators of implicitly learned patterns, their complexity, and internal structure, revealing learning that remains largely inaccessible to conscious awareness.

43.469 COMPETING SYSTEMS ACROSS DEVELOPMENT: EXECUTIVE CONTROL VERSUS VISUAL STATISTICAL LEARNING

Dezso Nemeth¹; ¹INSERM

Human predictive processing and learning rely on multiple cognitive systems supported by partially distinct brain networks. These systems do not always cooperate; instead, they may compete as the brain seeks to optimize performance. Prior research shows that reducing the engagement of prefrontal, attention-driven control processes can enhance visual statistical learning and prediction. Competition theory provides a powerful framework for interpreting developmental changes in implicit statistical learning and the formation of predictive representations. In our studies, we examined the developmental trajectory of statistical learning using both cross-sectional and longitudinal designs. Across both approaches, we consistently observed an age-related decline in statistical learning performance. Crucially, executive functions can predict the developmental trajectory of this fundamental learning mechanism. Stronger executive functions—typically improving with age—were associated with the decreasing trajectory of statistical learning. Together, our

developmental and cognitive neuroscientific results reveal a systematic antagonistic relationship between prefrontal executive control and visual statistical learning. These findings support a developmental “less-is-more” account in statistical learning and predictive processing, emphasizing that changes in the balance between cognitive control and statistical learning shape how prediction emerges across childhood, adolescence, and adulthood.

43.470 DISSOCIATING PAIR- AND ITEM-LEVEL REGULARITIES: AN EEG FREQUENCY-TAGGING STUDY

Çiçek Güney¹ (cicek.guney1@sabanciuniv.edu), Yasemin Gunindi¹, İlayda Efsane Algin¹, Nihan Alp¹; ¹Sabanci University, Istanbul

Statistical learning enables the visual system to extract regularities from probabilistic structure in the visual environment. When an object repeatedly occurs in a specific place, the visual system can form object-place associations. Yet underlying neural correlates of such statistical regularity remain elusive. Here, we investigate this by separating item-level and pair-level regularities and examining how integration processes evolve across repetitions (1st-10th). To this end, we used frequency-tagging EEG paradigm (object and place tagged at fundamental frequencies: $f_1 = 7.5$, $f_2 = 6$ Hz) and measured Steady-State Visual Evoked Potentials (SSVEP). Participants (N = 20) viewed objects and places, either as repeating object-place pairs (pair-level regularity) or as independently repeating items in different pair combinations (item-level regularity), followed by memory tasks. We quantified Signal-to-Noise Ratios (SNRs) at fundamental frequencies and their harmonics as an index of item-level processing, and their intermodulation (IM: e.g., $nf_1 \pm mf_2$) frequencies as an index of pair-level integration processing. As expected, behaviorally, participants showed higher accuracy for pairs learned under pair-level regularity, indicating stronger associative learning. SSVEP results revealed that place-related (mf_2) SNRs decreased more strongly within individuals across repetitions in item-level than in pair-level condition, indicating greater item-specific adaptation. Similarly, IM SNRs decreased only in item-level regularity condition. Moreover, across participants, memory association was significantly correlated with IM SNRs only in the pair-level regularity condition, suggesting a link between behavioral performance and SSVEPs. Contrary to our expectations, this correlation was negative ($r = -0.53$, $p = .017$). It may suggest that better performers integrate information more efficiently and require less neural effort, resulting in lower overall IM SNRs while maintaining stable IM SNRs across repetitions. Overall, these findings suggest that repeated exposure to item-level regularity may induce neural adaptation that weakens object-place integration. Accordingly, pair-level regularities seem to buffer against the adaptation effects induced by repeated items.

N.A. and Ç.G. are supported by the Scientific and Technological Research Council of Turkey (TUBITAK 1001) under the Grant Number 122K922.

43.471 RETROSPECTIVE LINKING IN VISUAL STATISTICAL LEARNING

Sophie D. Allen¹ (sophie.allen@yale.edu), Nicholas B. Turk-Browne^{1,2}; ¹Yale University, ²Wu Tsai Institute

Combining visual information across experiences enables novel inferences that support flexible behavior. Much of our understanding of how information is combined across experiences comes from studies of integrative encoding, in which related information is retrieved during a new experience and linked to that experience in memory. However, this mechanism does not work for situations in which seemingly unrelated experiences get linked retrospectively in light of new information. Here, we tested whether visual statistical learning supports such retrospective linking to enable multi-step inferences. Participants viewed a continuous stream of objects containing embedded pairs while performing a size judgment cover task. Unbeknownst to them, the sequence was structured into tetrads presented across two blocks: initially unrelated pairs in Block 1 (AB, CD) were followed by a bridging pair in Block 2 (BC). This structure allowed us to test not only whether the directly experienced pairs (AB, CD, BC) were learned but also whether the bridging pair retrospectively linked the initially unrelated pairs to enable a two-step inference (AD). In a two-alternative force-choice familiarity test, participants showed high accuracy for the directly experienced pairs, indicating successful visual statistical learning. However, accuracy for the two-step inference was substantially lower and not significantly greater than chance. This result stands in striking contrast with a second experiment in which retrospective linking occurred robustly in an explicit associative inference task where participants were instructed to learn associations. Thus, when associations are learned more implicitly during visual statistical learning, overlapping experiences do not automatically integrate earlier object associations. This dissociation between explicit and implicit learning suggests that multiple mechanisms can support the combination of experiences over time.

Perceptual Training, Learning and Plasticity: Category learning

MONDAY, MAY 18, 8:30 AM – 12:30 PM,
PAVILION

43.472 GIST-FIRST LEARNING FACILITATES CATEGORY ABSTRACTION ACROSS GAN-GENERATED SCENE SPACE

Hala Rahman¹, Maya R. Newton¹, Jeffrey D. Wammes^{1,2}; ¹Queen's University, Psychology Department, ²Queen's University, Centre for Neuroscience Studies

To organize information in our world, we are constantly using shared features to separate individual exemplars into distinct categories. A

hallmark of human intelligence is later generalizing this category learning to novel, relevant contexts. While current methods effectively capture this learning using simple stimuli, less is known about how we manage to parse continuous stimuli that approach the complexity of real world learning. Here, we use learning conditions of varying composition to test whether learning the most difficult aspects of a category promotes improved generalization. Following prior work (Son et al., 2024), category learning stimuli were developed by sampling points along a plane in the latent space learned by a generative adversarial network (GAN), describing the continuously varying features of indoor apartment scenes (e.g. kitchens). Participants were tasked with categorizing a subset of scenes from the innermost region of the plane, with feedback. A “hard” group received disproportionately more exemplars close to the decision boundary than the “easy” group. At test, we assessed generalization in three different ways: interpolation: untrained scenes in the innermost region of the plane; extrapolation: scenes from outside the innermost region; and parallel-plane: scenes from a different area in the latent space. Unsurprisingly, the “easy” group performed better, but unexpectedly this was especially true for the hardest exemplars, where the “hard” group had triple the experience. Even more striking, this group generalized better than the “hard” group in all but interpolation. This implies that the “easy” group extending their enhanced knowledge of coarse category distinctions to guide their judgments about even the hardest cases. Our work here uses GAN-generated scenes to fill theoretical gaps in our understanding of visual category learning, revealing distinct patterns of learning and generalization depending on the learning regimen.

43.473 THE ROLE OF FREQUENCY IN DIFFERENT CONFLICT TYPES IN COGNITIVE CONTROL

Emma Wiedenmann¹, Myeong-Ho Sohn; ¹The George Washington University, ²The George Washington University

Stimulus frequency has repeatedly been shown to enhance performance across a range of cognitive tasks. However, the mechanisms through which frequency shapes cognitive control remain insufficiently understood. Prior research distinguishes two major learning types: response learning, which is habit-based and reflects direct associations between stimuli and responses, and categorization learning, which is rule-based and requires abstraction over stimulus features (Goodman, 2021). In parallel, the Dimensional Overlap model (Kornblum et al., 1990) identifies two primary forms of conflict: Stimulus–Stimulus conflict, characteristic of the Stroop task, and Stimulus–Response conflict, characteristic of the Simon task. Integrating these frameworks, the present study examines how stimulus frequency influences the engagement of cognitive control across distinct types of conflict and task demands. Across four experiments, participants completed one of the following tasks: (1) a response-learning Stroop task, (2) a response-learning Simon task, (3) a categorization-learning Stroop task, or (4) a categorization-learning Simon task. In each experiment, two colors served as high-frequency

stimuli, appearing three times more often than the low-frequency colors, while overall congruency proportions were equated across frequency conditions. This design ensured that any observed effects could be attributed to stimulus frequency rather than differences in conflict frequency. Results showed that under response learning, stimulus frequency modulated cognitive control only in the Simon task, which engages Stimulus–Response conflict, but not in the Stroop task. Under categorization learning, the reverse pattern emerged: frequency modulated performance only in the Stroop task, which engages Stimulus–Stimulus conflict, but not in the Simon task. Taken together, these findings indicate that stimulus frequency attenuates conflict specifically within the domain most relevant to the learning demands of the task.

43.474 COLOR CATEGORIES IN MACAQUE MONKEYS ACQUIRED THROUGH LONG-TERM ENGAGEMENT WITH A SET OF COLORED OBJECTS

Helen E. Feibes¹, Spencer R. Loggia^{1,2}, Isabela Téllez¹, Bevil R. Conway^{1,3,4}; ¹National Eye Institute, National Institutes of Health, ²Brown University, ³National Institute of Mental Health, National Institutes of Health, ⁴University of Maryland, College Park

Where do color categories come from? Prior work suggests that consensus color categories are not innate, which raises the hypothesis that they derive from colors of consensus behavioral relevance (Garside et al., PNAS 2025). To test the hypothesis, we trained two monkeys over nine years to learn the color-shape associations of twelve 2-D objects whose colors were six hues evenly sampling CIELUV, at two luminance levels. We then tested if the monkeys' experience with these highly relevant objects impacted color matching in a delayed-match-to-sample task (Bae et al., J Exp. Psych 2015), as predicted by the hypothesis. We used 83 equally spaced colors in CIELUV. In each trial, the animals were shown a Gaussian colored blob ("sample"); in the subsequent frame they were rewarded for touching the exact match out of four match options. We analyzed errors (75,355 trials in monkey 1, 55,104 trials in monkey 2; across 25 sessions). Mixture modeling showed evidence for 4-6 category biases in each monkey, with individual differences in their precise location in color space. Control animals showed 2 biases, attributable to underlying stimulus space nonuniformities (Garside et al., PNAS 2025). The results provide evidence in support of the hypothesis that long-term experience with a limited set of highly relevant colors contributes to the construction of cognitive color categories, which manifest in tasks unrelated to the behavioral relevance of the colors. We then used target confusability competition modeling (Schurgin et al., Nat Human Behav 2020) to try to disambiguate categorical biases from stimulus nonuniformities, but the results were inconclusive, probably because categorical biases, stimulus space nonuniformities, and individual differences interact in an unknown way. These results imply that color categories reflect an interaction of stimulus space nonuniformities and individual differences.

Supported by NIH IRP (NIH, 1ZIAEY000558 to BRC), NSF (0918064 to BRC) and NIH (R01 EY023322 to BRC). Contributions of NIH authors are considered Works of the United States Government. The findings and conclusions are those of the authors and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

43.475 RAPID FORMATION OF A WORD-SELECTIVE AREA IN MACAQUE IT THROUGH VISUAL EXPERIENCE

Jia Yang¹ (jiayang@pku.edu.cn), Haoxuan Yao¹, Kesheng Wang¹, Jingqiu Luo¹, Qianhua Zhu¹, Shiming Tang¹, Zhonglin Lu², Pinglei Bao¹; ¹Peking University, ²New York University Shanghai

Category-selective regions in primate inferotemporal (IT) cortex are defined by strong responses to specific categories. Yet with accumulating experience, familiar stimuli typically elicit reduced responses. This apparent paradox raises a key question: how can visual experience, which often suppresses responses, drive the emergence of highly responsive, category-selective regions? Here, we used longitudinal widefield calcium imaging to track the anterior IT (AIT) cortex in two macaques across approximately 20 sessions of passive exposure to Chinese words. Thirty fixed words, each with 10 exemplars varying in size, position, and rotation, were presented consistently as trained words. In each session, 10 novel words with 10 exemplars were introduced as controls. Additionally, we paired each set of 10 trained words with a distinct reward magnitude of one, two, or three drops of juice. Remarkably, training induced the progressive formation and expansion of a word-selective area. This newly formed region exhibited enhanced selectivity for trained words, characterized by stronger responses and improved decoding performance, including coarse decoding of reward level and finer decoding of character identity within each reward condition. At the population level, learning reorganized representational geometry by increasing inter-word separation, accompanied by expanded yet more elongated individual-word manifolds. Pre- and post-tests further revealed generalization to untrained words, yielding stronger responses and higher decoding accuracy. Critically, these benefits were evident at the start of the post-test, suggesting direct transfer driven by signal enhancement rather than an improved ability to learn new words. Together, these findings indicate that during the early stages of category-selective area formation, experience can operate through signal amplification to rapidly boost selectivity and develop a general and structural representational space for a specific category in the primate IT cortex.

This work was supported by National Science and Technology Innovation 2030-Major Project 2022ZD0204803, Natural Science Foundation of China NSFC32230043 to P. B., Natural Science Foundation of China Grant 32200857, China Postdoctoral Science Foundation Grant 2023M740125, 2022T150021 to J.Y.

43.476 OVERCOMING LOCATION-SPECIFIC CATEGORY LEARNING

Luke Rosedahl^{1,2}, Isabella Cabaleiro¹, Yuka Sasaki², Takeo Watanabe²; ¹Florida Atlantic University, ²Brown University

Category learning is fundamental to how we interact with the world and supports performance in many important professional domains. Unlike visual perceptual learning (VPL), category learning was historically thought to generalize across locations, reflecting the belief that it uses location-invariant visual information. However, recent work found that learning for information-integration (II) categories (categories that combine information across different feature dimensions) is location-specific (Rosedahl, Eckstein, and Ashby, *Nature Human Behaviour* 2018; Rosedahl and Watanabe, *JCE* 2024). However, the mechanism of specificity remains unclear, partly because no paradigm has yet induced location transfer. Here, we tested whether performing a subsequent unrelated task with the categorization stimuli, combined with feedback, can induce location invariance. We utilized a paradigm shown to induce VPL location transfer (subsequent double-training) either with or without response feedback for the double-training task. Participants performed category learning in the trained visual field location followed by a 2IFC contrast discrimination task with the categorization stimuli on the opposite side of the visual field and finished with category learning post-test for both locations. Half the participants (N=15) received response feedback for the 2IFC task while the other (N=15) did not and eye-tracking was used throughout to enforce fixation. While the group without 2IFC feedback showed location specificity, the group with 2IFC feedback showed no location specificity. This work provides the first paradigm that induces location transfer of category learning, addressing a key gap in understanding location-specific learning. One possible explanation is that feedback during double-training triggered reinforcement-related neural processes (such as spatially-diffuse dopamine-mediated plasticity in the striatum) that enabled mapping from previous category learning to the new visual location. More broadly, this work adds to the growing list of similarities between category learning and perceptual learning and raises the question of whether feedback also plays a critical role in VPL location transfer.

This work was supported by the National Eye Institute of the National Institutes of Health under award numbers [K99EY034891, R00EY034891, R01EY019466, R01EY027841, and R01EY031705] and NSF-BSF under award number [BCS2241417].

TUESDAY MORNING POSTERS IN BANYAN BREEZEWAY

Perceptual Training, Learning and Plasticity: Neural mechanisms

**TUESDAY, MAY 19, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY**

53.301 A BEHAVIORAL MARKER OF STRUCTURAL REORGANIZATION IN ADULT V1 AFTER LONG-TERM DEPRIVATION

Jisu Kang¹, Daniel Dilks¹; ¹Emory University

When input to a region of adult primary visual cortex (V1) is removed, creating a scotoma, neurons within the deprived region begin responding to stimuli that normally activate adjacent cortex only. In typically sighted individuals, this reorganization can occur within minutes of reversible deprivation and produces a perceptual distortion: a square flashed for 150 ms next to the scotoma is seen as a rectangle stretching into it. Although such rapid reorganization has been linked to changes in GABA-mediated inhibition, little is known about adult V1 reorganization after decades of deprivation. To investigate long-term deprivation, we tested three one-eyed individuals whose V1 regions corresponding to the intact eye's natural blind spot had been deprived for over 30 years. With brief stimulus presentations (150 ms), long-term deprivation produced the same magnitude of stretching seen after short-term deprivation. But with longer presentations (750 ms) a key difference emerged: the distortion persisted only after long-term deprivation, not short-term deprivation. The persistent perceptual distortion may serve as a behavioral marker of structural – rather than purely functional – reorganization in adult V1 as a result of long-term deprivation.

53.302 DECODING V1 REPRESENTATIONS OF STRABISMIC-EYE INPUT IN DIPLOPIC VERSUS NON- DIPLOPIC STRABISMUS

Shuai Chang¹ (aaronchangshuai@outlook.com), Yan Yang², Shuying Liu¹, Xinping Yu¹, Ming Meng³; ¹Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China, ²Sichuan Provincial People's Hospital, School of Medicine, University of Electronic Science and Technology of China, ³University of Alabama at Birmingham

Strabismus can lead either to persistent diplopia or to chronic suppression of input from the strabismic eye, implying distinct cortical strategies for handling mismatched binocular signals. Yet it is unknown how primary visual cortex (V1) represents strabismic-eye input in these divergent outcomes. We combined task-based fMRI with multivariate pattern analysis (MVPA) to compare V1 representations in strabismic patients with diplopia, strabismic patients without diplopia, and visually normal controls. Participants viewed colored orthogonal gratings under three conditions: dominant eye alone, strabismic eye alone, and binocular stimulation. Retinotopic mapping defined an individual V1 region of interest. We analyzed univariate BOLD amplitudes and multivoxel patterns using representational similarity and decoding of eye-of-origin and viewing condition. In diplopic strabismus, V1 showed little reduction of binocular BOLD responses relative to monocular viewing, and MVPA revealed robust decoding of dominant versus strabismic eye as well as monocular versus binocular trials. Monocular and binocular patterns were less similar when the strabismic eye was stimulated, consistent with

parallel, separable V1 representations of inputs from the two eyes. In non-diplopic strabismus, dominant-eye responses were relatively preserved, whereas binocular BOLD responses were suppressed and patterns elicited by strabismic-eye and binocular stimulation were poorly decodable and more similar to dominant-eye patterns, indicating loss or remapping of strabismic-eye-specific information in V1. Controls showed well-separated patterns for all conditions with high decoding accuracy. These results suggest that the presence versus absence of diplopia in strabismus is linked to distinct V1 coding regimes for strabismic-eye input: maintained, decodable eye-of-origin representations versus early suppression and restructuring before signals reach conscious awareness.

53.303 TURNING OFF IMPLICIT LEARNING: PUPILLARY BIOFEEDBACK AS A SUPPRESSOR OF TASK-IRRELEVANT PLASTICITY

Aaron Cochrane¹, Takeo Watanabe¹, Yuka Sasaki¹; ¹Brown University

Human vision is shaped by our environments through both explicit goal-oriented and implicit goal-irrelevant interactions with stimuli, each of which can induce plasticity. The mechanisms underlying implicit perceptual learning have far-reaching implications due to the presumed ubiquity of such learning in development, rehabilitation, and occupational contexts. However, these mechanisms remain poorly understood. Here, we investigated the possibility of volitional influences on implicit learning. Specifically, we utilized a recently-developed biofeedback method for modulating pupil size, associated with noradrenergic activity, to investigate how such modulation affects implicit perceptual learning. Participants ($n = 8$) were instructed on some trials to dilate and on others to constrict their pupils, with real-time visualization of pupil diameter displayed in front of a random dot-motion stimulus. A third trial type (control) involved passive viewing of the stimulus without pupil-size manipulation. Over seven days of training, participants improved their ability to both dilate and constrict their pupils. Unbeknownst to participants, dilation, constriction and control trials were each associated with different directions of perithreshold coherent motion, separated by 120 degrees. No participants reported noticing coherent motion during the biofeedback procedure. From pre-test to post-test on a 3AFC motion discrimination task, participants improved their perceptual threshold for motion direction associated with pupillary constriction as well as for the control direction. In contrast, thresholds for the direction associated with pupillary dilation did not improve, and the change in threshold associated with dilation was reliably smaller than for either constriction or control. These results are compatible with previous findings that increased noradrenergic activity associated with pupil dilation may enhance goal-relevant stimulus processing in a way that suppresses goal-irrelevant features. Such volitional control trained via biofeedback could be useful in contexts where implicit, goal-irrelevant learning is undesirable, such as occupational training or sequential tasks wherein retrograde interference may be detrimental.

We thank funding to YS and TW: NIH (R01EY019466, R01EY027841, R01EY031705, P30GM149405); NSF-BSF (BCS2241417)

53.304 INTER-OCULAR PERCEPTUAL LEARNING IS LOCATION AND EYE-COMBINATION SPECIFIC

Kate Walsh¹ (kate.walsh@colorado.edu), Rosemary Cowell¹, David Huber¹; ¹University of Colorado Boulder

Monocular perceptual learning transfers to the untrained eye (e.g., Lu et al., 2005), indicating learning within binocular brain regions. This implies an absence of learning within monocular brain regions, but it is not clear if this finding applies to between-eye, dichoptic learning. To address this question, we developed a novel inter-ocular perceptual learning paradigm in which perception of a Kanizsa stimulus (an illusory rectangle) can only be achieved by combining information across the eyes. This was achieved with a pair of pacman-shaped objects with one presented to one eye and the other presented to the other eye. This between-eye training condition was compared to within-eye training. Over six days of training, observers ($N=14$) judged the alignment of a monocularly-presented stimulus pair ("within-eye") trained in one visual quadrant and a dichoptically-presented pair ("between-eye") trained in the diagonally opposite quadrant. Over two transfer sessions, observers judged alignment of stimulus pairs, trained and untrained, across all quadrants to investigate the eye- and location-specificity of monocular and inter-ocular learning. Psychometric functions exhibited an above zero flat portion as a function of increasing misalignment, prior to the first increase in accuracy, indicating a bias to respond 'misaligned'. A novel signal-detection based function was developed to capture this response bias, thus allowing measurement of the point of initial rise, which we termed the 'limen' (i.e., the smallest noticeable misalignment). Although performance was generally better for trained quadrants and trained stimuli, this learning did not reliably map onto specific parameters of the psychometric curve, with the exception of the limen, which revealed a hyper-specific training effect for the particular trained eye-combination in the quadrant that experienced between-eye training. This specificity supports the conclusion that between-eye learning occurred within monocular brain regions.

53.305 DISRUPTION OF RECURRENT PROCESSING BY BACKWARD MASKING ELIMINATES LOCATION SPECIFICITY IN VISUAL PERCEPTUAL LEARNING: EVIDENCE FROM PSYCHOPHYSICS AND TIME- AND FREQUENCY- RESOLVED EEG DECODING

Yusuke Nakashima¹, Akane Kinumaki¹, Jessica Ding¹, Angie Boza¹, Luciana Chiu¹, Eileen Y Jang¹, Yuka Sasaki¹, Takeo Watanabe¹; ¹Brown University

Specificity to the trained retinal location is a core characteristic of visual perceptual learning (VPL), yet the neural mechanisms underlying location specificity remain unclear. We previously showed that location specificity was eliminated when training on an orientation discrimination task was conducted with backward masking, whereas

specificity remained with no masking and with forward masking. Backward masking is thought to disrupt later processing phases, possibly including recurrent processing, whereas forward masking disrupts early phases where feedforward processing may predominate. If these differential roles are valid, the loss of specificity under backward masking suggests that recurrent processing may contribute to location specificity, although direct evidence remains limited. Here, we tested these roles by examining which frequency bands of neural oscillations are affected by backward and forward masking. Neurophysiological studies suggest that theta and gamma neural oscillations carry feedforward signals, whereas alpha and beta bands carry recurrent signals. Thus, if backward and forward masking affect decoding accuracy in recurrent-related (alpha or beta) and feedforward-related (theta or gamma) frequency bands, respectively, such a pattern would support the two-roles model and clarify the role of recurrent processing in location specificity. Participants performed an orientation discrimination task under no masking, backward masking, and forward masking. Decoding accuracy for orientation classification (clockwise vs. counterclockwise) was calculated for theta, alpha, beta, and gamma bands at each timepoint. In the no masking condition, decoding accuracy increased in theta and alpha bands. Backward masking reduced alpha accuracy while sparing theta, whereas forward masking reduced theta and alpha accuracy but left beta accuracy intact. These results suggest that backward masking disrupts recurrent processing reflected in alpha, whereas forward masking disrupts feedforward processing reflected in theta. Together with our previous findings, these results indicate that recurrent processing during training plays an important role in location specificity in VPL.

NIH R01EY019466, R01EY027841, R01EY031705, NSF-BSF BCS2241417

Perceptual Training, Learning and Plasticity: Neuroimaging, neurostimulation

TUESDAY, MAY 19, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY

53.306 DIFFUSION AND QUANTITATIVE MRI PROFILES OF HUMAN VISUAL WHITE MATTER TRACTS OF EARLY BLIND INDIVIDUALS

Kelly Chang¹ (kchang4@uw.edu), John Pyles¹, Ariel Rokem¹, Ione Fine¹, Woon Ju Park²; ¹University of Washington, ²Georgia Institute of Technology

Early visual deprivation alters the structural development of the visual system. Prior work found decreases in white matter diffusion fractional anisotropy (FA) in the visual pathways of congenitally blind individuals (Reislev et al., 2016; Schoth et al., 2006; Shimony et al., 2006; Shu et

al., 2009). Decreases in FA are often interpreted as evidence of reduced myelination in visual pathways following early blindness, but they may also reflect changes in white matter spatial configuration (e.g., axon tortuosity). To distinguish diffusivity-driven changes from myelin-sensitive tissue properties, we collected multi-shell diffusion MRI and quantitative R1 maps (multiparameter mapping sequences) from 8 early blind individuals (3 females, 55.6 ± 15.6 SD y.o.) and 8 sighted controls (4 females, 53.9 ± 13.2 SD y.o.). Vision-related white matter tracts, including the optic radiations (OR), vertical occipital fasciculus (VOF), and forceps major, were segmented using pyAFQ (Kruyer et al., 2021, 2025), along with two non-visual comparison tracts (acoustic radiations and motor corpus callosum). Diffusion metrics from the diffusion kurtosis model, such as FA, kurtosis FA (KFA), mean diffusivity (MD), and axial diffusivity (AD), were sampled at 100 equidistant nodes along each tract. Quantitative R1 maps were sampled along the same tract profiles to assess myelin-sensitive tissue properties. Early blind participants exhibited lower FA, KFA, and AD than sighted participants across all vision-related tracts, except the left VOF. In contrast, R1 values were not significantly different across groups in these tracts. No significant group differences were observed in non-visual tracts in either diffusion or R1 values. The combined diffusion and quantitative MRI results suggest that early blindness is associated with reduced axonal microstructural complexity rather than large-scale reductions in myelination along visual white matter tracts. The findings challenge the common assumption that reductions in diffusivity are associated with decreased white matter myelination following early blindness.

NEI R00 EY034546 and Georgia Tech Smithgall-Watts Early Career Award to WP; NEI R01EY014645 to IF.

53.307 FEASIBILITY OF MEASURING GABA+ IN THE VISUAL CORTEX OF CHILDREN TO INVESTIGATE CRITICAL PERIOD NEUROPLASTICITY IN AMBLYOPIA

Rebecca Willis¹ (rebecca.willis@ndcn.ox.ac.uk), Megan Groombridge¹, Aaron T. Hess¹, William T. Clarke¹, Juliet Semple², Jon Campbell¹, Pete R. Jones^{3,4}, Tessa M. Dekker^{4,5}, Dinesh K. Deelchand⁶, Pierre-Gilles Henry⁶, Ravi Purohit⁷, Holly Bridge¹, I. Betina Ip¹; ¹The University of Oxford Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford, UK, ²Oxford Centre for Human Brain Activity, Department of Psychiatry, University of Oxford, UK, ³Department of Optometry and Visual Sciences; School of Health and Medical Sciences; City St George's, University of London, UK, ⁴Institute of Ophthalmology, University College London, UK, ⁵Experimental Psychology, Division of Psychology and Language Sciences, University College London, UK, ⁶Center for Magnetic Resonance Research, Department of Radiology, University of Minnesota, Minneapolis, MN, USA, ⁷Oxford Eye Hospital, Oxford University Hospitals NHS Trust, UK

Amblyopia is a common neurodevelopmental condition, marked by extreme eye dominance. While amblyopia in adults has been

associated with GABAergic inhibition in the visual cortex (Mukerji et al. 2022; Ip et al. 2024), the role of GABA in childhood amblyopia is unexplored. We created a child-friendly protocol to measure GABAergic inhibition in children. Here we present the protocol and preliminary data from controls. 5-8 year old participants have two visits six months apart, between which amblyopes have occlusion therapy and controls have no intervention. Each visit includes an MRI scan and vision tests. The MRI comprises motion corrected T1-structural and magnetic resonance spectroscopy (MRS) sequences, and resting state fMRI. MRS is conducted in early visual cortex (EVC) and sensorimotor cortex (M1 control) voxels. Child-friendly vision tests are used to measure visual acuity (Keeler Crowded LogMAR), contrast sensitivity (PopCSF, Elfadaly et al. 2020 and FrACT, Bach 1996) and stereopsis (Frisby; Asteroid, Vancleef et al 2019). In controls (N=14, mean age=6.96 years), 91.3% of attempted MRS scans resulted in data acquisition. Of those data, 85.7% of scans could be analysed without further processing and are reported here. When analysed using a standard basis set, GABA+ concentration was comparable between voxels ($F(1,3)=4.32$, $p=0.13$) and timepoints ($F(1,3)=3.12$, $p=0.09$). Absolute CRLB indicated that MRS data quality was good (mean= 0.53, SD=0.70) and did not differ between voxels ($F(1,3)=0.79$, $p=0.44$) or timepoints ($F(1,3)=1.32$, $p=0.33$). Visual acuity (logMAR mean=-0.05, SD=0.08), contrast sensitivity (popCSF AUCSF, mean=29.4, SD=12.8), and stereopsis (Asteroid arcsec=77.8, SD=90.6) were within the range for normal vision. These preliminary results demonstrate that our protocol can reliably obtain high quality non-invasive measures of GABA+ concentrations from 5-8 year old children. This feasibility is an important step for improved understanding of neuroplasticity during the critical period in humans.

This research is funded by The Royal Society (studentship to R.W., Dorothy Hodgkin Research Fellowship to I.B.I.) and the Medical Research Council (MR/V034723/1).

53.308 SHORT-TERM PLASTICITY OF THALAMO-CORTICAL CONNECTIVITY MEASURED WITH DYNAMIC CAUSAL MODELLING OF ULTRA-HIGH FIELD FMRI

Miriam Acquafredda¹, Giacomo Mazzotta¹, Laura Biagi^{2,3}, Michela Tosetti^{2,3}, Maria Concetta Morrone¹, Paola Binda¹; ¹Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy, ²IRCCS Stella Maris, Calambrone, Pisa, Italy, ³IMAGO7 Center, Pisa, Italy

In adult humans, brief periods of monocular deprivation induce ocular dominance plasticity, transiently enhancing responses to stimuli in the deprived eye in the primary visual cortex V1 (Binda et al., 2018) and the ventral pulvinar (Kurzwaski et al., 2022). The connectivity among these regions measured in resting state is also affected, with a selective reduction of the pulvinar-to-V1 pathway (Acquafredda et al., Science Advances 2025). Here we investigated whether and how the connectivity between V1 and the visual thalamus is modulated by visual stimulation. We acquired ultra-high-field 7T fMRI EPI sequences from 25 normally sighted adults, pre- and post-2 hours of monocular

deprivation. Participants were presented with monocular band-pass noise stimuli with five contrast levels. Dynamic Causal Modeling (SPM implementation, Zeidman et al., 2019) was used to analyse BOLD signals from three regions of interest: V1, lateral geniculate nucleus, and ventral pulvinar – all showing reliable visually evoked activity. Short-term deprivation differentially modulated V1 connectivity with the two thalamic regions. During deprived eye stimulation, the connectivity between V1 and the pulvinar decreased post-deprivation (like in resting state), while the connectivity between V1 and LGN increased. The opposite was observed during stimulation of the non-deprived eye, resulting in a significant three-way interaction (time x eye x region of interest, $F(1,24) = 5.85$, $p = 0.024$). Across participants, the modulation of connectivity between V1 and the pulvinar (but not with LGN) was reliably correlated with the shift of ocular dominance, measured with short binocular rivalry sessions before and after the fMRI acquisitions. We conclude that monocular deprivation produces changes that extend beyond local V1 processes and affect the connectivity between the cortex and the visual thalamus. Our results indicate that the pulvinar participates in regulating activity in the visual cortex and setting its short-term plasticity.

European Research Council (ERC) Consolidator grant PredActive (101170249); European Union–Next Generation EU, grant PRIN 2022 (“RIGHTSTRESS - Tuning arousal for optimal perception” Settore SH4, 2022CCPJ3J - CUP I53D23003960006); MUR, program FARE-2 (grant SmilY); Fondazione Pisa (305/22).

53.309 HIGH-FREQUENCY, BUT NOT LOW-FREQUENCY, TRANSCRANIAL RANDOM NOISE STIMULATION (TRNS) SHARPENS VISUAL TUNING

Simay Uner^{1,2} (simayuner.nsc@gmail.com), Irem Akdogan^{1,2}, Hulusi Kafaligonul^{1,2,3}; ¹Department of Neuroscience, Aysel Sabuncu Brain Research Center, Bilkent University, Ankara, Türkiye, ²National Magnetic Resonance Research Center (UMRAM), Bilkent University, Ankara, Türkiye, ³Neuroscience and Neurotechnology Center of Excellence (NÖROM), Department of Anatomy, Faculty of Medicine, Gazi University, Ankara, Türkiye

Transcranial random noise stimulation enhances visual perception, a benefit most consistently observed with high-frequency (hf-tRNS) compared to low-frequency (lf-tRNS) protocols. Most existing work typically tests single stimulus parameters, limiting our understanding of how electrical noise influences broadband visual sensitivity. In addition, it remains unclear whether hf-tRNS and lf-tRNS modulate vision via similar mechanisms or if they induce distinct alterations in cortical tuning. To characterize frequency-dependent effects on overall visibility, we measured the foveal Contrast Sensitivity Function (CSF) in 57 participants across two experiments, using the quick CSF paradigm. tRNS was applied over the visual cortex at 1.5 mA. Participants received either lf-tRNS (n = 26) or hf-tRNS (n = 31) while performing an orientation discrimination task to estimate the full CSF. We analyzed key morphological parameters, including maximum contrast sensitivity (CS_{max}), peak spatial frequency (SF_{peak}),

bandwidth, and area under the curve (AUCSF). Results indicated a dissociation between stimulation protocols, such that hf-tRNS did not uniformly boost contrast gain. It rather selectively enhanced sensitivity at mid-spatial frequencies while suppressing sensitivity at the low- and high-frequency tails, effectively narrowing the CSF bandwidth. In contrast, lf-tRNS failed to induce such morphological changes, signaling a fundamental divergence in efficacy. This sharpening effect in the hf-tRNS group was confirmed by a significant negative correlation between the changes in bandwidth and CSmax, a relationship absent in the lf-tRNS group. These findings demonstrate that hf-tRNS but not lf-tRNS alters the fundamental tuning of the visual system. Unlike the broad effects often attributed to general excitability changes, hf-tRNS appears to sharpen visual channels around mid-spatial frequencies, suggesting a stochastic resonance mechanism that selectively enhances signal-to-noise ratios in optimally tuned processing channels. We discuss potential mechanisms underlying this dissociation, suggesting that high- and low-frequency noise may interact differently with cortical processing dynamics to modulate visual sensitivity.

This work was supported by The Scientific and Technological Research Council of Türkiye (ARDEB 424K278, BIDEB 2211 Program) and the BAGEP Award of the Science Academy-Türkiye.

53.310 THE EFFECTS OF PERCEPTUAL LEARNING ON STIMULUS-DRIVEN NEURAL ACTIVITY IN EARLY VISUAL AREAS

Pinar Demirayak¹, Pauline Stewart¹, Elam Cutts¹, Madeline Ragland¹, Elliot Maxwell¹, Marcello Maniglia², Samyukta Jayakumar³, Jaap Munneke⁴, Nicholas Turk-Browne⁵, Aaron Seitz⁴, Kristina Visscher¹; ¹University of Alabama at Birmingham, ²Rochester Institute of Technology, ³University of California at Riverside, ⁴Northeastern University, ⁵Yale University

People who experience central vision loss due to retinal damage in the macula frequently adopt a spontaneous “preferred retinal locus” outside the damaged area, which they rely on for visually demanding everyday activities. To investigate the neural plasticity that supports this compensatory strategy, we implemented a simulated scotoma paradigm in healthy participants. Sixty-eight participants were trained on contrast sensitivity and contour integration tasks, at peripheral locations (7.5° from the center) either to the left or right of an artificial scotoma. Peripheral training location – either right or left – was assigned to each participant before training. We collected a total of six runs of task-based fMRI data in each pre- and post-training MRI session. We used a General Linear Model to examine average BOLD activity during task performance for both pre and post MRI sessions. We identified locations in early visual areas where the stimulus drove a positive BOLD response. Moreover, we extracted cortical thickness for these regions. We observed that the stimulus-driven BOLD signal was increased in V1 following visual training. On the contrary, that signal decreased in V2. Our results indicate that experience influences stimulus-driven activity in V1 and V2 in different ways.

53.311 SHORT-TERM ANTI-HEBBIAN PLASTICITY OF VISUAL SURROUND SUPPRESSION THROUGH DECORRELATED CENTER-SURROUND STIMULATION

Nan Jin^{1,2}, Yongchun Cai¹; ¹Zhejiang University, ²Ludwig-Maximilians-Universität München

Recent computational and electrophysiological studies suggest that short-term co-activation of mutually inhibitory neural populations can strengthen their inhibitory interaction, consistent with anti-Hebbian plasticity, whereas temporally uncorrelated activation should theoretically weaken it. However, evidence for such inhibitory plasticity in the human visual system remains limited. Here, we tested whether the strength of visual surround suppression (a phenomenon in which responses to a central stimulus are reduced by peripheral stimulation) can be modulated by short-term decorrelated activity between central and peripheral populations. We recorded steady-state visual evoked potentials (SSVEPs) while participants (N = 30) viewed two 20-s flicker conditions: (1) central gratings flickering at 20 Hz, and (2) the same central gratings surrounded by a peripheral grating flickering independently at 24 Hz. The decorrelated 20/24 Hz condition induced sustained uncorrelated activity between center and surround populations. Surround suppression was quantified as the differences of the 20 Hz SSVEP signal-to-noise ratio (SNR) across the two conditions during the 20-second stimulation. To dissociate adaptation from inhibitory modulation, we implemented a divisive normalization model that accounted for independent adaptation within center and surround channels. As predicted, surround suppression progressively weakened during decorrelated (20/24 Hz) stimulation, indicating a rapid reduction of inhibitory influence between center and surround populations. Our results demonstrate that inhibitory interactions underlying visual contextual modulation are not static but dynamically adjust to the recent input statistics, providing neural evidence for short-term anti-Hebbian plasticity in the human visual cortex. This dynamic modulation may represent a fundamental mechanism by which the visual system maintains flexible sensitivity to contextual changes in the environment.

53.312 REACTIVATION DURING NREM SLEEP CONTRIBUTES TO ORIENTATION SPECIFICITY IN VISUAL PERCEPTUAL LEARNING—REVEALED BY DECODING OF FMRI SIGNALS DURING SLEEP

Theodore LaBonte-Clark¹, Teruaki Kido¹, Takeo Watanabe¹, Yuka Sasaki¹; ¹Brown University

While visual perceptual learning (VPL) — long-term improvements on visual tasks through experience — is known to be enhanced by post-training sleep, the underlying neural mechanisms of sleep-dependent improvement remain unclear. A leading hypothesis is that sleep facilitates learning through reactivation, the re-emergence of training-related neural activity during sleep. VPL is characterized as specific to the trained feature. Here, we tested a novel model proposing that reactivation contributes directly to such feature specificity. In the

experiment, participants were trained to detect Gabor patches with a single orientation. During the post-training sleep, we recorded fMRI activity from early visual areas simultaneously with polysomnography to determine sleep stages. We used an fMRI decoder, constructed in a previous session and capable of classifying four Gabor orientations, including the trained orientation, to evaluate representations in visual cortex during sleep. In the event that VPL occurred, the decoder identified neural activity corresponding to the trained orientation above chance, suggesting selective reactivation of the trained orientation during sleep. Furthermore, greater decoding performance for the trained orientation appeared to accompany greater performance improvement after sleep. This pattern suggests that neural reactivation during NREM sleep may reflect the trained orientation and support the proposal that reactivation plays a key role in mediating the orientation specificity characteristic of VPL.

NIH R01EY019466, R01EY027841, R01EY031705, NSF-BSF BCS2241417

Perceptual Organization: Intuitive physics

TUESDAY, MAY 19, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY

53.313 COGS IN A MACHINE: A “ROTATING GEARS” ILLUSION OF INCREASED SPEED AND CONTRAST FROM CAUSAL ENMESHMENT

Jacky Yueh¹, Joan Danielle K. Ongchoco¹; ¹University of British Columbia

We perceive not only objects and motions but also the relationships between them—as in the perception of causality. But causal interactions occur not only in simple collisions, but also in complex causal systems, such as pulleys and gears. These examples are typically associated with effortful and abstract reasoning, but here, in contrast, we introduce a “rotating gears” illusion in which sustained causal dynamics seem to reshape basic visual properties. When observers fixate on one of two rotating gears (with directly aligned teeth), the adjacent gear will appear to rotate faster even though they have the same speed, due to greater motion sensitivity in peripheral vision. But when equidistant gears are properly offset so their teeth slot nicely in between each other (making them *causally enmeshed*), the adjacent gear seems to rotate even faster. We verified this experimentally: Observers saw pairs of gears and reported how much faster the adjacent gear rotated relative to the one fixated on. Speed estimates were greater for causally-enmeshed gears versus not. These were reliably different from estimates in critical control conditions, such as when observers fixated at the center (so we shouldn’t expect an illusion of speeding-up) and when observers saw “poker chip”-like discs whose edges matched the alternating black and

white gear teeth and notches but could not be causally enmeshed (isolating causal connection from mere alternating black and white motion). Causal enmeshment also generates an entirely different illusion when we modify the transparency of the adjacent gear. In an independent experiment, we verified that a semi-transparent adjacent gear appears more visible when causally enmeshed with the gear that is fixated on, compared to when it is not. Causal enmeshment can sharpen visual properties of speed and contrast, suggesting that principles of more complex causal dynamics may be more deeply embedded in visual processing.

53.314 CAUSAL PROPERTY TRANSFER: A NEW PERCEPTUAL EVENT TYPE CONSTRAINED BY INTUITIVE PHYSICS

David Schwitzgebel¹, Brent Strickland¹, Alon Hafri²; ¹École Normale Supérieure—Paris Sciences et Lettres, ²University of Delaware

Causal perception typically concerns motion: when one object strikes another, momentum appears to be “transferred” from Agent to Patient. Yet many causal interactions involve non-spatiotemporal properties: a wet towel moistens skin, an ice cube cools a drink, and a paintbrush imparts its color to a surface. Does visual perception have automatic “expectations” of property transfer from Agents to Patients (but not the reverse), and are such representations constrained by intuitive physics? Across six experiments (n=869), we identify a novel class of physical events—causal property transfer—in which an Agent’s contact with a Patient produces a change in the Patient’s non-spatiotemporal properties (color, shape). In a rapid go/no-go task, participants viewed displays in which an Agent object launched a Patient object; unlike standard Michottean launching displays, one object changed color upon contact. Participants’ task was to detect whether that color matched a prespecified target color. They responded faster and more accurately on trials in which the Patient took on the Agent’s color relative to trials with incongruent color changes and trials in which the Agent took on the Patient’s color, revealing an automatic, selective visual expectation that Agents transfer their properties to Patients. This effect was disrupted when physical plausibility was eliminated (e.g., delaying the post-contact color change), suggesting sensitivity to intuitive physical constraints. Notably, it persisted when the Patient remained static, indicating reliance on a mechanism distinct from classic Michottean launching. Additional experiments showed similar effects for shape change, but again only when the transformation was physically plausible, suggesting that visual expectations about causal property transfer generalize beyond color to object properties more broadly. Taken together, these findings expand the domain of causal perception from motion to non-spatiotemporal properties and show that this novel form of causality, like classic motion-based causality, is constrained by intuitive physics.

53.316 ‘TIGHT-FITTING’ VS. ‘LOOSE-FITTING’ AS A VISUAL PRIMITIVE FOR EVENT PERCEPTION: EVIDENCE FROM CATEGORICAL PERCEPTION

Suppose you go jogging while wearing a ring on your finger, and earphones in your ears. Will the ring fall off, or will the earphones fall out? The answers will depend not so much on the relation itself ('on', 'in'), but on the degree to which the ring/earphones fit **tightly** vs. **loosely**. This distinction is so important that it can even have lexical priority — e.g. in Korean, where you might use the same term to indicate putting a ring **on** your finger, or earphones **in** your ears, as long as the fit is similarly tight. Given the power of this distinction, might it also be prioritized even in visual processing? To find out, we exploited the logic of **categorical perception**, with simple animations of a cylinder moving down into a container. English speakers viewed sequential pairs of such events, and simply reported whether they were identical, vs. different in any way. Both the cylinder and the container could vary in diameter across animations, to implement tight fits (e.g. narrow cylinders moving into narrow containers) or loose fits (e.g. narrow cylinders moving into wide containers). Observers were much better at noticing Cross-Type changes (e.g. from a tight-fit to a loose-fit animation) than Within-Type changes (e.g. two different tight-fit animations), even though (1) the tight/loose distinction was always task-irrelevant, and (2) Within-Type changes were always greater in objective magnitude. Moreover, these results were specific to the tight vs. loose fits themselves, since these effects disappeared when the same movements (and sizes) occurred in the context of occlusion (where fit is irrelevant) instead of containment (where fit is consequential). This work thus demonstrates a new form of intuitive physics that may serve as an underlying visual primitive for dynamic event perception.

53.317 ADAPTATION-INDUCED REVERSAL OF CAUSAL AGENT-PATIENT RELATIONSHIPS IN VISUAL SEPARATION EVENTS

Jiyoung Hwang¹, Jimin Ju^{1,2}, Sung-Ho Kim¹; ¹EWHA WOMANS UNIVERSITY, ²UNIVERSITY OF IOWA

People effortlessly recognize objects of various materials and predict their behavior from visual information, drawing on how different materials respond to external forces—fundamentally, pushes or pulls. Despite this, research on causal perception has largely underexplored dynamic interactions involving non-rigid objects, often treating spatiotemporal causality as independent of object properties. Using ambiguous separation events, Ju and Kim (2026) demonstrated that simple kinematic information can simultaneously shift material perception and reverse causal-role assignment: when a line segment elongates with a disc at its end and then separates, pre-separation acceleration evokes the impression of a rigid stick pushing the disc, whereas deceleration evokes that of a disc pulling an elastic band. To test whether these causal-role assignments arise from perceptual mechanisms, we employed a visual adaptation paradigm. Eighty participants viewed one of four adaptor types created by crossing acceleration profile (acceleration/deceleration) with the presence/absence of a 200-ms pause inserted during the line's

retraction at separation. They then judged whether ambiguous test events appeared to depict a push or a pull. Adaptors without a pause revealed bidirectional aftereffects: the acceleration adaptor (pushing-percept-dominant) increased subsequent pull responses, whereas the deceleration adaptor (pulling-percept-dominant) increased push responses. However, a 200-ms pause modulated these effects differentially depending on the acceleration profile. For the acceleration adaptor, the pause preserved the aftereffect. In contrast, for the deceleration adaptor, the pause increased subsequent pull responses, possibly because it disrupted the expected tension-release cue of elastic recoil, thereby inducing the perception of the line segment as rigid rather than elastic. These findings indicate that adaptation occurred at a level integrating motion segments into coherent force dynamics, rather than at the level of low-level motion features. More broadly, they suggest that push-pull interactions are perceptually encoded along a continuum that supports intuitive causal reasoning in everyday physical events.

53.318 CAUSAL SUPPORT SHAPES THE PERCEPTUAL SALIENCE OF DISRUPTIONS IN SINGLE-OBJECT MOTION

Sujin Lee¹ (chance0305@gmail.com), Sung-Ho Kim²; ¹Ewha Woman University

Human observers do not perceive object motion solely in terms of its physical trajectory but also in terms of the causal structure that generates it. Here, we investigated how different causal framings modulate the perception of brief pauses interrupting an otherwise continuous motion. On each trial, a single disc translated at a constant speed and turned once, with a 50-ms pause occurring at the turn on half of the trials. The turn occurred within one of three contexts: Bounce-Turn, where the disc contacted a small block and appeared to bounce; Pass-Over Turn, where the disc passed over the block and turned as it cleared the block's far edge; and Free-Turn, where no block was present and the turn occurred in empty space. Observers judged whether a pause occurred, and pause-detection sensitivity followed a systematic ordering: highest in Bounce-Turn, intermediate in Pass-Over Turn, and lowest in Free-Turn. Two interpretations may account for this pattern. A low-level account attributes the heightened sensitivity in Bounce-Turn to perceived impact force: the collision produces a vivid sense of an abrupt physical jolt, making a brief interruption a salient violation of expected motion continuity. The absence of such impact in the other conditions renders the same interruption less striking. A higher-level account suggests that pause salience depends on how the event modulates perceived animacy: a Newtonian violation in a mechanically constrained system (Bounce-Turn) may induce a larger transient shift toward agent-like interpretation than in Pass-Over Turn, where agentive construals are already more accessible. Although the present data do not decisively favor one account over the other, they call into question the idea that animacy itself drives attentional capture. Rather, the results indicate that salience may originate primarily from perceived violations of

Newtonian mechanics grounded in perceived force, independent of whether the motion is interpreted as animate.

53.319 CAUSAL ROLE ASSIGNMENT TUNES PREDICTION PRECISION IN UNFOLDING MOTION: HIGHER FOR PATIENTS THAN FOR AGENTS

Jimin Ju^{1,2} (jimin-ju@uiowa.edu), Sung-Ho Kim²; ¹University of Iowa, ²Ewha Womans University

In dynamic events, even minimal kinematic information can lead observers to assign different causal roles to the objects involved—agent versus patient—thereby specifying who acts on whom and how. Because such role assignments entail distinct expectations for how each object should continue to move, they should also modulate how precisely brief motion disruptions are detected. We evaluated this claim with a pause-detection task using ambiguous separation displays (Ju & Kim, 2026), in which a line segment with a disc attached at its end elongates and then separates: pre-separation acceleration evokes the impression of a rigid stick pushing the disc, whereas deceleration evokes a disc pulling an elastic band. On half the trials, a 100-ms pause was inserted into the post-separation motion of either the line or the disc (blocked), and observers (N = 30) judged whether a pause occurred. Signal-detection analyses revealed a robust Target × Pre-separation crossover: pause-detection sensitivity was higher for the disc following pre-separation acceleration—when a rigid stick pushing the disc is the dominant percept—than following deceleration—when a disc pulling an elastic band is dominant—whereas the line showed the reverse, with greater sensitivity after deceleration than acceleration; as such, in both cases, detection was enhanced when the probed object was perceived as the causal patient. These findings indicate that objects construed as inertial, patient-like recipients of external forces elicit tighter Newtonian predictions, making brief disruptions more salient as violations of expected motion. In contrast, when an object is interpreted as an active, self-propelled agent whose motion is less tightly constrained by external forces, the same kinematic irregularities become less noticeable. More broadly, within a force-resistance framework, the attribution of agent-patient roles can tune precision of motion predictions, altering sensitivity to transient deviations as dynamic events unfold.

53.320 CAUSAL CHAINS AS A DISTINCT TYPE OF EVENT IN VISUAL PERCEPTION

Emily Wedin¹, Joan Danielle K. Ongchoco¹; ¹University of British Columbia

The task of perception is to come up with a description of experience. Accordingly, visual processing seems to traffic in properties beyond color or motions, such as *causality*—e.g., one object launching another into motion. But what caused *you* to be here? The answer may be your flight to Florida, your December submission, or other choices along the way—but ultimately any causal chain can be traced all the way back to the Big Bang. Do causal chains belong solely to the

domain of reasoning, or might they also play a unique role in visual processing, stringing together events in time? Here, observers enumerated discs in fast, irregular sequences, with the display panning continuously to track up to 30 collisions. In the “causal chain” condition, each disc launched the next, while in the non-causal control, motions were identical, except for small spatial gaps between discs. Observers estimated *more* discs in the causal chain, demonstrating that chain structure helps retain more motions in working memory. To further isolate the “causal chain” from mere causality and from mere spatial continuity, we ran independent experiments with “paired” and “overlap” conditions. In the paired condition, collisions occurred between alternating disc pairs, preserving causality without chain structure. Estimates fell between causal chain and non-causal estimates, suggesting that causality alone does not allow the same degree of retention. In the overlap condition, each disc took over the next disc’s spatial position, providing continuity without causality. Estimates were again reliably different from the causal chain, in fact, even higher—suggesting that spatial continuity may generate a “one is more” illusion, but causality still provides additional structure. Causal chains—as sequences of temporally-linked causal interactions—may be a distinct type of event in vision, allowing us to *hold on* to more of what we see.

Perceptual Organization: Grouping

**TUESDAY, MAY 19, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY**

53.321 COMPROMISED SYMMETRY PERCEPTION FOLLOWING PROLONGED VISUAL DEPRIVATION

Priti Gupta¹ (gupta.priti.84@gmail.com), Stutee Narang¹, Purva Sethi¹, Shraddha Sharma¹, Naviya Lall¹, Meghna Rao¹, Manvi Jain¹, Chetan Ralekar², Lukas Vogelsang², Marin Vogelsang², Michal Fux², Suma Ganesh², Pawan Sinha²; ¹Project Prakash, Dr Shroff's Charity Eye Hospital, New Delhi, India, ²Department of Brain and Cognitive Sciences, MIT, USA

The human visual system is exquisitely sensitive to bilateral vertical symmetry. Although the roots of this proficiency are not definitively known, they may relate to the kinds of regularities present in ecologically important entities such as animals, faces, and plants. Even neonates are reported to exhibit sensitivity to vertical symmetry, suggesting that visual experience may play only a limited, if any, role in the genesis of this ability. Against this backdrop, here we ask whether the perception of bilateral symmetry is resilient to severe restrictions of early visual experience. To this end, we worked with children who had early onset blindness and received sight surgeries at the age of several years. We recruited two patient groups. The first (n = 15, mean age = 16.4y) had their sight surgeries two to four years prior to this study and therefore had a significant period of visual experience post-surgery. The second group (n=5, mean age = 12.6y)

was newly operated and had only had up to fourteen months of visual experience. The performance of these two patient groups was compared with controls ($n=13$, mean age = 17.2y). Our stimulus images comprised ten colored discs arranged in a vertically symmetric or non-symmetric configuration. Results reveal a significant compromise in symmetry-based classification across newly-sighted patients. For patients who had long periods of visual experience post their surgery, symmetry-based classification improves but does not reach control level. In summary, the data thus far indicate a profound role of early visual experience on symmetry perception and, perhaps, long-range visual analyses more broadly.

We are grateful to NIH for supporting this work

53.322 OLDER ADULTS' PATTERN DETECTION IN 2×2 GRAPHS

Nestor Matthews¹ (matthewsn@denison.edu), Megan Broderick²; ¹Denison University, ²The Pennsylvania State University

Introduction: Detecting statistically significant patterns in graphs requires ensemble perception—the ability to extract summaries rapidly from groups of similar elements. Prior work found systematic ensemble perception biases in younger adults (<65 years) when learning to detect statistical significance patterns in graphs (Matthews, Broderick & Kozlowski, 2025). Here, we examined whether those same biases extend to adults aged 65 years or older. Methods: We recruited 480 older adults (aged 65+) from Prolific. Participants viewed 2×2 graphs that plotted a y-axis variable against two predictors: one on the x-axis (Factor A) and one in the legend (Factor B). After random assignment to evaluate Factor A, Factor B, or their interaction, participants learned by trial-and-error to classify line or bar graphs as statistically significant or non-significant. Line and bar graphs displayed identical numeric information. Results: The older adults showed the same ensemble perceptual biases observed in younger adults. First, line graphs yielded significantly greater accuracy than bar graphs when detecting interactions or Factor B main effects. No such line graph advantage emerged for Factor A main effects, which produced low accuracy for both graph types. Second, participants tasked with detecting main effects in bar graphs often applied an incorrect cognitive strategy, misconstruing significant interactions as significant main effects. This incorrect classification rule produced more systematic errors (both false positives, and false negatives) in bar graphs than in line graphs, whether detecting Factor A or Factor B main effects. Conclusion: The findings show that these perceptual biases in graph reading persist across the adult lifespan. They also guide data visualization efforts to optimize graph readability and inform educators about what students likely miss—and incorrectly “see”—when learning to interpret 2×2 graphs.

Denison University Research Foundation

53.323 THE FOLLOW PARCEL PARADIGM: REVISITING VISUAL STIMULI IN LAGRANGIAN OBJECT-CENTRED COORDINATES

Francesca Peveri¹, Andrea Canessa¹, Silvio Paolo Sabatini¹; ¹University of Genoa

Traditional psychophysics treats participants as relatively passive observers: stimuli are generated by the experimenter, and responses are measured afterward. Recent continuous psychophysics approaches accelerate data collection and enable dynamic response analyses, yet they still consider screen-fixed coordinates, and position the observer in a largely reactive relationship to the stimulus. By analogy with fluid mechanics, current methods follow an Eulerian approach by defining visual input at fixed spatial locations and recording how observers respond to changes at those locations. Here, we propose a methodological shift: instead of limiting participant to generating responses, we engage them as active agents who explore and manipulate 3D visual stimuli in a closed perception-action loop. In this framework, sensory properties change contingently with the observer's actions, allowing us to characterize how exploration dynamics modulates sensory encoding and perceptual sensitivity. The key innovation is to follow a specific visual “parcel” as it evolves over time and to evaluate perceptual transformations along its trajectory. This treats the moving object, or feature cluster, as the reference frame of the experiment, corresponding to a genuinely Lagrangian description of visual input. We developed a combined experimental and computational framework to instantiate this paradigm. A virtual object was implemented in Unity3D and placed in peripersonal space. Its time-varying 3D pose was yoked to that of a real handheld display operated by the observer. Thus, the display serves simultaneously as the object's physical manipulator and as the aperture through which the object's appearance is sampled according to the contingent vantage point of the observer. We show that identical stimulus properties can produce different sensory statistics depending on the observer's actions, emphasizing the need for perceptual measures that surpass existing continuous paradigms. Incorporating sensorimotor contingencies, our method offers a quantitative framework for perception in true naturalistic action loops.

53.324 PERCEPTUAL AND SEMANTIC BOUNDARIES PLAY DISTINCT ROLES IN SHAPING TEMPORAL MEMORY

Bahar Aykut¹ (bahar.aykut@sabanciuniv.edu), Nihan Alp¹; ¹Sabanci University, Istanbul

A core challenge for perception and memory is organizing continuous streams of input into meaningful units. Prior research has shown that changes in perceptual context (frame color) or semantic content (category) influence how we structure experience. For example, changes in frame color can act as a visual context shift, grouping or separating information. In parallel, semantic changes are known to create event boundaries that often disrupt temporal memory.

However, naturalistic studies frequently conflate perceptual context and semantic content, and controlled paradigms rarely incorporate broader category change. Here, we investigated how perceptual context and semantic content (broader object categories such as animals or tools) interact to shape temporal memory. Participants ($N = 23$) viewed sequences of real-world objects in colored frames, where color and category changed separately (only color or only category) or simultaneously (both changed at the same time). Later, participants completed the temporal-order and temporal-distance judgment tasks. The results showed that perceptual context shifts driven by color changes significantly impaired temporal-order accuracy relative to a within-context baseline, suggesting that perceptual context changes may trigger an attentional reset that disrupts temporal associations. Crucially, the presence of semantic content change—whether alone or simultaneously with perceptual context—enhanced temporal-order accuracy, despite eliciting significantly longer reaction times and increased subjective temporal-distance judgments. These findings reveal a distinction between boundary types, indicating that not all boundaries might function identically in temporal memory. Semantic content provides a higher-level schematic scaffold that enhances temporal organization, serving as a cognitive landmark that effectively helps segment the timeline into distinct groups. However, in the absence of semantic content change, perceptual context shifts disrupt temporal continuity, perhaps by interfering with temporal association. Our findings highlight how multiple sources of organization—both perceptual and semantic—modulate how we group, segment, and remember continuous streams of input.

N.A. is supported by the Scientific and Technological Research Council of Turkey (TUBITAK 1001) under the Grant Number 122K922.

53.325 EXPLORING THE MENTAL REPRESENTATION OF VISUALIZATION COMPLEXITY THROUGH MEASUREMENT METHODS

Kylie Lin¹, Maxwell Paik², Cindy Bearfield¹; ¹Georgia Institute of Technology, ²New York University

How do people mentally represent the visual complexity of scenes and visualizations? One possibility is that visualization complexity is perceived holistically, as a gist impression that compresses detailed structure into a global summary (Sun & Firestone, 2021; Sun & Firestone, 2022). Alternatively, visualization complexity may be perceived as more fine-grained, involving feature-based processing that integrates local structure and interactions among visual elements (Attneave, 1957; Rosenholtz, Li, & Nakano, 2007). To examine how visual complexity is perceived in visualizations, we compared two methods for measuring perceived complexity: a numerical rating task and a two-alternative forced-choice (2AFC) comparison. The 2AFC paradigm is typically more discriminative as it reduces individual differences in scale use and can reveal subtler distinctions between stimuli (So, 2022). If perceived complexity reflects only a coarse global impression, we would expect both methods to produce similar results; if it encodes finer visual structure, 2AFC should yield greater

differentiation. We conducted an experiment where 43 participants judged the complexity of a diverse set of data visualizations using one of the two tasks. Using bootstrapped samples, we found a significant difference between the two methods (reduced chi-squared=1.36, $p < 0.05$). This effect was largely driven by image pairs where all participants agreed on which image was more complex in the 2AFC task. Rating-scale data underrepresented these differences in perceived visualization complexity, suggesting that complexity perception preserves some fine-grained information beyond a global gist. These results demonstrate how methodological sensitivity can reveal the granularity of perceptual representations and inform broader understanding of how people process and communicate visual information.

53.326 FROM SUBSYMBOLIC TO SYMBOLIC ENCODING IN CONTOUR INTERPOLATION

Angela Huang¹, Patrick Garrigan², Philip Kellman¹; ¹University of California, Los Angeles, ²Saint Joseph's University

Recent work on contour interpolation in visual completion distinguishes a bottom-up contour-linking process—producing an intermediate representation based on edge and junction geometry—and a subsequent, more probabilistic, constraint satisfaction-based scene description process that operates on this representation to resolve border ownership, closure, and crossing interpolations (Kellman & Fuchser, 2023). Models of the contour-linking process based on neurally plausible filters have been proposed (Heitger et al., 1998; Kalar et al., 2010). However, these models produce pictorial activation maps—discrete points of edge, keypoint, and interpolation activity. Observers viewing these outputs see connected edges, illusory contours, closure, and shape, but the models themselves lack any symbolic tokens corresponding to these entities. In contrast, symbolic descriptions appear necessary for implementing subsequent scene constraints. Here we present a new model of the transition from subsymbolic filter outputs to a symbolic representation of the linked contour representation. We employed the biologically plausible front end of Heitger et al. (1992) to generate initial activation maps and applied simple mathematical operations to extract precise edge locations and orientations, as well as keypoint locations (contour junctions). Interpolation is implemented using abstract curvature elements hypothesized to encode smooth contours (Baker, Garrigan & Kellman, 2021). The model thus embodies a concrete proposal about the transition from subsymbolic to symbolic representations. Empirical tests show improved agreement with human perception. Whereas prior neurally-based completion models used only positive cases—stimuli in which human observers perceive illusory contours—we tested both positive and negative cases. The results indicate that the new model closely matches human judgments of the presence or absence of illusory contours. In contrast, models lacking a transition to symbolic coding generate substantial interpolation activity both for displays in which human observers see illusory contours and those in which they do not.

53.327 FROM GRIDS TO GODS?: PERCEIVING VISUAL STRUCTURE VIA 'SCAFFOLDED ATTENTION' IS RELATED TO SPIRITUAL EXPERIENCES

Dawei Bai¹ (dawei.bai@yale.edu), Joan Danielle K. Ongchoco², Kara Weisman³, Eleanor Schille-Hudson³, Elliott Ihm⁴, Ann Taves⁴, Tanya Marie Luhmann³, Brian Scholl¹; ¹Yale University, ²University of British Columbia, ³Stanford University, ⁴University of California Santa Barbara

The input to visual perception is a continuous wash of light, but what we see are structured scenes of discrete objects. Discrete structure is so fundamental that the visual system may effectively create it in its absence. A particularly striking example occurs in the phenomenon of 'scaffolded attention': when staring at a regular grid of squares (as in graph paper, or floor tiles), many people report spontaneously seeing the squares as grouped into a shifting array of structured patterns (e.g. a block-letter E). (This phenomenon does not occur without the grid, which serves as a 'scaffold' for attention to group subsets of squares.) Might such 'everyday hallucinations' of structure relate to higher-order experiences and beliefs? To find out, we explored its relationship to some of the most profound and life-changing experiences in many people's lives: experiences of supernatural presences (as when seeing a vision of a deity). Observers viewed a uniform grid of squares, and were then asked if they saw the squares as grouped into any shapes or patterns, after which they deliberately tried to see such patterns. The same observers then completed a nuanced measure of supernatural experiences (the Spiritual Events scale). Across a large sample (n=1000), these two forms of experience were related: people who experienced 'supernatural presence' were considerably more likely to spontaneously experience scaffolded attention. And yet, these people were actually *less* able to do so when explicitly instructed — suggesting that the connection may be specific to *spontaneous* experiences (and cannot be due merely to suggestibility). Crucially, these links could not be explained by general degrees of religiosity, or by demographic factors such as education level. These findings suggest a robust link between two types of experiences that may seem maximally distant: perceiving visual structure, and feeling the presence of supernatural beings.

This study was funded by a grant from the Templeton Foundation.

53.328 MULTIPLE FOREGROUND OBJECTS IN COMPLEX SCENES ENHANCE FMRI V1/V2 RESPONSES BEYOND SALIENCE

Shaozhi Nie¹ (nie00043@umn.edu), Daniel Kersten¹, Stephen A. Engel¹; ¹University of Minnesota, Twin Cities

Segmented objects show enhanced neural response relative to background in early visual cortex. However, it is unclear whether such response enhancement simply reflects salience, and whether it is visible in natural scenes and for multiple objects. Using fMRI, we tested the condition under which V1/V2 locations whose population receptive fields (pRFs) overlapped with objects showed higher

response than when they overlapped the background. We analyzed the Natural Scenes Dataset (NSD), which consists of 7T-fMRI data from eight participants, each viewing 10,000 images from the Common Objects in Context dataset (COCO). NSD also includes pRFs of the image pixels driving each cortical location. The COCO images have rich annotations of segmented pixels. We operationalized objecthood using the "thing" annotation, which included most objects, and contrasted it with "stuff" including common backgrounds, such as sky and grass. In addition, we used a state-of-the-art deep network (FOCUS, developed to identify foreground, You et al., 2025) to calculate the salience of COCO segments. For each image, we computed the objecthood and salience at each pixel and projected these onto cortical locations by dot product with pRFs. Average V1/V2 responses were higher for pRFs on an object than background (3.4% signal change for object vs 2.9% change for background, $p < 0.01$ by permutation test). To control saliency, we used FOCUS scores to predict each vertex's response. Residuals were higher of vertices on an object, confirming an effect beyond salience. A similar analysis accounted for low-level features by fitting a V1-like Gabor filter-bank model, and again found effects of objecthood in the residuals. Effects of objecthood were also found when analysing separately the first, second, and third most salient objects. Our results demonstrated that object segmentation increases response in the human visual cortex for multiple objects in complex naturalistic scenes.

53.329 AN AI-HUMAN SIMILARITY DATABASE FOR STUDYING CONTEXT-DEPENDENT REPRESENTATIONAL ALIGNMENT IN MEDICAL IMAGING

Giovanna C. Del Sordo¹ (delosordo@nmsu.edu), Eben W. Daggett¹, Jon Art¹, Michael C. Hout¹; ¹New Mexico State University

Developing artificial intelligence systems for medical image interpretation requires not only high performance but also alignment with human perceptual reasoning. Although AI can provide rapid and scalable diagnostic assistance, systems trained only on categorical labels often form internal representations that do not reflect how humans reason about images, reducing interpretability and introducing potential risks in clinical use. Understanding human perceptual reasoning, however, is not straightforward. Similarity perception is highly flexible: similarity ratings (e.g., between two images) change depending on task context, methodology, and observer expertise. To date, no large-scale resource systematically captures these contextual influences for use in training or evaluating representationally-aligned medical AI. We constructed an open-access database of human and AI similarity judgments (using a triadic comparison task) for a large set of histological medical images. Participants viewed three images and selected the one most dissimilar to the others. This method enables control over context, allowing us to model how framing shapes perceived similarity. Three context conditions were tested: Narrow (all images depict the same pathology), Intermediate (same tissue type but mixed pathology), Broad (different tissues and pathologies). These conditions induce shifts in perceived similarity, providing a structured

tested for studying similarity “flexing.” In parallel, we collected similarity judgments and natural-language rationales from vision-enabled large language models (ChatGPT Vision, Claude Max, Gemini Vision) using the same triadic tasks. The resulting database therefore includes both human- and machine-generated similarity structures, for direct comparison of psychological spaces of where AI and human perception converge/diverge. The database will support researchers in computer vision, cognitive science, and medical AI by providing a “sandbox” for evaluating model robustness to context, benchmarking AI-human alignment, and exploring methods for incorporating human perceptual structure into model training. Ultimately, this resource aims to advance the development of safer, more interpretable, and human-aligned AI systems for medical diagnostics.

This research is supported by the Institute for Applied Practice in AI and Machine Learning at New Mexico State University

53.330 VISUALIZING ILLUSORY CONTOURS PERCEIVED BY DEEP NEURAL NETWORKS TRAINED WITH OBJECT RECOGNITION AND NEURAL RESPONSE PREDICTION

Yuki Kobayashi^{1,2,3} (y-kobayashi@nict.go.jp), Satoshi Nishida^{1,3,4,5}; ¹National Institute of Information and Communications Technology, ²Ritsumeikan University, ³The University of Osaka, ⁴Hokkaido University, ⁵Ministry of Internal Affairs and Communications

Deep neural networks (DNNs) are considered powerful tools for modeling the human visual system. This raises the question: do they share the visual experiences characteristic of human perception? In this study, we aimed to reconstruct the visual percepts from the internal representations of a DNN when processing illusory contours, a phenomenon where humans perceive a subjective shape without physical edges. Our proposed method optimizes an input image to match the DNN's internal representations, while constraining pixel values in the inducer regions. Applying this method to DNNs trained on object recognition, we successfully visualized illusory shapes from early layers of multiple convolutional neural networks. This finding suggests that these models perceive illusory contours in a similar manner to humans, addressing the inconsistent results reported in previous studies. Importantly, these shapes were not visualized with either untrained DNNs or control images, suggesting that perception of illusory contours naturally arises through exposure to the natural world. We further applied our method to DNNs trained to predict neural responses in the human early visual cortex. These models also produced blurry but discernible surfaces, suggesting that current neural predictive models capture key features of the biological brain required to perceive illusory contours. Taken together, our results indicate that DNN architectures can yield plausible models of human vision through training on either object recognition or neural response prediction. The present study provides new evidence of DNNs' similarity to human vision by visualizing their illusory percepts, thereby enabling an intuitive interpretation of their internal processes.

Supported by the Japan Society for the Promotion of Science (grant number: 25K21532).

Scene Perception: Ensemble

TUESDAY, MAY 19, 8:30 AM – 12:30 PM,

BANYAN BREEZEWAY

53.331 STABLE CORRELATIONS BIAS ENSEMBLE PERCEPTION

Richard Bailey¹, Jefferson Ortega¹, Andrey Chetverikov², David Whitney¹; ¹University of California, Berkeley, ²University of Bergen

Ensemble perception allows observers to rapidly extract summary statistics from complex visual scenes. However, it remains unclear whether local correlations within an ensemble bias observers' ensemble judgments. Utilizing spatially weighted average models (Pascucci et al., 2021), we investigated whether observers' perceptual judgments in an orientation ensemble judgement task are susceptible to biases towards correlated information in the visual environment across three experiments. In each experiment, participants reported the mean orientation of 36 lines. In Uncorrelated trials, orientations in the ensemble were drawn from the same underlying distribution with a common mean. However, in the Correlated trials, one half of the ensemble display (e.g., left or right) contained orientations sampled from a distribution with a smaller standard deviation than the other side, leading to orientations on one half of the display to have lower variance (i.e., highly correlated orientations). In Experiments 1 and 2, the location of the correlated signal was fixed (always Left or Right for Experiment 1 and 2, respectively). We found that observers had significant spatial biases, consistently over-weighting the side of the display containing the correlated, low-variance information. This occurred regardless of laterality, ruling out inherent visual field anisotropies. In Experiment 3, we investigated whether observers could dynamically shift their spatial weights on a trial-by-trial basis by including both left and right side correlated trials, in addition to uncorrelated trials, interspersed across all trials. Interestingly, we found that participants' spatial weights did not significantly shift towards the correlated side of the display when the correlated signal switched randomly from trial to trial. These results suggest that the visual system is biased towards correlated signals in the environment when making ensemble judgements, but this bias is not dynamic. Instead, observers might update their spatial weights across time through statistical learning, accumulating evidence of environmental regularities over time.

53.332 SPATIAL ENSEMBLE PERCEPTION IN DERMATOLOGICAL JUDGMENTS

Ning-Ting (Michelyne) Hsu¹, Tiffany Tai¹, Haley G Frey², Katrina A Wolters², Shao-Min (Sean) Hung³, Po-Jang (Brown) Hsieh¹, David Whitney²; ¹Department of Psychology, National Taiwan University,

Dermatologists demonstrate remarkable efficiency when examining large areas of patients' skin, such as the arms or back, spending only brief glances at individual spots. This efficiency suggests they may use ensemble perception, a cognitive process in which clinicians extract overall statistical patterns from groups of lesions rather than examining each one individually. Understanding whether this mechanism operates in dermatology is crucial, as it would reveal how clinicians achieve rapid, wide-field anomaly detection under limited time. Currently, no direct evidence confirms that ensemble perception applies to skin lesion detection, representing a significant gap in our understanding. If dermatologists rely on ensemble processing, it will indicate that successful diagnosis depends not just on individual lesion features but on how a lesion compares to the statistical structure of the whole visual field. However, this introduces specific biases—malignant lesions may appear more benign when surrounded by benign ones, and higher variability may reduce sensitivity to malignancy. Contrarily, if ensemble perception is weak, clinicians would rely on serial search, potentially reshaping how diagnostic errors are understood. The research test whether observers can perceive ensemble characteristics of skin lesion malignancy, bridging vision science with real-world cancer detection. We recruited participants without dermatological training. Seven experimental conditions included both benign and malignant lesion images from the ISIC Challenge Datasets 2019, with malignant images ranging from zero to six. Each trial randomly assigned set sizes of one, two, four, or six images. After exposure, participants judged average malignancy and confidence using a continuous rating slider. Results showed a strong positive correlation ($r = 0.64$, $p < 0.001$) between participant responses ($N = 6$) and average actual malignancy ratings, demonstrating that observers can use ensemble coding to perceive malignancy during rapid screening, opening important questions about training this mechanism and diminishing its inherent biases.

supported in part by NIH R01CA236793

53.333 REVISITING THE DISSOCIATION OF EXPLICIT AND IMPLICIT ENSEMBLE REPRESENTATIONS

Daun Jung¹ (daunjung@yonsei.ac.kr), Junghyun Park¹, Sang Chul Chong¹; ¹Yonsei University

Ensemble perception enables efficient extraction of statistical structure. Prior work has suggested that explicit and implicit ensemble tasks rely on distinct representations (Hansmann-Roth et al., 2021), but this claim was based on paradigms that differed in both stimuli and computational demands. Using identical stimulus arrays and a large sample, we tested whether the two tasks relied on a shared representation. Ninety-six participants completed the explicit and implicit tasks in separate sessions spaced 1–3 days apart, with session order counterbalanced across participants. Each pair consisted of a Priming and a Following streak (5–7 trials each).

Priming distractors were sampled from either a broad Gaussian or a uniform distribution and Following distractors from a narrower Gaussian with nine levels of mean shift (-80° to 80°). In the implicit task, we replicated the characteristic RT–distribution–distance function reported by Chetverikov et al. (2016), confirming robust sensitivity to distractor statistics. More importantly, in the explicit task the accuracy of judging ‘whether a change occurred’ showed an intersecting pattern as a function of $|\Delta\mu|$ depending on the form of the priming distribution (Uniform vs. Gaussian). This hit-rate curve closely resembled a theoretical distribution-overlap distance, suggesting that explicit judgments are sensitive not only to mean differences but also nonlinear similarity arising from the overall distribution shape. We also computed within-participant correlations between explicit and implicit reaction times. After Fisher-z transformation, the mean correlation was significantly negative, revealing weak but reliable covariation, consistent with partially shared noise sources. These results challenge claims of dissociating explicit and implicit ensemble representations. Explicit judgments also incorporate information from the full feature distribution. Overall, the findings indicate that both tasks rely on overlapping probabilistic representations, suggesting that explicit and implicit tasks are two different ways of accessing the same ensemble representation.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT)(RS-2022-NR070542).

53.334 BEYOND FACES: THE EFFECT OF CONTEXTS ON ENSEMBLE EMOTION PERCEPTION

Woojeong Lee¹, Hyeri Yoh¹, Sang Chul Chong¹; ¹Yonsei University

Judging the overall emotion of a crowd is essential for effective social interaction. While previous studies showed that people rapidly judge ensemble emotion based on facial expressions, real-world contexts such as background information and the observer's own emotional state have not been explored. To understand how each observer integrates these sources of information, this study investigated how facial expressions, background emotion, and observer mood influence ensemble emotion perception using naturalistic photographs. In Experiment 1, 10 participants judged the overall emotion of scenes containing crowd figures superimposed on positive, negative, or neutral backgrounds. Facial emotions were evaluated using DeepFace, and background emotions were rated in a pilot survey. Bayesian generalized linear mixed-effects models revealed robust main effects of both facial emotion and background emotion on ensemble judgments. Positive and negative backgrounds shifted ratings in their respective directions relative to neutral backgrounds. Model comparisons strongly favored the full model including interactions. However, interaction terms showed little evidence, suggesting that background emotion primarily influences the overall bias in ensemble judgments. Experiment 2 ($N = 43$) examined whether the observer's own emotional state further influences these judgments. Participants completed an autobiographical recall task to induce either

a positive or a negative mood before judging scenes with positive or negative backgrounds. Results replicated the strong main effects of facial and background emotion from Experiment 1. While the observer's mood showed no main effect, model comparisons revealed a notable interaction with background emotion, suggesting internal emotional states may modulate ensemble perception. Together, these findings demonstrate that emotion ensemble perception integrates multiple sources of information beyond facial expressions. Specifically, contexts substantially influence how we perceive the collective mood in social scenes, underscoring the importance of studying ensemble perception in naturalistic settings that reflect the complexity of our environments.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT)(RS-2022-NR070542).

53.335 SUBSAMPLING STRATEGY IN ENSEMBLE CODING REVISITED

Long Ni¹, Michael S. Landy¹; ¹New York University

When confronted with an influx of sensory inputs, our visual system often relies on ensemble coding to rapidly extract summary statistics from a spatial array of stimuli. A prevailing view in the field is that observers use only a small, selective subset of the items—roughly the square root of the ensemble size—to compute these summary estimates (Whitney & Yamanashi Leib, 2018). Proponents for this subsampling strategy argue for its “efficiency”, as integrating information from a small subset would presumably require less resource expenditure, and for its “efficacy”, as observer models that employ such a strategy can quantitatively predict human performance of ensemble coding across varying set sizes. Here, we revisit both aspects of the subsampling strategy. The prediction that integrating a few items is sufficient to capture human data is based on a problematic assumption. It assumes that sensory encoding noise of each ensemble item remains constant across set sizes. A corollary of this assumption is that the total resource expenditure—measured in Fisher information—scales linearly with set size in ensemble coding. Under a more realistic assumption that the total coding resources are limited, such that the encoding precision for each item decreases as set size increases, we demonstrate that an observer model that computes summary statistics from the entire stimulus ensemble can achieve equal decision accuracy with a consistently lower total resource budget (i.e., lower Fisher information) for any set size $N > 1$. We further show that this resource-constrained full-integration model can better account for data from Baek & Chong (2020) in which subjects discriminated the average item size in ensembles with varying set sizes. Our analysis challenges the “efficiency” argument for this subsampling strategy and opens the door to further evaluating its “efficacy” against the resource-constrained model in explaining existing data.

NIH EY08266, NYUAD Center for Brain and Health, funded by Tamkeen under NYU Abu Dhabi Research Institute award CG012

Scene Perception: Virtual reality

**TUESDAY, MAY 19, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY**

53.336 PREDICTIVE SCENE MEMORY ENHANCES PERIPHERAL COLOR AWARENESS DURING ACTIVE VISION

Ben Chamberlain Zivsak¹, Anna Mynick², Caroline Robertson³, Michael Cohen⁴; ¹Dartmouth College, ²Dartmouth College, ³Dartmouth College, ⁴Amherst College

During active real-world viewing, observers frequently fail to notice when color is removed from most of their visual periphery, revealing profound limits on perceptual awareness (Cohen et al., 2020). Yet, our perceptual experience feels richly saturated across our visual field. What explains this contradiction? Here we ask whether memory for familiar environments supplies the missing color that perception itself fails to deliver (Cohen et al., 2024), building on recent work showing that scene memory selectively speeds judgments for expected adjacent views during planned head movements (Mynick et al., 2025). Participants ($N=12$) studied five immersive real-world scenes drawn in head-mounted Virtual Reality (VR). In a subsequent task, they viewed either a 45° scene prime of a studied scene that was spatially adjacent to an upcoming target view (valid prime condition, $N=6$) or a grey neutral prime with no relevant visual information (neutral prime condition, $N=6$), before executing a head-turn to a corresponding target scene viewpoint. Critically, on the final trial, the target image was completely desaturated except for a central 25 DVA circular aperture. Participants were immediately asked a series of questions to assess awareness of the absence of peripheral color. Participants who received familiar scene primes were significantly more likely to report noticing the loss of peripheral color (33% of trials) than participants in the neutral prime group (0%), despite identical visual input on the target trials. This suggests that visuospatial memory for familiar environments may support the subjective impression of a richly colored visual world during active scene viewing. Taken together, these results suggest that our visual awareness in naturalistic visual settings is actively constructed from predictive memory representations of the surrounding environment (Mynick et al., 2025), shaping both what we expect to see and what we become aware of when those expectations fail.

53.337 CHANGES IN HEAD AND EYE MOVEMENTS DURING NAVIGATION WITH PERIPHERAL VISUAL FIELD LOSS

Sam Beech¹ (samuel.beech@utah.edu), Maggie McCracken¹, Bobby Bodenheimer², Sarah Creem-Regehr¹; ¹University of Utah, ²Vanderbilt University

Simulated and glaucoma-related peripheral visual field loss (PVFL) impairs navigation. Effective navigators optimally integrate visual cues with body-based self-motion cues according to their relative reliability. PVFL may reduce the spatial reliability of vision, but compensatory visual search behaviors could preserve reliability. To investigate this, we simulated gaze-contingent PVFL during a virtual reality navigation task while recording head and eye movements. We present preliminary data for 15 healthy adults across three visual field groups: 110°, 30°, or 20°. Participants walked to three markers with landmarks in the background. The markers were removed, and participants walked to their estimated position of the first marker (homing task) in three conditions: Full Cues – no manipulation; Self-Motion Only – landmarks removed; or Vision Only – participants rotated to disrupt directional self-motion cues. We recorded error and variability. Across groups, error and variability were greater in single-cue conditions: Full Cues (110° = 59 (40) cm, 30° = 55 (37) cm, 20° = 56 (31) cm); Vision Only (110° = 76 (46) cm, 30° = 54 (37) cm, 20° = 77 (53) cm); and Self-Motion Only (110° = 92 (49) cm, 30° = 89 (48) cm, 20° = 94 (52) cm), $d \geq 0.411/0.459$. In the Vision Only condition, average absolute yaw-plane head-trunk deviation was greater with 30° (25.0°) compared to 110° (16.0°) and 20° (18.5°) vision, $d \geq 0.39$. Additionally, head and eye pitch were lower with 20° (-19.4°/-37.5°) and 30° (-14.5°/-34.2°) compared to 110° (-7.9°/-22.0°) vision, $d \geq 0.66/1.00$. The lack of an effect of PVFL on navigation accuracy and variability could be explained by compensatory visual search. The 30° and 20° groups showed increased fixation on landmark-ground intersections, and the 30° group exhibited increased side-to-side head movements between landmarks. These results show preliminary support for distinct visual search behaviors during navigation with PVFL.

53.338 VIRTUAL REALITY ASSESSMENT OF FUNCTIONAL VISION USING DYNAMIC, TASK-ORIENTED SCENARIOS

Brian Bartley¹, Rayyan W Saidi², Michelle Harter¹, Dana Aravich¹, Galen Holland¹, Rakié Cham¹, Alessandro Fascetti¹; ¹University of Pittsburgh, ²Fox Chapel Area High School, Pittsburgh, PA

Advances in vision restoration technologies such as visual prostheses, gene therapies, and whole-eye transplants have outpaced our ability to rigorously measure their impact on functional vision - specifically how people perform real-world tasks like mobility, instrumental activities of daily living (IADLs), and social interaction. Existing tools (patient reported outcomes, observational ratings, and performance-based tests) each address parts of the problem but suffer important limitations, including lack of ecological validity and standardization, as well as limited suitability for ultra-low vision populations. To this end, we propose a virtual reality, task-oriented assessment, VISTA, to directly address these gaps by providing objective, repeatable, and

modular performance measures embedded in high-fidelity parametrically controlled environments that can be deployed across sites. The initial implementation focused on constructing and evaluating a kitchen scene, the first of five planned interactive environments. Task flows were derived from validated low-vision assessments and the HOVER consensus framework, incorporating functional subtasks such as simulated meal prep, detecting room lighting, visual scanning, and navigating within a furnished space. Scene layouts were parameterized to allow controlled variations in object placement across trials, enabling assessment of visual performance rather than memorization. The platform extracts outcome measures, including task completion, absolute and relative speed, accuracy (e.g., misidentification errors), and biomechanical metrics such as head movement, step count, and gaze patterns. Preliminary results demonstrate that VISTA can deliver realistic environmental conditions while maintaining experimental control, producing outcome measures with precision generally unachievable in physical tests. Early feasibility testing confirmed successful execution of all task flows, with expected gradations in task completion, speed, and efficiency for visually demanding subtasks such as structured searches, obstacle avoidance, and detection of lights or motion. These findings support VISTA as a promising tool for standardized, objective, and ecologically valid assessment of functional vision and motivate continued development for future clinical studies.

This research was, in part, funded by the Advanced Research Projects Agency for Health (ARPA-H). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the United States Government.

53.339 DISCRIMINABILITY VARIES ACROSS THEMES OF VISUAL SCENES IN A 2AFC TASK ON THEME-MATCHED REAL-VS-AI IMAGE PAIRS

Alan L. F. Lee^{1,2} (alanlee@ln.edu.hk), Haonan Kong¹, Elaine Jiaxi Fan¹, Kevin Kaiwen Guo³, Siu Lung Jacky Tse¹; ¹Department of Psychology, Lingnan University, Hong Kong, China, ²Cognitive Science Research Centre, Lingnan University, Hong Kong, China, ³School of Engineering, Capital Normal University, Beijing, China

Images generated by artificial intelligence (AI) can be indistinguishable from photographs of real scenes. While many recent studies have focused on face images, few have studied non-face images. This gap raises an important question about how our generic ability to discriminate between real photographs and AI-generated images varies across visual scene types. Here, we created pairs of theme-matched, real-vs-AI images to address this question using a two-alternative, forced-choice (2AFC) task. From online photo databases, we obtained 60 real photographs for 12 scene categories, every 3 of which belong to the following 4 themes: humans, nature, interactables, and man-made structures. We then used commercially-available multimodal large language models to transform each real photograph first into detailed text descriptions and then into an AI-generated

image. These theme-matched image pairs were then evenly divided into 5 batches, with 40 participants (total = 200 from Prolific.com) completing the following tasks for each batch. On each trial, participants viewed the two images side by side simultaneously. They chose the one that was more likely to be AI-generated and then rated their confidence on a continuous scale based on their decision. In general, participants achieved above-chance discriminability, with an average 2AFC d' of ~ 0.70 . They were also metacognitively sensitive about their decisions, with significantly higher confidence for correct than for incorrect responses. Critically, we found that discriminability varied across scene categories: with food being the highest (0.96), followed by images with humans (~ 0.81), and indoor scenes being the lowest (0.50). Furthermore, performance and confidence were dissociated depending on scene category, e.g., despite the high discriminability, confidence for food images was significantly lower than that for images with humans. These findings suggest that the theme of a visual scene can influence our perception and metacognition of realism judgment on images.

This work was partially supported by the Research Matching Grant Scheme from the University Grants Council of HKSAR, China.

53.340 EFFECTS OF VISUAL COMPLEXITY ON WAYFINDING PERFORMANCE IN PEOPLE WITH ULTRA-LOW VISION USING VIRTUAL REALITY

Dinesh Venugopal¹ (dineshoptometrist16@gmail.com), Batuhan Erkat¹, Rokhsana Sadeghi^{2,3}, Chau Tran⁴, Will Gee⁴, Brittnee Livingston⁵, Gislin Dagnelie², Arathy Kartha¹; ¹SUNY College of Optometry, Department of Biological and Vision Sciences, New York, NY, United States, ²Johns Hopkins Medicine Wilmer Eye Institute, Baltimore, MD, United States, ³Biomedical Engineering, Johns Hopkins University, Baltimore, MD, United States, ⁴BaltiVirtual, Baltimore, MD, United States, ⁵Central Association for the Blind and Visually Impaired, Utica, NY, United States

People with ultra-low vision (ULV) perceive limited visual information and experience substantial challenges in complex environments. However, the impact of visual complexity on wayfinding performance in ULV remains largely under-explored. This study aimed to quantify the effects of visual complexity on walking speed and turn deviations during visual wayfinding in individuals with ULV using virtual reality (VR). Participants with normal vision (NV) ($n=10$) and ULV ($n=14$) completed visual wayfinding tasks in a virtual cafeteria under three complexity levels: ordered tables (A), disordered tables (B), and cluttered surroundings (tables, seated avatars, balloons, posters) with a moving avatar (C). At each level, participants were instructed to locate and walk towards the cashier at the opposite side ($\sim 4.3\text{m}$ away) at a comfortable speed. The NV participants also completed wayfinding tasks with ULV-simulating filters (sULV; using Bangerter foils). Head positions were tracked; turns were segmented using the simplified Ramer–Douglas–Peucker algorithm ($\epsilon=10\text{cm}$) and compared to an optimal path to compute turn deviations. Walking speed (m/s) and turn deviations were compared across participant

groups and levels using Generalized Estimating Equations. The ULV group showed the greatest impairments, with walking speed at C decreased by 0.10m/s compared to A and 0.05m/s compared to B (both $p<0.05$), and turn deviations increased by 5.1 and 3.3 turns, respectively (both $p<0.05$). In sULV, walking speed at C decreased by 0.05m/s compared to levels A and B (both $p<0.05$), while turn deviations increased by 2.9 turns compared to A ($p<0.001$) but showed no difference compared to B. The NV group showed no differences in walking speed or turn deviations across levels. In both ULV and sULV groups, walking speed and turn deviations were significantly correlated ($\rho = -0.79$ and -0.67 , respectively; both $p<0.001$). Visual complexity significantly impaired wayfinding performance in individuals with ULV and sULV.

53.341 DOES SPATIAL KNOWLEDGE OF ROAD LAYOUT PLAY A ROLE IN THE VISUAL CONTROL OF STEERING?

Grace Roessling^{1,2}, Brett Fajen¹; ¹Rensselaer Polytechnic Institute, ²Carnegie Mellon University

Previous research has extensively explored the visual information used to control steering during driving, but less is known about the role of spatial knowledge of the road. The present study investigated whether drivers rely on knowledge of road geometry and the nature of such knowledge. Subjects performed a simulated driving task, using a steering wheel and foot pedals to drive along a set of ten winding tracks. Each track comprised straight segments interleaved with curved segments, some that were immersed in fog. Experiment 1 ($N = 30$) compared performance of subjects who drove the same track on all ten trials against subjects who drove a different track on each trial. Analyses focused on performance on the 10th trial, which used the same track in both groups. Subjects in the Constant Track group exhibited significantly greater steering stability and reduced lane position variability, but only when visibility was reduced by fog. They also performed better on post-tests that measured knowledge of road geometry in egocentric and allocentric reference frames. Experiment 2 ($N = 60$) tested whether the spatial knowledge used by drivers is better characterized as a globally cohesive cognitive map or as local landmark-to-road-segment associations. Three groups performed the same task as Experiment 1: a Control group similar to the Constant Track group from Experiment 1, a Scrambled Landmark group in which the track was the same on each trial but landmarks were scrambled, and a Scrambled Segment group in which the order of road segments was scrambled but landmark-road segment pairings were preserved. Steering performance was most affected by the manipulation of landmarks. We conclude that knowledge of road layout does play a role in steering control but only when visibility is reduced. Further, such knowledge is best characterized in terms of local associations between landmarks and road geometry.

NSF 2218220

Scene Perception: Categorization, memory

TUESDAY, MAY 19, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY

53.342 CONTRIBUTIONS OF STATISTICAL REGULARITY AND IMAGE COMPLEXITY TO RAPID AESTHETIC EVALUATION

Kunlin Jin¹ (linkunjin2000@outlook.com), Diane M. Beck^{1,2},
¹Department of Psychology, University of Illinois Urbana-Champaign,
²Beckman Institute, University of Illinois Urbana-Champaign

Observers form aesthetic impressions within a fraction of a second, and processing fluency accounts propose that these immediate evaluations depend on how easily an image is processed (Reber et al., 2004). Two factors known to shape fluency are real world statistical regularity, which includes features such as scene prototypicality and canonical object viewpoints (Beck et al., 2024) and image complexity (Forsythe et al., 2011). Although both factors have been linked to aesthetic preference, their relative contributions during rapid judgments remain unclear. We examined whether statistical regularity predicts attractiveness ratings more strongly than two complementary measures of visual complexity: fractal dimension and edge density. Participants rated the attractiveness of scenes (N = 33) and objects (N = 28) on a 1–5 scale. Scene images were from four categories and were either highly representative of their category (statistically regular) or poor representatives (statistically irregular). Object images were either canonical views of the object (statistically regular) or non-canonical views (statistically irregular). Initial paired t tests showed that statistically regular images received higher attractiveness rating than irregular images for both scenes, $t(32) = 9.44$, $p < .001$, $d = 0.64$, and objects, $t(27) = 7.42$, $p < .001$, $d = 1.24$. We then modeled all images with a linear mixed effects approach including statistical regularity and the visual complexity factors as fixed effects. Statistical regularity emerged as the strongest predictor of attractiveness, with good exemplars rated higher than bad exemplars. Beyond this dominant effect, images that were more fractally complex received higher ratings, whereas images with greater edge density received lower ratings. These findings demonstrate that rapid aesthetic evaluations of our real-world scenes and objects were shaped primarily by statistical regularity, with image complexity exerting additional but weaker influences.

53.343 REAL-WORLD STATISTICAL REGULARITY SPEEDS UNCONSCIOUS BUT NOT CONSCIOUS VISUAL PROCESSING: EVIDENCE FROM BCFS AND REV-BCFS

Yiwen Wang¹, Ling Lee Chong¹, Diane M. Beck¹; ¹University of Illinois at Urbana-Champaign

Real-world statistical regularities, learned through a lifetime of exposure to structured environments, allow the visual system to form predictions about incoming stimuli. Under predictive coding models, statistically regular images generate smaller prediction errors and should therefore be processed more efficiently than statistically irregular images. One important form of such regularity is scene representativeness, where highly representative “good” exemplars are detected more efficiently than “bad” exemplars (Caddigan et al., 2017). The present study tested whether the influence of representativeness arises during unconscious or conscious stages of visual processing. We combined the standard Breaking Continuous Flash Suppression (bCFS) task with its reverse variant (rev-bCFS; Ciorli, Pia, & Stein, 2025). In bCFS, the image is initially suppressed, and the time it takes to break into awareness reflects early, unconscious processing efficiency. In contrast, rev-bCFS presents the image fully visible at trial onset and then introduces suppression, measuring how long a consciously perceived image resists suppression, a marker of conscious-level properties. Our results showed that good exemplars broke suppression faster than bad exemplars in bCFS, demonstrating a representativeness advantage during initial, unconscious processing. However, in rev-bCFS, good and bad exemplars exhibited comparable suppression-resistance times, showing no effect once images were consciously available. These findings align with the predictive coding framework, indicating that good exemplars gain an early processing advantage because they produce smaller prediction errors, but this benefit does not extend into conscious perception, where prediction errors no longer drive performance.

53.344 LOW SPATIAL FREQUENCY INFORMATION MORE EFFECTIVELY ACTIVATES SCENE PREDICTIONS THAN HIGH SPATIAL FREQUENCY INFORMATION

Ling Lee Chong¹ (llchong2@illinois.edu), Yiwen Wang¹, Kara D. Federmeier¹, Diane M. Beck¹; ¹University of Illinois at Urbana-Champaign

Our prior work showed that the N300 amplitude is smaller in response to good scene exemplars than bad scene exemplars, reflecting a better match between good scene exemplars and our internal predictions. The frame-and-fill hypothesis proposes that coarse, low spatial frequency (LSF) information first generates initial predictions of an object’s identity, which are then refined by detailed high spatial frequency (HSF) information to improve object recognition. If the prediction described by the frame-and-fill model is the same one indexed by the N300, then LSF cues should be more effective than HSF cues in activating these predictions. To test this, we used a cueing paradigm in which participants viewed either LSF- or HSF-filtered scenes as cues that either matched or mismatched the subsequent full-spectrum target scene, which was either a good or bad scene exemplar. Participants indicated whether the filtered scene and full-spectrum scene belonged to the same category while EEG was being recorded. As in Kumar et al. (2021), we hypothesized that if the cue was effective, we would see a diminished difference between good

and bad exemplars on mismatch trials relative to match trials. Our results showed that only the LSF cues were effective, as the LSF cue trials showed a significant modulation of match versus mismatch, with a larger N300 good/bad difference when the categories matched than when they did not. However, this effect was not present when HSF scenes were used as cues, as the N300 good/bad difference was similar across match and mismatch conditions. These findings suggest that LSF scenes were more effective at activating scene predictions than HSF scenes in keeping with the frame-and-fill hypothesis.

53.345 ARE SCENE REPRESENTATIONS ORGANIZED IN PART BY OBJECT FUNCTION?

Thomas Sanocki¹ (sanocki@usf.edu), Teresa P. Pham²; ¹University of South Florida, ²University of Minnesota

Cognitive theories assume that perceiving and acting within a familiar scene is supported by top-down influences driven by internal knowledge about the scene. The knowledge may include typical objects and actions — information about object function. Here, we test whether viewing a scene can prime knowledge about object function. The primes were an image made from related, same-scene pieces, or from unrelated pieces of different scenes. That is, same-scene primes consisted of 3 different pieces of the same room (3 phrases of a living room, in terms of scene-grammar theory); different-scene primes consisted of 3 pieces that came from different scenes. A prime was presented for one second, followed by a copy of the image with one piece and one object within the piece highlighted. Observers indicated whether the object in this region (piece) was positioned appropriately to function. For example, a lamp functions if upright on a table top (YES response), but not if oriented sideways (NO response). (The stimuli were validated in a pilot study, to encourage rapid and accurate function-judgments. All objects were appropriate for the scene-piece they were depicted in, so that scene-belonging was not relevant to the response. Orientation was changed for some NO objects, and position within their scene was varied for others. 31 students participated.) Function judgments were overall accurate (97%) and rapid (Mean RT = 1171 ms), with RT improving across the session. The critical comparison was the mean RT after same-scene and different-scene primes. RT was not different between the two prime conditions, indicating a null effect of scene-organization. This similarity in RT held across the session. The results suggest that function perception can't be primed by scene meaning; function perception is independent of larger scene-organization.

53.346 AD HOC CATEGORIZATION OF HAZARDS IS NO SLOWER THAN BASIC LEVEL CATEGORIZATION

Gillian Rosenberg¹, Caitlyn McGlashan², Benjamin Wolfe³, Michelle Greene¹; ¹Barnard College, ²Bates College, ³University of Toronto Mississauga

Safe driving requires constant assessment of potential hazards. Observers can correctly report the presence of a hazardous road condition in ~400-600 ms (Wolfe et al., 2020), but reaction times

combine perceptual processing, decision time, and motor response. While object and scene categorization can be assessed from M/EEG in 150- 250 ms, these studies use well-defined basic-level categories. However, hazards are not a fixed category. For example, a moose is only hazardous when it is located too close to a road. Thus, hazards can be understood as an ad hoc category where category members depend on context (Barsalou, 1983). Is ad hoc hazard categorization slower than basic-level categorization? We presented participants (N=16) with 8-second dash cam videos that either showed a hazard or no hazard at some point in the visual epoch. We used 64-channel EEG to measure brain activity while observers viewed each video. We then performed cross-validated logistical regression analyses at each time point (from 1000 ms before to 1000 ms after the hazard) to determine how well the EEG data predicted the presence of a hazard using a log likelihood metric. To control for the possibility that videos with hazards have more motion, we calculated motion energy for each video and fit a full model using both motion energy and EEG features, as well as a model with only the motion energy. We subtracted the R2 of the motion energy model from the R2 of the full model to obtain the prediction from EEG alone. We found that significant hazard decoding was possible starting at 142 ms after hazard onset. These findings indicate that hazard categorization has the same lower bound as basic-level categorization, showing that the ad hoc and contextual nature of the hazard category does not strongly impact the time course of category processing.

NSF CAREER 2240815

53.347 THE BALANCE OF SEMANTIC AND SPATIAL ASPECTS OF SCENE UNDERSTANDING IN OLDER ADULTS.

Elissa Aminoff¹, Bernadine Gurning; ¹Fordham University

Visual perception undergoes dynamic changes throughout the lifespan, in part due to changes in brain development. As individuals age, the category specific organization of the ventral visual stream becomes less specific – a process known as dedifferentiation. This study examined how this might affect older adults' scene perception. We predict that as one ages, the focus of scene perception may shift more towards the semantic and content aspects of a scene – processing that typically occurs in more anterior areas of the ventral stream. In contrast, younger adults may process more spatial aspects of the scene, which may rely more on category specific scene processing, and more posterior areas of the ventral stream. To test this, both older and younger participants viewed scenes and described them as accurately and descriptively as possible. Their responses were then coded and scored along three dimensions: level of detail, semantic information, and spatial information. Each response was scored by two reviewers whose ratings showed strong inter-rater reliability (ICC = 0.8-0.9). Results demonstrated that, critically, the level of detail matched across both age groups, and thus, a similar level of specificity in scene perception was observed. However, a significant interaction was found between groups and the spatial and

semantic information included in the responses ($p < .03$). Older adults focused more on semantic information. In comparison, younger adults focused more on spatial information. We suggest that this pattern of results indicates not a decline in scene perception as one ages, as demonstrated by the equal amount of detail, but rather a shift in strategy that may be triggered by changes in the ventral stream's dedifferentiation.

53.348 AUTOMATICALLY PERCEIVING PATHS THROUGH A SCENE

Giacomo Aldegheri¹, Başak Güvensin², Roland W. Fleming¹; ¹Justus Liebig University Giessen, ²Bilkent University

As we look at scenes in the real world, we can perceive not just “what is out there” but also “what is possible”. In an apparently automatic and effortless way, we can extract information about actions that are possible in a scene, locations that can be reached, or events that can happen. For example, we can quickly perceive the possible paths to reach a door, or whether an object at the edge of a table is likely to fall. It is still unknown, however, how we automatically determine the paths leading to these possible states, taking environment dynamics and constraints into account. Here, we use simple maze-like scenes as a case study, using a Go/No-Go task to probe which states are perceived as “leading up to” a given goal state. We instruct participants to press a button whenever they see a scene in which an agent has reached the end of the maze and is next to a target. Importantly, we do not describe the scene in these terms, but simply show them the image they have to respond to. We find that they make a substantially larger number of false alarms on scenes in which the agent is closer to the goal along the path delimited by the maze. Other scenes, in which the agent has the same Euclidean distance to the goal but would have to cross a wall to reach it, do not lead to false alarms. These results provide preliminary evidence that scene constraints (such as the walls of a maze) determine the perceived similarity of different scene states. We plan on expanding this research to more complex scenes, and other domains beyond maze navigation, in order to elucidate the mechanisms by which we determine the paths leading up to a possible state.

This research was funded by the European Research Council (ERC) Advanced Award ‘STUFF’ (ERC-ADG-2022- 101098225)

53.349 A LEXICON OF PERCEIVED VISUAL COMPLEXITY

Shuning Jiang¹, Meng Ling², Rober S. Laramée³, Michael Sedlmair⁴, Jian Chen¹; ¹The Ohio State University, USA, ²The Washinton Post, USA, ³University of Nottingham, UK, ⁴University of Stuttgart, Germany

Words people use have long served as a window into perception. Despite decades of research on visual complexity (Oliva et al. 2004; Rosenholtz et al. 2005; Chu et al. 2025), most quantitative metrics are derived from natural scenes rather than data visualizations. To fill in the gap, we focus on the degree to which specific features, attributes, and factors are more or less relevant. We discuss lay terms, provide definitions, and suggest their potential effects on observers, by

conducting a qualitative analysis of ~700 human verbal descriptions of visualization image pairs to identify the perceptual dimensions that observers spontaneously use when describing complexity. From these descriptions, supported by human expert coding and LLM-assisted analysis (GPT-5mini), we developed and defined a lexicon of six dimensions: (1) Data Density and Clutter, (2) Visual Encoding Clarity, (3) Color, Symbol, and Texture Detail, (4) Schema, (5) Semantics and Text Legibility, and (6) Aesthetics. We mapped these dimensions to expressions of Immediacy and Cognitive Load (e.g., “unclear where to look,” “takes longer to interpret”) and linked them to memorability outcomes, following that of Kyle-Davidson et al. 2024. Using this lexicon, we characterized the complexity of ~700 visualization images and provided the annotated dataset. In contrast to findings from natural scenes we observed no strong relationship between complexity and memorability in visualization images (r -squared=0.07). Our results establish a principled vocabulary for perceptual drivers of visualization complexity and offer a foundation for developing objective, computational measures tailored to data visualization images.

Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2120/1 – 390831618; DFG, German Research Foundation) – Project-ID 251654672; EPSRC UKRI157, APP17227

Visual Working Memory: Models, neural

**TUESDAY, MAY 19, 8:30 AM – 12:30 PM,
BANYAN BREEZEWAY**

53.350 HIERARCHICAL REPRESENTATIONAL TRANSFORMATIONS OF WORKING MEMORY IN HUMANS AND MACHINES

Qingqing Yang¹ (yang.6118@osu.edu), Hyewon Willow Han², Bogeng Song³, Julie Golomb¹, Dobromir Rahnev³, Yalda Mohsenzadeh², Hsin-Hung Li¹; ¹The Ohio State University, ²Western University, ³Georgia Institute of Technology

Working memory (WM) allows humans to maintain past information even as new inputs arrive, yet how WM representations transform between encoding and retrieval remains unclear. Moreover, whether and how the format of WM representations, and their transformations across phases, are aligned between the human brains and artificial agents remains largely unexplored. To address these questions, we analyzed WM representations in the public human 7T fMRI Natural Scenes Dataset and compared them to the representations in recurrent neural networks (RNNs) trained on the same naturalistic 1-back task. Using representational similarity, cross-decoding, and subspace geometry analyses, we compared rotational and non-rotational transformations between WM encoding and retrieval phases in different brain regions and model layers. Our analyses revealed convergent evidence for a gradient of WM coding schemes in both the

human brain and neural networks. In humans, early visual regions (V1–hV4) underwent large representational changes from encoding to retrieval, including both rotational and non-rotational transformations. In contrast, representations were more stable in higher-order regions in the prefrontal cortex (FEF, dlPFC). Applying the same analyses to models showed a similar hierarchical pattern, but it critically depended on the learning objective and the recurrent architecture. We examined two different encoder architectures (ResNet and Vision Transformer ViT) with two learning objectives (supervised and self-supervised learning), followed by three recurrent architectures (vanilla RNN, GRU, and LSTM). Supervised encoders exhibited an increasing stability along network layers paralleling the cortical gradient in terms of both rotational and non-rotational transformations, while self-supervised encoders lacked such a hierarchy. Among recurrent architectures, gated architectures (GRU, LSTM) better reproduced the brain-like hierarchical patterns of WM transformations. Together, these results revealed hierarchical shifts between flexibility and stability in WM representational transformation in both humans and machines, with supervised learning objectives combined with gated recurrent dynamics most closely resembling human WM mechanisms.

53.351 PRESERVED FEATURE GEOMETRY OF PERCEPTION AND WORKING MEMORY

Waragon Phusuwan^{1,2} (waragonphoosuwat@gmail.com), Xue-Xin Wei³, Chaipat Chunharas^{1,4,5}; ¹Cognitive Clinical and Computational Neuroscience Center of Excellence, Chulalongkorn University, Bangkok, Thailand, ²Medical Sciences, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand, ³Department of Neuroscience, Department of Psychology, Center for Perceptual Systems, Center for Learning and Memory, Center for Theoretical and Computational Neuroscience, The University of Texas at Austin, Austin, Texas, U.S.A., ⁴Division of Neurology, Department of Medicine, Faculty of Medicine, King Chulalongkorn Memorial Hospital, Thai Red Cross Society, Bangkok, Thailand, ⁵Chula Neuroscience Center, King Chulalongkorn Memorial Hospital, Thai Red Cross Society, Bangkok, Thailand

Visual working memory is often conceptualized as the persistence of sensory-like representation. Consistent with this view, decoding analysis in early visual cortex reveals cross-generalization between perception and memory, suggesting that their representations may be shared. However, quantifying this alignment directly from the voxel-level activity has been challenging, as noise and task-irrelevant variance can obscure underlying geometry. To address these limitations, we develop an approach that analyzes the geometry of the task-relevant subspaces inferred from the Inverted Encoding Model (IEM). We applied our approach to analyze a published fMRI dataset collected while six participants performed a visual localizer task (perception) and a delayed-orientation recall task (WM). IEM models were fitted to both task conditions to obtain voxel-to-IEM channel loadings as an effective coordinate basis. While previous work on IEM generally focused on the response of individual channels, our

approach instead analyzes the geometrical properties of the subspaces defined by these IEM channels. We evaluate the subspace geometry using Canonical Correlation Analysis (CCA) to identify the principal angles, and Basis Procrustes Analysis to calculate the principal rotation angle between the coordinate axes of perception and WM feature subspaces. In contrast to the raw BOLD signal, where subspaces for perception and working memory appeared orthogonal, the feature-specific latent subspaces revealed significant alignment. Moreover, CCA and Basis Procrustes methods showed stable shared geometry from V1 through IPS0. Interestingly, while alignment of the IEM axes remains stable in early visual cortex, it increases slightly in higher-order parietal regions (IPS1). Our analysis of task-relevant subspaces revealed substantial alignment between perception and working memory representation – an effect that appears to be preserved across visual hierarchies. The alignment of the subspaces may inform computations during tasks that involve both perception and working memory.

This research project is supported by the Second Century Fund (C2F), Chulalongkorn University.

53.352 HOLISTIC REPRESENTATION AS AN ALTERNATIVE FORMAT OF VISUAL WORKING MEMORY

Ziyi Su¹, Brad Wyble¹; ¹Penn State University

Working memory representations are commonly assumed to store information as separate objects. However, recent work emphasizes the flexible nature of working memory representation such that different types of representations might be formed according to distinct task requirements (Lecerf & Ribaupierre, 2010; Gonthier, 2021). One alternative format of representation is a holistic representation: representing the whole display as a single object without decomposing it into distinct objects or features. This type of representation could be strategic in tasks that favor retrieving or manipulating the entire display. Change detection, for example, can be conceivably accomplished by comparing a holistic representation of the remembered array against the test display. Studies of memorizing complex scenes reported a similar strategy of encoding “the gist” as one part of the memory (e.g., Rensink, 2000). However, holistic representations have not been widely recognized as a format readily applied for other visual working memory tasks. Therefore, this study probes evidence for holistic representation with colored square arrays. Participants encoded 6 colored squares for each trial and completed a change detection test in which 1, 3 or 6 of the squares would be probed, unpredictably. The squares were initially presented simultaneously during the encoding phase for one group, which encouraged holistic representation, and were presented sequentially for the other group, which encouraged item-based encoding. We predicted better performance for probe 6 exclusively in the simultaneous condition because a holistic memory could be used in its entirety while the probe 3 condition required selecting a subset of the squares. Results confirmed this: Simultaneous presentation elicited higher performance when probing 6 rather than 3 items; but for

sequential presentation, subjects had similar performance for probing 3 and 6 squares. These results suggest strategic use of holistic representations in certain task configurations, and support an alternative to object-based storage.

53.353 NO TRUE REPRESENTATIONS: STORING OBJECTS AS EITHER ITEMS OR HOLISTIC REPRESENTATIONS

Ian Deal¹ (imd5205@psu.edu), Ziyi Su¹, Brad Wyble¹; ¹The Pennsylvania State University

Visual working memory (VWM) is the process that stores temporary visual information needed by other cognitive processes. There is some debate as to the true nature of representations in VWM, but it is typically described as maintaining discrete, item representations. Through current advances in deep-learning architectures and training, we model the interaction between the visual system and VWM using a generative deep-learning neural network: Model for Latent Representation (MLR). MLR consists of four main components: a spatial transformer for attending to objects or groups, a modified variational autoencoder for compressing visual representations, a classifier for categorizing colors and objects, and finally a memory binding pool to store information. This architecture allows for simulations to encode, store, and reconstruct visual scenes with multiple discrete elements. This model offers various ways to encode a single scene, including the use of multiple items or alternatively encoding the whole scene as a holistic representation. Importantly, even when using holistic representations, MLR can simulate effects traditionally viewed as supporting item-based storage. For example, when simulating multi object change detection experiments, MLR exhibits a set-size effect even when using a single holistic representation due to the loss of details as the complexity of a stored item is increased. Moreover, holistic encoding replicates the item-based conjunction advantage of (Olson, Jiang 2002). We also explain data that challenge item-based encoding using a traditional change detection task. When subjects are shown 6 items simultaneously, accuracy is higher when probed on the whole display as opposed to half of the items and this impairment disappears when the initial display is shown sequentially. MLR implies that memory representations can be configured for a specific task, suggesting that a true, underlying representation of VWM may not exist.

NSF grant number BCS-2216127

53.354 RAPID ORGANIZATION OF NON-SPATIAL INFORMATION INTO SPATIAL STRUCTURES FOR ADAPTIVE WORKING MEMORY

Sihan Yang¹ (siy009@ucsd.edu), Yueying Dong¹, Anastasia Kiyonaga¹; ¹University of California, San Diego

Successful working memory (WM) must be flexible and constructive, allowing visual inputs to be manipulated and organized to optimize behavior. In long-term memory, this structuring may be supported by

cognitive maps, which represent non-spatial information in an internal spatial framework to exploit navigation circuits. Such a mapping strategy may also be an efficient means of retaining and indexing novel information for WM. But it is unclear how quickly, and under what constraints, such maps are formed. Here, we show that people construct cognitive maps on the fly, organizing novel visual inputs that vary along multiple feature dimensions into an indexing scheme for WM. Participants viewed sequences of stimuli whose features varied along two dimensions – for instance, animal profiles varying in ear length or tail bushiness. The experiment comprised blocks that emphasized different types of judgments. ‘Match’ blocks required recognition judgments about whether the probe display depicted the exact features of stimuli from the preceding sequence. ‘Relative’ blocks asked participants to make relational judgments about how a feature changed from one stimulus to its successor in the sequence. We specified several hypothetical models of strategies for mentally representing and accessing the stimulus space. These included strategies for traversing a 2D cognitive map (where feature combinations are indexed as map coordinates) and a simple ‘sequence’ model (where the stimuli from a given trial are represented in their presentation order). We found that, in ‘match’ blocks, response times (RTs) corresponded best to the simple sequence strategy. However, in ‘relative’ blocks, RTs corresponded to the path length of traversing the hypothetical cognitive map. Likewise, in a parallel eye-tracking study, gaze dynamics during ‘relative’ blocks also tracked the hypothetical map paths. Thus, cognitive maps may be rapidly constructed for visual WM, especially when the task encourages relational encoding among novel pieces of information.

53.355 USING GENERAL RECOGNITION THEORY TO CHARACTERIZE VISUAL AND AUDITORY FEATURE BINDING IN WORKING MEMORY

S. Newaz¹ (shah.newaz-1@ou.edu), M. P. Dungan², M. J. Wenger¹; ¹University of Oklahoma, ²Florida International University

When we temporarily store information about a multidimensional stimulus and keep it in an active state, are the features of that stimulus bound together or are they maintained independently? Most approaches to addressing this question have relied on operational definitions of working memory binding, allowing for the possibility of circular reasoning, rather than relying on theoretically-driven approaches. General recognition theory (GRT), a formal framework that explains multidimensional feature binding using three core constructs, provides a means to address this limitation. The violation of these three constructs—informational independence (II), informational separability (IS), and decisional separability (DS)—can be used to test for varieties of binding in working memory representations. In our previous study, we observed violations of II, inferences regarding which were limited by high marginal accuracy in our sample. To address this limitation, we used a modified n-back task to examine feature binding between color and pitch, with multiple levels determined by assessing each participant’s discrimination

thresholds. This resulted in higher but interpretable error rates with marginal accuracy rates off ceiling and above chance. The analysis of marginal response invariance and the marginal signal detection measures revealed few if any violations of either IS or DS. In addition, analysis of report independence showed limited violations of II. These results indicate that the features of color and tone were generally not bound together in this n-back task and raises questions about the extent to which strong evidence for feature binding can be obtained across vs. within sensory modalities.

OU Undergraduate Research Opportunities Program

53.356 DISTINCT EFFECTS OF RETRO-CUE RELIABILITY ON VISUAL AND MOTOR PRIORITIZATION IN WORKING MEMORY

Sisi Wang^{1,2}, Freek van Ede²; ¹Texas Tech University, ²Vrije Universiteit Amsterdam

Working memory enables us to maintain visual information to guide upcoming behavior, bridging recent sensations to anticipated actions. While much prior work has focused on the storage mechanisms of visual working memory, recent studies highlight the dynamic processes through which stored contents are prepared for action. When visual items held in memory are associated with specific manual responses, visual and motor prioritization jointly emerge after retro-cues that indicate with certainty which memory item will be tested. However, real-world cues often vary in the degree of certainty they provide. It remains unclear whether visual prioritization and the prioritization of associated actions are similarly modulated by retro-cue reliability. To address this question, we used a visual-motor working memory task in which each of two memory items was linked to a different response hand. Critically, retro-cue reliability was manipulated across three levels (100%, 80%, 60%). Using EEG, we measured visual and motor prioritization during the delay period via posterior alpha (8–12 Hz) and central mu/beta (8–30 Hz) lateralization related to the cued memory item's location (visual) or response hand (motor), respectively. Our results reveal dissociable effects of cue reliability on visual and motor prioritization in working memory. Visual prioritization was robust and remained stable across reliability levels. In contrast, motor prioritization profoundly scaled with cue reliability, showing weaker – and slower – motor prioritization following less reliable cues (lower certainty). Direct comparisons further showed distinct temporal profiles whereby visual and motor prioritization developed together following 100% reliable cues, while motor prioritization developed after visual prioritization with 80% reliable cues (and was absent with 60% reliable cues). These findings demonstrate a functional dissociation between visual and motor prioritization in working memory and provide new insight into how certainty shapes the dynamic processes by which working memory contents are prepared for behavior.

This work was supported by an NWO Vidi Grant (14721) and an ERC Starting Grant from the European Research Council (MENTICIPATION, 850636) to F.v.E.

53.357 EEG DECODING REVEALS SENSORY STORAGE FOR BOTH ACTIVE AND PASSIVE STATES IN VISUAL WORKING MEMORY

Xinran Chen¹, Mowei Shen¹, Hui Chen¹, Yingtao Fu¹; ¹Zhejiang University, Hangzhou, China

Previous research has shown that contents of visual working memory (VWM) can be decoded from neural activity in early visual cortex, supporting the sensory recruitment hypothesis. However, it remains debated about whether different representational states of VWM—namely active and passive states—rely on sensory storage. Here, we used electroencephalography (EEG) to demonstrate that memory load of both states could be decoded from multivariate pattern of posterior neural activity. In a VWM task, participants were required to memorize either two or four colors, and the representational states were manipulated by a retrospective cue: cued items were maintained in the active state, whereas uncued items were in passive state. The memory load of both states was independently manipulated as either low or high. We found that activity from posterior channels during the delay period reliably classify high versus low load for both active and passive state. Besides, decoding for passive state showed less robust result compared to the active state. Together, these findings suggest that both active and passive state of VWM recruit sensory storage, although passive representations may be maintained with weaker memory strength.

53.358 NEURAL DISSOCIATIONS BETWEEN CUE-DRIVEN AND METACOGNITION-BASED SELECTION IN WORKING MEMORY

Ying Zhou¹ (yz5725@nyu.edu), Daryl Fougine¹; ¹Department of Psychology, New York University Abu Dhabi

We can flexibly control the contents of working memory. Retro-cue paradigms reveal that explicitly cued information in working memory can be selectively prioritized (Souza & Oberauer, 2016); 'Choose the best' paradigms indicate that participants have metacognitive knowledge of memory quality and can use this knowledge to select the best remembered information in working memory (Fougine et al., 2012). While both selections can enhance behavioral performance of the selected item, it remains unclear whether they rely on the same neural mechanism. We used fMRI to measure the neural activity during metacognition-based and cue-driven selections (8 participants). Specifically, on each trial, participants memorized the orientation of three items at different locations on the screen, and a cue during the memory delay would prompt a specific type of selection: either a spatial retro-cue indicating the location of the to-be-tested item or a self-report cue indicating that they would report the best-remembered item. The visual input across the two selection conditions is comparable, and the only difference is whether participants selected

an item based on a retro-cue or on their metacognitive knowledge of memory quality. We found different brain regions are recruited by these two types of selection. Cue-driven selection, which required top-down spatial attention, preferentially engaged the precentral sulcus, whereas metacognition-based selection, which required an internally generated readout of memory quality, selectively recruited the dorsolateral prefrontal cortex. These findings indicate that there are separate neural mechanisms for internally- (metacognition-based) versus externally-driven (cue-driven) working memory selection.

This work was supported by the NYUAD Center for Brain and Health, funded by Tamkeen under NYUAD Research Institute grant CG012

53.359 NEURAL MECHANISMS OF VISUAL WORKING MEMORY CAPACITY LIMITATIONS IN THE COMMON MARMOSET

Tsz Wai Bentley Lo¹ (tlo57@uwo.ca), Susheel Vijayraghavan¹, Lyle Muller², Julio Martinez-Trujillo^{1,3}; ¹Department of Physiology and Pharmacology University of Western Ontario, ²Department of Mathematics University of Western Ontario, ³Department of Psychiatry, University of Western Ontario

Visual working memory (vWM) is a fundamental cognitive function essential for goal-directed behaviour; however, it is characterized by strict capacity limits. While the lateral prefrontal cortex (LPFC) is critical for vWM maintenance, the computational mechanisms constraining capacity in New World primates remain understudied. We investigated these limits in the common marmoset (*Callithrix jacchus*), leveraging their lissencephalic cortex to implant 3D volumetric arrays (N-Form, Plexon) in LPFC areas 8/46. Uniquely, we utilized wireless electrophysiology to record neural activity while animals performed a touchscreen-based Delayed Non-Match to Position (DNMP) task under freely moving conditions. Behaviourally, performance declined significantly with increasing load, reflecting classic capacity limitations. We recorded from 994 neurons in 2 marmosets. We identified single neurons (33%) exhibiting spatially selective persistent activity during memory maintenance. A topographical analysis identified a distributed "salt-and-pepper" functional architecture in LPFC, distinct from the retinotopy observed in early visual areas. To investigate the computational basis of capacity limits, we compared coding models of population activity. Neural firing rates scaled sublinearly with load, consistent with a divisive normalization model, which was statistically preferred over Summation and Winner-Take-All models based on model comparison metrics (Friedman $p < 0.05$). Principal Component Analysis (PCA) revealed that population activity for multi-item trials occupied a low-dimensional subspace intermediate between single-item subspaces. This "geometric averaging" suggests that normalization compresses the representational space to maintain dynamic range stability. However, it reduces the signal-to-noise ratio for representations of individual items. These findings suggest that vWM capacity limits are governed by population-level normalization dynamics in the LPFC.

TUESDAY MORNING POSTERS IN PAVILION

Attention: Capture 1

TUESDAY, MAY 19, 8:30 AM – 12:30 PM, PAVILION

53.401 BEHAVIORAL ALPHA RHYTHMS REVEAL SUPPRESSION OF PREDICTABLE DISTRACTORS

Suyeon Kim¹, Oakyoon Cha¹; ¹Department of Psychology, Sogang University

Salient visual objects tend to capture attention, even when they are irrelevant. For example, a red apple on a desk can draw attention away from a document you are searching for, unless you are prepared to ignore it. In Experiment 1 of Betteto et al. (2025), participants responded more slowly when a color singleton unexpectedly appeared in the search array (singleton-standard) compared to when no singleton was present (singleton-absent). In contrast, responses were comparable to the singleton-absent condition when the singleton repeatedly appeared at a predictable location (singleton-repeated). Predictable singleton distractors may allow top-down suppression, resulting in reaction times comparable to those without a distractor. However, similar reaction times do not reveal whether additional suppression processes are engaged when the distractor is predictable. To address this, the present study examines whether the strength of suppression differs depending on the presence and predictability of a singleton distractor. To estimate the strength of distractor suppression, we examined rhythmic fluctuations in response times at the alpha frequency (8–12 Hz), which is associated with attentional control. We reanalyzed behavioral data from Experiment 1 of Betteto et al. (2025) using PATS (parametric analysis of temporal structure; Cha & Blake, 2024), a method that detects rhythmic temporal patterns from relatively small datasets. PATS showed that the singleton-repeated condition produced larger alpha-frequency fluctuations within the response time distribution than the singleton-standard and singleton-absent conditions. This pattern was not observed in our reanalysis of Experiment 2, where the distractor did not consistently appear at a fixed location. Overall, these findings suggest that predictable distractor locations trigger stronger inhibitory control, and that the magnitude of behavioral alpha rhythms reflects phase-resetting associated with this suppression.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) (No. RS-2023-00211668).

53.402 THROUGH THE STUDENT'S LENS: CAPTURING VISUAL ATTENTION IN REAL-WORLD CLASSROOMS

Xiaorui Xue¹, Ido Davidesco¹, Na'ama Av-Shalom¹, Marina Vasilyeva¹, Nanyu Zhang¹, Rachael Sabelli¹, Jason Geller¹; ¹Boston College

Attention is highly dynamic, and students frequently shift their focus from the learning task to unrelated thoughts—a phenomenon known as mind wandering (MW). Although MW can sometimes be beneficial, it is generally viewed as detrimental to learning. Prior estimates suggest that students spend roughly 30% of class time MW, but these estimates rely on self-reports, which could be inaccurate. Laboratory studies suggest that there is a tight link between attention shifts and eye movements, but it remains unclear whether eye-movement patterns can reliably index MW in authentic learning environments. To address this gap, we conducted a classroom-based study using wearable eye-tracking glasses (Pupil Labs Neon). Data were collected from 15 college students over an 8 week period. Each student completed two regular class sessions while wearing eye tracking glasses. During each session, students received 6–8 probes asking whether they had been mind wandering and to report their cognitive load immediately beforehand. At the end of each session, students completed a short quiz to assess their learning. Preliminary analyses show that students reported MW on approximately 28% of the probes. Furthermore, the frequency of mind wandering reports was negatively correlated with quiz scores ($r = -.27$, $p = .035$), suggesting that greater mind wandering was linked to poorer learning outcomes. Reports of MW were associated with longer fixations and reduced blink frequency in the 60 second period before each probe. These preliminary results support the use of eye-tracking metrics as indicators of attention fluctuations in real-world environments. This pilot study demonstrates the potential of using eye-tracking measures to capture the dynamics of attention in real-world classrooms and highlights opportunities to advance methodological approaches for detecting MW.

This material is based upon work supported by the National Science Foundation under Grant No. 2526250

53.403 AN EVIDENCE ACCUMULATION AND REINFORCEMENT LEARNING MODEL OF DISTRACTOR SUPPRESSION

Isaac Savelson¹, Tzu-Yao Chiu¹, Brandon M. Turner¹, Andrew B. Leber¹; ¹The Ohio State University

The visual world is full of irrelevant but distracting information that must be efficiently handled to support effective visual processing. Fortunately, distracting stimuli often exhibit regularities, which can be exploited to minimize their interference. One such example is humans' ability to quickly learn to ignore spatial locations where distractors commonly appear. However, despite extensive empirical research on the subject, the mechanisms behind learning and implementing spatial suppression are still poorly understood. Here, we sought to enhance our understanding of distractor suppression mechanisms using a computational evidence accumulation model. To this end, we developed a model of the latent decision process associated with overt attentional selection. The model accounts for visual search behavior

through a combination of attentional priority values for the target stimulus, salient distractors, and frequent distractor locations. The attentional priority levels of specific stimulus properties were estimated as a set of attentional weights, which determined the rate at which evidence for an attentional shift accumulated toward candidate display locations. When fit to saccadic response behavior collected during an additional singleton task, our model accounted for individuals' choice-response time distributions remarkably well. To capture the inter-trial dynamics of learned suppression we integrated a reinforcement learning mechanism into our basic diffusion model, allowing us to estimate changes in attentional weights across trials. The model output showed a rapidly learned down-weighting of the frequent distractor location across trials, consistent with the literature's assumptions about learned suppression. Our approach demonstrates the utility of computational modeling techniques in investigating the latent mechanisms underlying attentional capture and distractor suppression.

Funding: NSF BCS-2021038 to ABL

53.404 UNDERSTANDING THE RELATIONSHIP BETWEEN EYE MOVEMENTS AND SALIENCE ARISING FROM ONE OR TWO VISUAL DIMENSIONS

Alejandro Lleras¹ (alleras@illinois.edu), Zoe J. Xu², Jun-Ming Yu¹, Howard J. H. Tan¹, Simona Buetti¹; ¹University of Illinois, ²University of Washington

The impact of salience on eye movements has been studied mostly via manipulations of luminance, a relatively easy to manipulate feature dimension. Humans as well as monkeys show an inverse relation where the stronger the luminance is, the smaller the saccadic initiation time. This finding has been modeled in terms of the time it takes for signals to reach the superior colliculus. Here, we explore the relationship between salience and the eye movement system along three different visual dimensions: color, shape, and orientation. We measured the time taken to initiate and land on a unique element. The unique element was randomly presented in a circle around fixation, embedded in a 12 x 12 grid of identical objects. In Experiment 1, the unique element was defined by a difference along a single dimension, with four different levels of salience, and all three feature dimensions, randomly varied along trials. We replicated the finding that saccadic initiation times are inversely related to salience in all three dimensions. In Experiments 2A-C, the unique element was defined along two dimensions (color-orientation, color-shape, and shape-orientation). Saccadic initiation times were not associated with the precision of the saccades in neither unidimensional ($R^2=0.24$) or bidimensional conditions ($R^2=0.31$). However, saccadic initiation times were strongly and linearly predicted by the strength of the guidance signal in both unidimensional ($R^2=0.93$) and bidimensional conditions ($R^2=0.85$), with the slope of the bidimensional condition being twice that of the unidimensional condition. Importantly, in Experiment 1, all data points fell along the same function, suggesting all visual dimensions produced signals that have equivalent impacts on the eye movement

system. This was also observed in Experiment 2: the bidimensional data fell along the same function, irrespective of which two dimensions defined the unique element. Surprisingly, saccadic initiation times were faster in unidimensional conditions.

53.405 MOVING BEYOND INSTANTANEOUS SALIENCE IN THE ATTENTION-CAPTURE DEBATE

John McDonald¹, Rebecca Carson¹, Daniel Tay¹; ¹Simon Fraser University

A recent attempt to resolve a longstanding debate failed to reach consensus about whether and when salient visual distractors capture attention. Part of the problem may stem from the limited way in which salience is conceptualized in the majority of attention-capture studies. In particular, although people can rapidly detect salient changes in visual stimulation across time, salience is usually defined solely in term of stimulus contrast within individual displays. Here, we integrated the visual search paradigm with the oddball paradigm to investigate the relative contributions of instantaneous spatial salience and temporal salience in the determination of visual selection. In all experiments, infrequent target and nontarget displays were randomly interspersed with frequent standard displays that contained three blue, vertical bars on each side of fixation (herein called standard bars). Participants were instructed to press a button when they noticed a display containing a specific number of yellow items, which varied across experiments. In Experiment 1, all standard bars on one side of fixation were replaced with yellow vertical bars (colour change) or blue horizontal bars (orientation change) to create target and nontarget displays, respectively. Neither of these displays contained any instantaneously salient item, but the target and nontarget bars (herein called change items) still popped out because they differed substantially from standard displays (temporal salience). To determine how the change items were processed, we looked for event-related potential activities associated with deviance detection (visual mismatch negativity), attentional selection (N2pc) and suppression (PD). In Experiment 1, target and nontarget change items triggered the N2pc, which indicates that temporally salient items can capture spatial attention when they are not (instantaneously) salient within individual displays. These results will be discussed in relation to current theoretical perspectives, including salience-driven selection theory, signal suppression theory, and dimension weighting theory.

Funded by NSERC

53.406 EARLY SENSORY PROCESSING SIGNATURES OF OCULOMOTOR CAPTURE ACROSS THE LIFESPAN

Anthony W. Sali¹ (saliaw@wfu.edu), Emily O. Oor¹, Terrence R. Stanford², Emilio Salinas²; ¹Wake Forest University, ²Wake Forest University School of Medicine

During natural behavior, the relevant timescale over which attention varies is on the order of a few tens of milliseconds, and although non-invasive measures of brain activity such as electroencephalography

(EEG) and event related potentials (ERPs) have the requisite temporal resolution, traditional psychophysical measures of attention do not. Across two experiments, we combined ERP measures of attentional selection with a novel oculomotor decision paradigm capable of resolving the contributions of both voluntary and involuntary attention to behavioral performance with unprecedented temporal resolution. Participants completed a compelled saccade task in which they made an eye movement toward or away from a salient color change based on the color. By cuing participants to initiate their movement before the color cue, varying the temporal gap between this go signal and the color cue, referred to as the processing time (PT), and placing a strict limit on response times, we measured the rise and fall of stimulus-driven attentional priority. In both experiments, participants' behavior was characterized by guesses at the shortest PTs, reflexive orienting toward the salient color onset at intermediate PTs, and goal-directed override of stimulus salience at the longest PTs. Furthermore, the amplitude of the P1 ERP component was significantly larger for trials with reflexive responding PTs relative to those associated with goal-directed control. In Experiment 2, aging was associated with a slower rise in reflexive responding as PT increased as well as a delay in goal-directed override of salience-based priority. Together, our results suggest that the ability to override reflexive oculomotor orienting takes longer in older adults and is more broadly associated with increased early sensory processing.

Wake Forest University Intercampus Pilot Grant

53.407 GENERALIZATION OF LEARNED DISTRACTOR SUPPRESSION

Baitong Mu¹ (mub25@wfu.edu), Brad Stilwell¹; ¹Wake Forest University

Recent research suggests that feature-based distractor suppression is a learned skill rather than a passive consequence of statistical learning. This study investigated whether learning to suppress a specific distractor color in one task facilitated suppression of that same color in a novel task, analogous to athletic skill transfer across similar sports (e.g., using a racquet in badminton and tennis). To test, participants completed two paradigms (with task order counterbalanced) while their eye movements were tracked: a gaze-contingent Visual Search task and a Rapid Serial Visual Presentation (RSVP) task. In the Visual Search task, participants searched through an array of two-color squares to find a target letter "F" among "O's" and reported its orientation. In the RSVP task, participants were instructed to identify a centrally presented target letter, rendered in a unique color, within the stream. Prior to the target appearing, at varying lags, task-irrelevant hashtag symbols were flashed surrounding the central stream. Critically, for each participant, across both tasks, the target never appeared in one color (i.e., the consistent distractor color). The remaining pool of colors served as variable distractor colors (i.e., randomly serving as the target or nontarget colors). The key comparison was between the consistent and variable distractor conditions (i.e., distractor suppression) as a function of time. We

hypothesized that for both tasks, participants would show robust learned distractor suppression, such that suppression would build with experience via either task. Crucially, if learning to suppress the consistent distractor color transferred across tasks, distractor suppression should continue from the first task into the second. However, if learned distractor suppression was context specific, then distractor suppression should have “reset” during the second task. We found that learned distractor suppression transferred between tasks, suggesting that learned distractor suppression operates like a trained skill, generalizing across contexts.

53.408 SKILL-BASED LEARNED OCULOMOTOR AVOIDANCE

Tsion A. Eshetu¹ (eshet25@wfu.edu), Brad T. Stilwell¹; ¹Wake Forest University

Sometimes salient stimuli, like a uniquely colored object or flashing neon sign, capture our attention, but other times they are suppressed. Recent evidence suggests that people learn to suppress salient, color singleton distractors via skill-based learning. Though the question remains, how do people achieve this learned distractor suppression, mechanistically? Previous investigations into this question have been limited by end-of-trial metrics such as response times. Here, we use eye tracking to examine the real-time dynamics of learned distractor suppression. To test these dynamics, undergraduate student participants searched for a unique target shape (e.g., a circle) among homogenous nontarget shapes (e.g., diamonds) in the additional singleton paradigm. On half of all trials, one nontarget item was rendered in a unique color, forming a color singleton distractor. Critically, when this singleton distractor was present, it appeared more often in one color than any other color, forming two conditions of distractor frequency (i.e., high versus low). To measure learned distractor suppression, we recorded eye movements, specifically, the destination of first saccades and their duration (i.e., dwell times). This combination of “where” and “for how long” participants directed their gaze allowed us to test three mechanisms. The first, the look-away hypothesis, predicts fewer fixations toward the high- than low-frequency distractors. The second, the rapid disengagement hypothesis, predicts shorter dwell times on the high- than low-frequency distractors. The third, learned oculomotor avoidance hypothesis, predicts the combination of both: fewer first fixations and shorter dwell times on the high- than low-frequency distractors. Consistent with the learned oculomotor avoidance hypothesis, participants were both less likely to direct their initial eye movements toward, and faster to disengage from, the high- than low-frequency distractors. These findings of learned oculomotor avoidance, suggest that when people suppress salient distractors, they do so through skill-based learning of oculomotor control.

53.409 VISUAL DISTRACTION AND EXPERIENCE-BASED SUPPRESSION IN ADHD

Mor Sasi¹ (mor.sasi1992@gmail.com), Tair Vziel, Nitzan Shahar, Dominique Lamy; ¹Tel-Aviv University, ²Sagol School of Neuroscience, Tel Aviv University

Although “easily distracted” is a core diagnostic criterion for attention-deficit/hyperactivity disorder (ADHD), its underlying mechanisms are poorly understood. The few studies on ADHD that relied on experimental tasks specifically designed to study visual distraction in the general population have reported mixed findings. Here, we suggest that failure to consider experience-based suppression—that is, the suppression of irrelevant information following repeated exposure—may explain these inconsistencies. Accordingly, we investigated distraction in ADHD while disentangling distraction by novel salient distractors from experience-based suppression of these distractors. Participants searched for a target shape while a salient color distractor was either present or absent, with the distractor color changing at the beginning of each block. We compared baseline distraction (performance costs on distractor-present vs. distractor-absent trials in the first block halves) and feature-based suppression efficiency (reductions in these costs from the first to the second block halves) in participants with ADHD and matched controls. The results showed that participants with ADHD exhibited greater distraction by novel salient distractors than controls, yet were equally efficient at suppressing these distractors following repeated exposure. This dissociation explains why studies using rare distractors report heightened distractibility in ADHD, whereas studies using frequently repeated distractors do not.

Support was provided by the Israel Science Foundation (ISF)

53.410 CHANGES IN THE ABSOLUTE POSITION OF ATTENTIONALLY TRACKED TARGETS TRIGGER A HEMISPHERIC EXCHANGE AS ITEMS MOVE BETWEEN VISUAL FIELDS

Huiqin Chen¹ (huiqinchen@uchicago.edu), Piotr Styrkowiec², Edward K. Vogel¹; ¹University of Chicago, ²Uniwersytet Wrocławski

Tracking objects that move between visual fields requires coordination between the two hemispheres. The CDA is a sustained negative EEG voltage over the hemisphere contralateral to the positions of the tracked targets. When targets cross the vertical midline, the CDA “flips” in polarity, revealing a handoff of information between the hemispheres (Drew et al., 2014). More recent work has shown that this CDA flip reflects the immediate demands of tracking the target position rather than reflecting prospective allocation of attention to the new region because it still occurs even when the target is expected to return to its original hemifield (Chen et al., VSS 2025). However, it is unclear which aspect of the immediate tracking demands triggered this hemispheric handoff because, during the crossing there was both movement of the target’s absolute position in space as well as relative movement between the positions of the target and its accompanying distractor. Here, we tested whether the change in absolute position on its own was sufficient to trigger the CDA flip even in the absence of

relative changes in the positions of the target and distractor during the crossing. In a condition with both types of motion, we replicated the CDA flip as the target moved between fields and then “flipped back” as it returned to its original field. We compared this to a “no relative motion” condition in which the relative positions of the target and distractor were “frozen” as the items moved between visual fields. The CDA flip still occurred as the target crossed visual fields suggesting that this interhemispheric exchange did not depend on resolving the relative positions of the target and distractor. Thus, the need to track changes in the absolute position of the target appears to be the critical feature for triggering the handoff between hemispheres.

Attention: Models

TUESDAY, MAY 19, 8:30 AM – 12:30 PM,
PAVILION

53.411 SPATIOTEMPORAL NORMALIZATION INCENTIVIZES SELECTIVE ATTENTION ACROSS SPACE AND TIME

Angus Chapman¹, Rachel Denison¹; ¹Boston University

Motivation: Visual attention can be deployed to spatial locations or points in time, resulting in enhanced processing of attended information and impaired processing of unattended information. Such tradeoffs are thought to reflect competition among stimuli for limited attentional resources. Normalization-based models implement attention by enhancing neural excitation prior to divisive suppression, but do not explain why tradeoffs arise. Here, we tested whether normalization, when coupled with an optimal attentional allocation policy, could incentivize tradeoffs in attentional allocation without appeal to an explicit limited attentional resource. Methods: We adapted a dynamic spatiotemporal normalization model to incorporate attentional modulation that varied across space and time. Spatiotemporal normalization was implemented through neurons with spatiotemporal receptive fields, with excitatory and suppressive drives that depended on inputs pooled across space and time. We simulated orientation discrimination performance for two stimuli appearing at varying locations and times, with different degrees of proximity, and 75% valid spatial or temporal cues. Attentional gain was applied independently to both stimuli, without limit, but adjusted to optimal levels that maximized the model's overall performance. Results: Across a range of conditions, model performance was best when attentional gain was allocated selectively to the cued stimulus, even compared to maximally attending both stimuli. Thus, the optimal attention policy led to focused attention in space and time. Importantly, these tradeoffs only occurred when stimuli were close enough in space or time that they mutually suppressed each other via normalization. Additionally, when stimuli were closer in spatial proximity, it was optimal to direct attention slightly away from the cued location, minimizing simultaneous enhancement of the uncued stimulus.

Conclusions: In the presence of normalization, attentional tradeoffs emerge from the optimal allocation of attention in space and time, with no attentional resource limit. Thus, normalization suffices to incentivize focused spatial and temporal attention.

This work was supported by National Institutes of Health National Eye Institute (R01 EY037358) and BU startup funding to R.N.D.

53.412 SPATIAL ATTENTION AND CONTRAST ENHANCEMENT ARE PHENOMENOLOGICALLY SIMILAR BUT MECHANISTICALLY DISTINCT

Peter Neri^{1,2}; ¹Italian Institute of Technology, Erzelli Campus, Genova (Italy), ²Laboratoire des Systèmes Perceptifs, École Normale Supérieure, Paris (France)

When an object is attended, its characteristics are enhanced: for example, its apparent contrast increases. This phenomenological observation suggests that visual attention and contrast perception may share common perceptual mechanisms. We performed experiments in which these two phenomena were independently manipulated within the same experimental paradigm: contrast was manipulated by simply varying the contrast of the entire stimulus display; attention was manipulated via valid spatial cues that could appear either before or after the display. The perceptual process was characterized via a combination of standard metrics (e.g. sensitivity) and a psychophysical variant of reverse correlation. Our results confirm previous findings: at the zeroth order of characterization (measurements of performance and reliability), the two phenomena behave in similar manner (more contrast/attention leads to better performance); at the first order of characterization (measurements of linear filtering processes), they again behave in similar fashion (more contrast/attention leads to higher gain). However, when we examined these phenomena at the second order of characterization (measurements of nonlinear processes), we found that they behave in qualitatively different manners (more attention increases gain, while more contrast decreases gain). The observed differences are captured by a computational model in which visual stimuli are mapped to a perceptual surface, and the process of selecting a target is dictated by distance measurements on said surface. Within this framework, contrast manipulations vary the curvature of the surface, while attentional deployment modulates the degree of noisiness associated with traveling across the surface. Our results therefore point to two distinct mechanisms underlying attentional deployment and contrast perception: if we conceptualize perceptual decisions as agents exploring high-dimensional sensory representations, varying stimulus contrast alters the geometry of such representations, while attending to a stimulus increases the precision with which perceptual agents can track shortest paths across the representation.

Supported by IIT (grant IVXX000701), the Agence nationale de la recherche (grants ANR-16-CE28-0016, ANR-19-CE28-0010-01, ANR-10-LABX-0087, ANR-10-IDEX-0001-02, ANR-22-CE93-0013), and CNRS.

53.413 ATTENTION ENABLES FLEXIBLE ENERGY-EFFICIENT VISION

Eivinas Butkus^{1,2}, Zhuofan Ying^{1,2}, Nikolaus Kriegeskorte^{1,2};
¹Columbia University, ²NSF AI Institute for Artificial and Natural Intelligence

Primate vision is thought to rely on attention to overcome capacity limitations, including severe energetic constraints. Yet how attention enables efficient use of energy—and whether the benefits outweigh the costs of attentional machinery—has never been rigorously demonstrated. Here we provide a mechanistic account of how attention supports energy-efficient vision and flexible energy-accuracy trade-offs. Our model, EAN (“Energy-efficient Attention Network”), implements attentional selection as recurrent top-down multiplicative gain varying over features, space, and time within a convolutional neural network. We optimize a joint objective balancing high accuracy and low energy costs. The energy cost accounts for action potentials and synaptic transmission throughout all model components, including the attention controller. We test EAN on a novel visual-category-search (VCS) task requiring identification of a digit’s class (what, 0-9) and location (where) among distractor letters. EAN learns to attend broadly in an initial pass and then dynamically focuses its energy on task-relevant locations and features of the particular scene at hand. The attention mechanism substantially improves energy efficiency and enables flexible trading of accuracy against energy on a trial-by-trial basis. The model variant with both feature and spatial gain is most efficient and best explains human errors and difficulty judgments. EAN readily generalizes to classical attention tasks and captures canonical single-cell electrophysiological attentional effects on firing rate, Fano factor, and noise correlations. Our work connects a cognitive function (attention), a neural mechanism (gain modulation), and a neurobiological constraint (metabolic costs) in a unified mechanistic account of how recurrence and selection enable flexible, energy-efficient vision.

53.414 REAL-WORLD GAZE IS STRUCTURED BY TASK: EVIDENCE FROM LARGE-SCALE EGOCENTRIC DATA

Lloyd A. Thompson¹, Stephanie M. Shields¹, Kathryn Bonnen¹;
¹Indiana University Bloomington

Understanding gaze allocation during real-world behavior requires measurements that go beyond static image-viewing and controlled laboratory tasks. The Visual Experience Database (Greene, Balas, Lescroart, et al., 2024) provides the opportunity to analyze how eye movements unfold during natural everyday activities via head-mounted video and binocular gaze data collected from 56 participants across hundreds of labeled tasks. Here, we asked whether task-specific fixation priors reflect stable task structure and improve gaze prediction. We developed an automated pipeline for processing VEDB sessions, temporally aligning binocular gaze to video frames and identifying recordings that met inclusion criteria (>10k gaze samples, ≥40% of samples localized with high confidence). This yielded 375 usable sessions containing one or more labeled task segments. Tasks

were grouped into 10 categories: locomotion/navigation, household chores, food preparation, sports/games, cognitively demanding tasks, social interaction, object manipulation, artistic tasks, technology use, and other. We computed a global egocentric fixation prior by pooling valid gaze samples across all task segments, providing a static baseline for the overall spread of gaze. We then aligned the global prior to the center of gaze for each task segment to generate a dynamic set of context-specific priors. We evaluated the predictive value of static and dynamic priors using AUC (area under the curve) and NSS (normalized scan-path saliency), computing results per task segment before pooling per task category. Dynamic priors consistently outperformed the static prior. Gains were strongest for tasks involving strong visuomotor structure (e.g., object manipulation, household chores, sports/games, and cognitively demanding tasks) and weakest for tasks like social interaction and locomotion/navigation. These results demonstrate that gaze in natural behavior is organized around stable, task-dependent fixation strategies rather than a single uniform center bias. Our findings highlight the importance of dynamic, context-conditioned priors for next-generation models of naturalistic gaze behavior.

53.415 DIFFERENT ATTENTIONAL REGIMES MAINTAIN GLOBAL VISUAL COHERENCE: A UNIFIED SIGNAL DETECTION ACCOUNT OF SUBJECTIVE INFLATION AND ATTENTIONAL BOOSTING

Lana Okubo^{1,2} (lanakubo@gmail.com), Shin'ya Nishida¹, Kiyofumi Miyoshi¹;
¹Kyoto University, ²Japan Society for the Promotion of Science

Scholars have long debated the relationship between attention and consciousness. An important challenge is explaining how we maintain coherent conscious experience across the entire visual field, despite that attention can only be allocated to specific locations or features. We propose that this is achieved through the concurrent operation of two attention-related phenomena: (1) attentional boosting, in which an attended stimulus appears more intense when compared to an unattended stimulus, and (2) subjective inflation, in which a stimulus can paradoxically become more detectable under inattention when tested in yes/no detection. Using the signal detection theory (SDT) framework, we demonstrated that these seemingly conflicting phenomena can indeed occur simultaneously under the identical set of parameter values. Specifically, we conducted an SDT simulation of comparative judgment between two stimuli (relevant to attentional boosting) and yes/no detection of a stimulus (to examine subjective inflation). Evidence from each stimulus was assumed to be normally distributed across trials. The critical assumption was that attention (attended vs. unattended) modulates both the mean and variance of evidence distributions. Following our previous findings, we posited that attended stimuli are encoded with higher mean strength and reduced variance, whereas unattended stimuli are represented with lower mean strength and increased variance. Importantly, these encoding regimes interacted with each task’s demands, simultaneously

accounting for both phenomena. In yes/no detection, increased variance in the unattended condition raised the probability of a weak stimulus producing evidence exceeding the threshold, providing a mechanistic explanation for subjective inflation. In contrast, variance showed a limited impact on comparative judgment, and the elevated mean of attended compared to unattended distributions yielded the typical attentional boosting effect. Our computational perspective bridges classical attention theory with contemporary consciousness research, showing how our visual system prioritizes specific locations while maintaining broadly stable perception by preserving stimulus visibility in unattended areas.

This work was supported by JSPS KAKENHI Grant Number 24KJ0137 and 25K00896.

53.416 A TASK-OPTIMIZED VISION-LANGUAGE NEURAL NETWORK PRODUCES HUMAN-LIKE LINGUISTIC CUEING DESPITE LACKING BUILT-IN ATTENTION MECHANISMS

Jonathan Skaza¹, Sana Shehabi¹, Miguel P. Eckstein¹; ¹University of California, Santa Barbara

Introduction: Feedforward CNNs trained for target detection exploit visual cues and show emergent human-like behavioral signatures of covert attention (Srivastava et al., 2024) and associated mechanisms that parallel neurophysiological findings (Srivastava et al., 2025), despite lacking built-in attention mechanisms. Human covert attention is not limited to visual information; linguistic cues can modulate spatial attention (Barnas et al., 2024). Here, we assess whether linguistic cueing can emerge in neural networks that combine visual and language input (a task-optimized CNN fused with an MLP operating on language embeddings) and compare these to linguistic cueing effects in humans. **Methods:** Humans performed a peripheral detection task in which a tilted line target ($15^\circ \pm 3.5^\circ$ SD) appeared at one of two locations with 50% probability, alongside a distractor ($7^\circ \pm 3.5^\circ$ SD). Target-absent trials contained only distractors. Each trial began with an auditory cue predicting the target location with 80% validity (e.g., “attend left”, “monitor 3 o’clock”). To model this experiment, a task-optimized CNN processed the visual display, and a pre-trained language encoder produced embeddings of the linguistic cue. These representations were transformed via an MLP and fused with the CNN features before classification, trained on the same detection and cue-validity structure as the human experiment. **Results:** Both humans and the CNN-MLP fusion model showed linguistic cueing, with higher hit rates on valid versus invalid cue trials ($\Delta_{\text{human}} = 0.102$, $\Delta_{\text{model}} = 0.098$). Humans and the model displayed similar variation in cueing magnitude depending on the language formulation: cues phrased directly (e.g., “focus left”) produced larger effects than negated or less-direct cues (e.g., “ignore 3 o’clock”). **Conclusion:** A task-optimized CNN-MLP model reproduces human-like linguistic cueing effects, suggesting that linguistic modulation of spatial detection can emerge from multimodal integration under task optimization, without requiring explicit attentional mechanisms.

Temporal Processing: Neural mechanisms, models

**TUESDAY, MAY 19, 8:30 AM – 12:30 PM,
PAVILION**

53.417 LONG-RANGE TEMPORAL INTERFERENCE REVEALS DISTINCT FORWARD AND BACKWARD EFFECTS

Ilanit Hochmitz¹ (ilanit57@gmail.com), Yaffa Yeshurun; ¹University of Haifa

Target identification is impaired when it is preceded and/or succeeded by task-irrelevant items. When such temporal interference occurs at relatively long intervals, it is termed temporal crowding. Temporal crowding has been observed at stimulus-onset-asynchronies (SOAs) that clearly exceed the limits of visual masking ($\text{SOA} > 150\text{ms}$), emerging even when items are separated by almost half a second. Here, we aimed to advance our understanding of the temporal dynamics that give rise to temporal crowding by dissociating the roles of backward and forward interference. Across four experiments, we used an orientation-estimation task, in which observers viewed sequences of two or three oriented items separated by relatively long intervals (five SOAs: 175–475ms). In three-item sequences, the target appeared either first, second, or third, depending on the experiment. These different target positions introduced different types of interference: primarily backward, both backward and forward, or primarily forward, respectively. In two-item sequences, the target always appeared first, followed by a single distractor, providing a minimal case of backward interference. Mixture-model analyses revealed that temporal crowding was strongest for the second item, with significant SOA effects on target-encoding precision and substitution errors, but not on guessing rate. A similar, though weaker, pattern of effects was found when the target was first, whereas minimal interference was observed when the target was third. In two-item sequences, even a single succeeding distractor was sufficient to elicit interference: precision increased with SOA, and substitution errors decreased. Thus, temporal crowding arises from a combination of strong backward and weaker forward interference, and this long-lasting impairment can emerge even with a single distractor. Surprisingly, we found stronger forward interference in substitution errors, manifested as higher rates of reporting the preceding than succeeding distractor. Moreover, source confusion did not reflect a complete loss of temporal-order information but rather a coarser representation of items’ order.

53.418 TOWARD A UNIFIED MECHANISM LINKING ALPHA OSCILLATIONS, APERIODIC ACTIVITY, AND VISUAL TEMPORAL PROCESSING

Michele Deodato^{1,2} (md5050@nyu.edu), Vincenzo Romei², Jason Samaha³; ¹New York University Abu Dhabi, ²University of Bologna, ³University of California Santa Cruz

EEG alpha oscillations are thought to impose pulsed inhibition, producing rhythmic fluctuations in neural excitability with clear sensory consequences. Numerous studies show that visual events are more likely to be perceived near the excitatory phases of the alpha cycle (typically around the trough) and suppressed near the peak, suggesting that alpha rhythms define temporal windows of enhanced sensory processing. The aperiodic structure of the EEG power spectrum has independently been linked to cortical excitation–inhibition (E/I) balance and correlates with alpha power, implying overlapping circuit-level mechanisms. Recent evidence indicates that multiple measures of temporal processing are predicted not only by individual alpha frequency and phase but also by the aperiodic exponent. For example, visual temporal fusion thresholds are jointly predicted by alpha dynamics and aperiodic activity: faster alpha rhythms correspond to finer temporal resolution, whereas flatter aperiodic slopes correspond to broader integration windows and shallower psychometric functions. These influences are partially independent, with both periodic and aperiodic EEG components contributing unique variance to perceptual timing. Here, we replicate these findings in a new dataset and propose a novel mechanistic account that unifies periodic and aperiodic effects through a common process. We suggest that global shifts in excitation–inhibition alter alpha-cycle dynamics such that a greater portion of each cycle falls below the excitability threshold required for sensory processing. This mechanism results in a compression of the high-excitability phase that shortens the effective temporal integration period without requiring changes in the frequency of rhythmic inhibition. This unified account provides a robust, replicable framework for understanding how cortical inhibition shapes the temporal structure of visual perception.

53.419 DYNAMIC SCALING OF TEMPORAL NORMALIZATION IN HUMAN EARLY VISUAL CORTICES

Minsun Park¹, Tomas Knapen^{2,3}, Sam Ling¹; ¹Boston University, ²Spinoza Center for Neuroimaging, KNAW, ³Vrije University Amsterdam

To process dynamic inputs efficiently in the temporal domain, the visual system does not simply accumulate sensory inputs linearly. Instead, neural responses to stimuli have been shown to exhibit compressive temporal summation, typically explained by a delayed normalization (DN) model. This model assumes a single decay constant to account for temporal normalization, but prior work has largely restricted testing to sub-second timescales –regimes within which a single decay constant is likely to suffice. However, it is known that neural adaptation doesn't adhere to a single fixed timescale, but exhibits power-law behavior. Here, we tested a modified DN model in which the decay constant scales as a power-law function of duration across a wider range of stimulus durations. We used fMRI to measure BOLD responses in human early visual cortex (V1–V3) to arrays of

gratings (45% contrast), which were presented for eight log-spaced durations (33 msec to 4.2 sec), using a fast event-related design. We observed strong sub-additivity in the BOLD response as stimulus duration increased, across all visual areas. The standard DN model with a fixed decay constant successfully predicted responses for brief durations, replicating previous findings. Interestingly though, it failed to account for the response compression at longer timescales. That is, the DN model significantly underestimated the magnitude of compression at longer durations (e.g., > 1 sec). In contrast, a modified DN model with a scaled decay constant does capture visuocortical response amplitudes across a full range of durations, challenging the assumption of a unitary temporal normalization pool. This dynamic DN model operates over a broader continuum of timescales, by flexibly adjusting to the temporal structure of visual inputs.

This work was funded by the National Institutes of Health (NIH) Grand R01EY028163 to S. Ling.

53.420 STRAIGHTENING OF NATURAL VIDEOS THROUGH LOCAL TEMPORAL INTEGRATION

Anne Zonneveld¹ (a.w.zonneveld@uva.nl), Pascal Mettes¹, Iris Groen¹; ¹University of Amsterdam

Predictions about future states of the world play an important role in guiding human behavior. In primate vision, temporal representational trajectories are straightened in neural and perceptual space relative to input space (Hénaff et al., 2019; 2021), supporting linear extrapolation and temporal predictability. This straightening enables time-aware representations that distinguish fine-grained temporally opposite actions, such as opening versus closing a door (Bagad & Zisserman, 2025). To characterize the computational principles underlying straightening, we evaluated a range of deep neural networks, differing in temporal integration (image- vs. video-based), architecture (convolutional vs. Transformer), and training (trained vs. untrained). Using more than 1,000 natural videos from the Bold Moments Dataset (Lahner et al., 2024), we identified properties that facilitate straightening, quantified as reduction in curvature from pixel to feature space. We further assessed temporal coherence in model feature space, a prerequisite for straightening, by demonstrating significantly higher curvature for temporally shuffled compared to unshuffled features. Consistent with prior work, we show straightening is absent in standard image models. However, we find it emerges in late layers of video convolutional neural networks, facilitated by local operations such as 3D convolutions, as a consequence of training. In contrast, we do not observe any straightening in video Transformers. Notably, global-attention based video Transformers also lack temporal coherence in feature space. Together, these findings suggest that temporal coherence is a necessary, but not sufficient, condition for straightening and highlight that local temporal integration of continuous visual information is critical to the straightening of visual trajectories, as performed by the brain.

This work was supported by the UvA Data Science Centre, as part of the Human Aligned Video AI Lab.

53.422 EEG NEUROMETRIC FUNCTION MIRRORS PSYCHOMETRIC PERFORMANCE IN VISUAL MOTION PERCEPTION

Seungbeom Seo¹ (ssb3304@unist.ac.kr), Oh-Sang Kwon¹, Sung-Phil Kim¹; ¹Department of Biomedical Engineering, Ulsan National Institute of Science and Technology, South Korea

Perceptual sensitivity to global motion in random-dot kinematograms (RDK) improves as motion coherence increases, yielding a characteristic psychometric function of the relationship between coherence and accuracy. Neurophysiological studies have identified neural correlates of this effect at multiple stages of visual processing. An early occipito-temporal event-related potential (ERP) component becomes more negative, and a later centro-parietal ERP component becomes more positive at higher coherences. Single-neuron recordings in monkeys likewise reveal neurometric response functions that parallel the animal's psychometric performance. However, it remains unclear whether non-invasive human electroencephalography (EEG) signals can exhibit a similar quantitative correspondence with behavioral performance. We recorded EEG from human participants performing a motion-direction discrimination task with varying levels of motion coherence in RDK. We analyzed ERP components previously linked to sensory evidence coding and decision-related processing, focusing on a late occipital slow-wave negativity (~550-800ms post-stimulus) and a centro-parietal positivity (~400-700ms). Behavioral accuracy increased monotonically with coherence and was well fit by a psychometric curve. Concurrently, EEG measures varied systematically with coherence, closely mirroring behavioral accuracy. The occipital component became less negative at higher coherence levels, while the centro-parietal positivity increased in amplitude with increasing coherence. Fitting a similar curve to the EEG data yielded neurometric functions that closely resembled the psychometric function and were broadly consistent across participants, indicating a robust correspondence between the brain signals and perceptual performance. ERP indices putatively linked to sensory coding and decision-related processing showed partially dissociable neurometric profiles, suggesting that sensory and decisional stages track motion coherence with distinct dynamics. Our findings demonstrate that EEG-derived neurometric functions can parallel human perceptual performance in a motion-direction discrimination task. In this sense, ERP components provide a quantitative link between EEG signals and perceptual decision-making, offering a useful framework for future investigations in visual cognition research.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2025-00522357).

53.423 HEMISPHERIC ASYMMETRY IN THE VISUAL AND AUDITORY THALAMIC ACTIVITY IN DYSLEXIA

Josiane Mukahirwa¹, Qianli Meng, Keith Schneider; ¹University of Delaware, ²University of Delaware, ³University of Delaware

The magnocellular (M) pathway plays a critical role in both the visual and auditory systems, supporting the rapid processing of transient stimuli. Abnormal development of the magnocellular system has been proposed as a potential neural origin of developmental dyslexia, but remains debated, with alternative accounts emphasizing cortical malformations linked to language processing. One way to advance this debate is through machine-learning-based classification that evaluates the predictive power of thalamic versus cortical structures. Anatomical work by Galaburda & Livingstone (1993) demonstrated reduced magnocellular neuron size in the LGN and a hemispheric asymmetry in the MGN, with controls showing leftward dominance reversed in dyslexia. These findings suggest that thalamic magnocellular asymmetry may serve as a diagnostic marker, motivating fMRI approaches that model hemispheric differences. This study tests whether LGN/MGN asymmetries improve dyslexia classification. Nineteen participants (9 with dyslexia, 10 controls) underwent 3T fMRI during presentation of transient stimuli (abrupt onsets/offsets) and sustained stimuli (smooth temporal transitions). Regions of interest for the LGN and MGN were manually delineated on T1-weighted and proton-density images. Transient minus sustained (T-S) beta maps were computed. The MGN and LGN were each anatomically divided into five inferior-to-superior bins in the left and right hemispheres, generating feature sets designed to capture potential magnocellular-parvocellular gradients and hemispheric asymmetries. Feature selection was performed using LASSO regression with 500 bootstrap resamples; features selected in more than 70% of models were retained and entered into a logistic regression classifier. Classification performance was evaluated using leave-one-out cross-validation. The most inferior MGN bin in the left hemisphere, consistent with putative magnocellular layers, was the most consistently selected feature across bootstrap LASSO models. Logistic regression using this feature classified dyslexia with 78% accuracy. These findings provide functional evidence for thalamic hemispheric asymmetry in dyslexia, aligning with anatomical reports and supporting the subcortical magnocellular deficit hypothesis.

53.424 INVESTIGATING THE NEURAL MECHANISMS UNDERLYING THE TEMPORAL STRUCTURE OF VISUAL AWARENESS

Giulia Gennari¹, Biyu J. He¹; ¹New York University Grossman School of Medicine

Our subjective reality corresponds to a continual stream of perceptual updates. Although our experience seems to unfold in real time, psychophysical evidence supports a two-stage model of perception: conscious updates are preceded by long-lasting periods of unconscious processing. This model posits a divergence in representational formats, with unconscious analysis preserving high

spatiotemporal fidelity to the incoming input and conscious percepts reflecting simplified, post-hoc interpretations. Yet, the neural substrates of these two phases remain unknown. Here, we used magnetoencephalography and multivariate decoding to delineate the neural underpinnings of unconscious processing versus phenomenology. Subjects performed a motion-induced blindness (MIB) task, in which the illusory disappearances and reappearances of a salient target reflect perceptual updates that follow periods of unconscious processing. Decoders were trained on brief localizer trials where the target was physically present or physically absent. Because these trials involved no task, the decoders remained blind to MIB-related decisional or motor processes. Our analyses uncovered a dissociation between slow cortical potentials (SCP, 0.05-5Hz) and the amplitude of gamma-band activity (30-60Hz). Whereas classifiers trained on SCP successfully discerned target disappearances and reappearances in awareness, classifiers trained on gamma amplitude could do so only in phenomenally-matched control trials, where perceptual changes had a direct physical counterpart. During illusory blindness, gamma-band classifiers always predicted the target as “present”, indicating faithful tracking of the physical input. Leveraging systematic manipulations of the target’s visual attributes, we further examined sensory processing in relation to the currently perceived content. Neural patterns in the gamma band encoded target’s veridical features and their evolution with high precision, irrespective of conscious access. Conversely, SCPs supported the decoding of perceived features as they (re)entered awareness, irrespective of their sensory accuracy. Overall, these findings reveal a frequency-domain division of labor in the brain, with distinct activity orchestrating the temporal structure of perception.

This work was supported by the Leon Levy Foundation, by the Charles H. Revson Foundation and by the National Institute of Health (R01EY032085).

53.425 MODELLING THE NEURAL DYNAMICS OF VIDEO PERCEPTION: FROM INCREASINGLY COMPLEX STATIC OBJECT FEATURES TO MID-LEVEL DYNAMIC ACTION FEATURES

Christina Sartzetaki¹, Anne W. Zonneveld¹, Pablo Oyarzo², Alessandro T. Gifford², Radoslaw M. Cichy², Pascal Mettes¹, Iris I.A. Groen¹; ¹Informatics Institute, University of Amsterdam, ²Department of Education and Psychology, Freie Universitat Berlin

Human visual perception unfolds in a dynamic world, yet most vision research relies on static images that lack temporal context, which shapes both cognitive processing and neural dynamics. Deep neural networks (DNNs) have proven effective models of visual cortex processing for static images, but their ability to capture dynamic visual processing in the brain remains less explored. To investigate DNN alignment to the neural dynamics of natural video perception, we here employ a newly collected EEG dataset with the same video stimuli as Bold Moments Dataset (Lahner et al., 2024). We compare these EEG recordings to representations from 100+ DNNs varying in temporal

integration, classification task, architecture, and pretraining, using Cross-Temporal Representational Similarity Analysis (CT-RSA) to identify the best-correlating model time-points and layers with evolving neural responses. We find that responses in posterior electrodes initially correlate best with low-to-high level layers of static object-trained models, reflecting processing of increasingly complex features across time similar to image perception. Strikingly however, this is followed by an extended period of posterior activity best correlating with mid-level layers of temporally-integrating action models, while exhibiting high temporal correspondence to the unfolding video content. In contrast, frontal electrodes show only early correlations with high-level action model layers, whilst lacking temporal correspondence to the video content. Regarding the variation factors of architecture and pretraining, state-space models as well as self-supervised pretraining show increased alignment to intermediate posterior activity, pointing to potential benefits of recurrent processing and general pretext tasks. Overall, our findings suggest that during continuous visual perception the brain integrates information beyond the canonical temporal hierarchy of low-to-high-level feature processing seen in static image perception, uncovering a novel stage of posterior processing related to mid-level dynamic action features and following early frontal engagement with high-level action features.

CS is supported by an ELLIS Amsterdam Unit grant to IIAG. AWZ acknowledges support from the UvA Data Science Centre, as part of the Human Aligned Video AI Lab.

53.426 TEMPORAL SIGNATURES OF COLOR ARE SUFFICIENT TO DECODE COLOR FROM A SINGLE MEG SENSOR

Quinn Battagliese¹ (quinn.battagliese@nih.gov), Sabrina Ripsman¹, Bevil Conway¹; ¹National Institutes of Health

Color can be decoded using multi-sensor magnetoencephalography (MEG), with average peak decoding of 51% accuracy (chance 12.5%) 120 ms after stimulus onset (Rosenthal et al., Current Biology, 2020; Hermann et al, Nature Communications, 2022). To what extent does decoding depend on the temporal versus spatial pattern of activity? If temporal information is sufficient, decoding should be possible using data from one sensor; if spatial information is exploited, then temporal structure should differ across sensors. We addressed these questions by analyzing the MEGco data set (8 colors, 375 trials/color/subject, 18 subjects) from Rosenthal et al. and Hermann et al., which describe the details of the color-viewing task and decoding strategy. In the original analysis, for each subject, the 25 sensors whose activity co-varied most significantly with the training labels were analyzed. Here we evaluate the extent to which decoding is successful using data from each of these top 25 sensors by itself. Single-sensor decoding in at least one sensor was successful in all 18 participants, over 47 consecutive time bins (SD=21), with average peak decoding of 32% [30-38%, 95% CI] at 95 ms after stimulus onset [range:85ms-270ms]. Across subjects, 22 (SD=3) sensors yielded significant decoding accuracy lasting more than 15 consecutive time bins. These results

show the brain represents color by the temporal structure of the response. To determine the underlying temporal structure that enables single-sensor decoding, we analyzed the univariate responses measured by each sensor. A multivariate repeated measures ANOVA showed significant differences ($p < 0.05$) in peak latency of the colors across sensors in 14/18 subjects ($F(7,168)$ range: 3.32-18.51) and significant differences in peak magnitudes of the colors in 11/18 subjects ($F(7,168)$ range: 2.77-7.51). These results show, in most participants, multi-sensor MEG can detect differences in the spatial representation of colors across sensors.

This research was supported by the NIH IRP (NIH, 1ZIAEY000558 to BRC). Contributions of NIH authors are Works of the United States Government. Conclusions in this paper are of the authors and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

53.427 WHICH VISUAL FEATURES SHAPE THE REPRESENTATIONAL GEOMETRY FOR PREDICTION?

Jiaming Xu¹ (jiaming.xu@utexas.edu), Sai Prashanth Raja Sundaram¹, Youjin Oh¹, Mary Hayhoe¹, Robbe L. T. Goris¹; ¹University of Texas at Austin

Under natural circumstances, visual inputs have a rich and complex statistical structure. Our visual system must exploit this structure to support key tasks such as temporal prediction—the ability to anticipate upcoming states of the environment. Recent work has argued that this ability is rooted in an information processing strategy known as temporal straightening. From a geometrical perspective, visual prediction is difficult because the stream of images on the retina evolves along irregular, complex temporal trajectories. The temporal straightening hypothesis proposes that the visual system exploits natural statistical regularities to transform this input into representations that follow straighter temporal trajectories, thereby facilitating prediction through simple linear extrapolation. In natural vision, these statistical regularities are grounded in the spatial organization of the scene, with predictive structure arising from the distribution of luminance, chromatic, and binocular information. Here, we ask which of these information sources are essential for temporal prediction and how the geometry of the representational trajectory changes when specific cues are removed. We hypothesize that selectively removing individual information sources diminishes the statistical specification of the scene and progressively impairs predictive processing, as reflected by a reduction in temporal straightening. We generated stimuli from egocentric videos recorded during locomotion by reconstructing environments in 3D using Gaussian Splatting and replaying the original camera path in Virtual Reality. For each video, we created three conditions: fully naturalistic, color removed, and both color and stereo removed. We used an AXB discrimination task to measure perceptual distances across all frame pairs and computed perceptual curvature from these data. Our initial results (two environments, three conditions, one subject) suggest that the impact of removing chromatic or binocular information depends on

scene structure, indicating that the neural computations underlying temporal straightening flexibly adapt to environmental statistics to support context-dependent temporal prediction.

Motion: Mechanisms, models

TUESDAY, MAY 19, 8:30 AM – 12:30 PM,
PAVILION

53.428 STATE-SPACE ENCODING OF DIRECTION AND ORIENTATION DIVERGES BETWEEN V1 AND PMLS AFTER EYE OPENING

Brandon R. Nanfity^{1,2,3}, Kristina J. Nielsen^{1,2,3}; ¹Johns Hopkins School of Medicine, ²Zanvyl Krieger Mind/Brain Institute, ³Kavli Neuroscience Discovery Institute

Sensory systems must explicitly signal information from stimulus inputs, which are usually composed of multiple feature dimensions (e.g. the orientation and direction of drifting gratings). Traditionally, tuning functions are used to quantify how informative single neurons are about a particular stimulus parameter. However, the response pattern from a population of tuned neurons is necessary to disambiguate stimulus identity, particularly in the face of higher dimensional stimuli. In adult animals, the population response has been used to evaluate how well an area distinguishes between stimuli across multiple feature dimensions. Here, we extend this approach to evaluate and compare the function of primary visual cortex (V1) and higher motion area PMLS in developing ferrets. Single neuron responses to gratings drifting in twelve different directions were recorded using extracellular electrophysiology in animals ranging from post-natal day 30 (around eye opening) to 52 (when orientation and direction tuning has matured). We then used principal component analysis to capture any low-dimensional structure in the population response, followed by linear discriminant analysis on the full dimensional data to further evaluate what information (e.g. orientation vs. direction) can be read out at each stage of development. Our results indicate that around eye opening, both areas represent orientation but not direction. In adult animals, however, they have distinct representational geometries. Adult PMLS responses maximally separate gratings moving in opposing directions despite having a common orientation. Thus, PMLS is optimized for encoding direction at the expense of orientation because the representations of each orientation is split across an axis of motion. V1 however balances separating different directions with clustering similar orientations. This difference in functional specialization appears just a few days after eye opening. The decoding analysis demonstrates this comes at the expense of retrieving orientation information in PMLS, whereas V1 maximizes performance in both feature dimensions.

This work was supported by the NIH (1R01EY035807) and the Kavli NDI

53.429 WIDEFIELD AND 2-PHOTON MAPPING OF MARMOSSET AREA MT USING BLOOD-BRAIN-BARRIER CROSSING CAPSIDS

Joseph Wekselblatt^{1,3}, Penny-Shuyi Chen^{1,2}, Declan Rowley^{1,3}, Andrew Steele⁵, Alex Huk^{1,4}; ¹Fuster Laboratory for Cognitive Neuroscience,, ²Neuroscience Interdisciplinary Program UCLA, ³Department of Ophthalmology UCLA, ⁴Department of Psychiatry & Biobehavioral sciences UCLA, ⁵Department of Biological sciences, California State Polytechnic University, Pomona

Introduction: Visual area MT has been studied extensively using electrophysiological techniques, but direct and comprehensive assessment of its functional architecture would benefit from imaging approaches. Here, we relied on the lissencephalic brain of marmosets, combined with novel tools to allow for easier, widespread deployment of calcium indicators (GCaMP) to characterize the functional architecture of MT at both population and cellular levels, assessed via single-photon widefield imaging, and 2-photon imaging, respectively. **Methods:** We packaged a calcium indicator (GCaMP8s) into a capsid engineered to (a) cross the blood brain barrier, (b) target neurons, and (c) spare the liver. We injected this construct (capb10-cag-GCaMP8s) via intravenous tail vein injection into 2 common marmosets. We then conducted MT functional mapping experiments in large fields of view (a 7 mm diameter window, and 3.5x3.5 mm² FOV for widefield imaging). Full-field moving dot stimuli (on vs off) were used to identify motion responsive areas, and then additional moving dot stimuli were used to measure direction tuning, speed tuning, and retinotopy. Widefield imaging was performed over the entirety of MT, and 2-photon cellular-resolution imaging was performed within different smaller ROIs within MT. **Results:** We identified MT in both areas due to strong and relatively homogenous expression of GCaMP throughout the imaging window. This demonstrates the viability of intravenous injections, supporting easier application and more homogenous signal over space. Widefield imaging revealed clustered direction tuning, with motion direction columns on the scale of 400-500 um per 180 degrees of direction, consistent with other estimates primarily from electrophysiological samples. The 2D spatial autocorrelation of direction preference dropped significantly over 500um, further confirming this estimate of the size of a direction hypercolumn. 2-photon imaging: identified 1336 (functionally) visual-motion stimulus driven cell across 13 sessions, and confirmed that the widefield measurements are consistent with the responses of individual neurons.

Grant/Funding support: Fuster endowment, UCLA

53.430 ISOLATING PARVOCELLULAR PATHWAY CONTRIBUTIONS TO THE PERCEPTION OF MOVING REAL-WORLD STIMULI

Matthew N. Martino¹ (mattnmartino@gmail.com), Jess M. Smith¹, Bradford Z. Mahon¹; ¹Carnegie Mellon University

Research on psychophysical dissociations between magnocellular and parvocellular pathways has largely used screens to present visual stimuli to maintain precise control of all stimulus properties. Here we took a step to advance understanding how different visual pathways contribute to perception of real world stimuli. Parvocellular pathways have differentially high sensitivity for stimuli defined by high spatial frequency red-green isoluminant contrast when compared to magnocellular pathways. We exploited this asymmetry between the parvocellular and magnocellular pathways to test motion detection of a ruler (oriented and dropped vertically, at either 2° or 30° to the right of fixation, monitored with an eye tracker). On each trial, the ruler would drop at a random time point (controlled by PsychoPy), and participants' task was to push a button as soon as they detected the drop. Care was taken to control and test for any auditory cues associated with the ruler dropping. The experimental design represented the factorial combination of spatial frequency (low vs high), contrast (isoluminant heterochromatic red/green vs black/white), and eccentricity of the ruler (2° vs 30°). Rulers were painted with bands of alternating contrast perpendicular to the direction of motion. Equivalent luminance (reflectance) between the red and green colors on the ruler was ensured with a photometer. A 2x2x2 ANOVA of response times demonstrated a significant main effect of spatial frequency (low < high), contrast (black/white < red/green), eccentricity (center < periphery), as well as a significant 3-way interaction. Follow-up hypothesis testing confirmed the key prediction that the slowest motion detection was for the high-frequency and red-green contrast ruler presented in the periphery. These findings demonstrate sensitivity to differences between the magnocellular and parvocellular pathways, using physical real-world stimuli. Ongoing work is applying this paradigm to study visuomotor function after focal lesions in the geniculostriate and non-geniculostriate pathways.

This research was supported in part by the Carnegie Mellon University Summer Undergraduate Research Fellowship

53.431 CROSS-DECODING PERCEIVED AND INTENDED VISUAL MOTION IN HMT+

Wentao Si¹ (wentao.si.gr@dartmouth.edu), Eunhye Choe¹, Nathan H. Heller², Patrick Cavanagh³, Viola S. Störmer¹, Peter U. Tse¹; ¹Dartmouth College, ²Johns Hopkins School of Medicine, ³Glendon College, York University

The apparent motion quartet is a bistable stimulus in which two dot pairs alternate in succession: one appearing in the upper-left and lower-right positions, and the other in the opposite corners of an invisible rectangle. This produces alternating percepts of horizontal and vertical motion. Prior psychophysical work has shown that volitional intention – choosing to perceive vertical or horizontal motion – modulates the perceptual bistability of this ambiguous stimulus (Kohler et al., 2008; Sun et al., 2017). Here we investigated the neural basis of intending to see a specific motion direction, using fMRI with a cue-delay, “single-shot alternation” paradigm. **Methods:** We cued observers to perceive either vertical or horizontal motion and

measured neural activity using 3T fMRI during the variable delay period (6s-10s) prior to an instantaneous dot-pair switch. In separate scans, we also recorded neural responses while observers passively viewed apparent or physical motion along the illusory trajectories. Results: Using MVPA and cross-decoding techniques, we found that apparent motion elicited activation patterns in hMT+ similar to those elicited by physical motion, broadly replicating a previous 7T high resolution study (Schneider et al., 2019). In addition to replicating this finding, we also found robust cross-decoding between perceiving physical motion and intending to perceive motion during the delay period in which neither physical nor apparent motion was present. Finally, we also observed reliable (though weaker) cross-decoding performance between apparent motion percepts and intended motion within hMT+. This suggests that similar representations in hMT+ exist while observers intend to see vertical versus horizontal motion and when they actually perceive corresponding motion paths.

NSF Award ID 2241975

53.432 REPRESENTATION OF AUDITORY MOTION IN HMT+ OF EARLY BLIND INDIVIDUALS

Yang YANG¹ (yyang3077@gatech.edu), Kelly Chang², Ione Fine², Woon Ju Park¹; ¹Georgia Institute of Technology, ²University of Washington

INTRODUCTION: hMT+ responds to auditory motion in early blind people. However, it remains unknown whether these responses engage spatiotemporally non-separable motion selectivity mechanisms analogous to those that mediate visual motion processing in the sighted. Recent evidence suggests that both sighted and early blind listeners rely on separable perceptual filters that emphasize stimulus onset and offset to hear auditory motion (Park and Fine, 2023). Here, we tested whether early blind hMT+ fMRI responses are primarily determined by onset/offsets, or monotonically increase to motion coherence, as predicted by non-separable motion mechanisms. METHODS: Seven early blind and eight sighted adults discriminated the direction of auditory motion stimuli that varied in motion coherence. Stimuli were nine broadband noise bursts (500-14000 Hz, 90 ms each). At 100% coherence, bursts changed location coherently over time, producing continuous motion. At 50% coherence, the first and last two bursts matched the coherent sequence, and the middle bursts were shuffled. At 0% coherence, all bursts were shuffled. BOLD responses were measured in functionally defined Planum Temporale (PT) and hMT+. RESULTS: Within PT, a region associated with auditory motion processing, the strongest fMRI responses were observed at 0% coherence and the weakest at 100% coherence in both groups. In hMT+, responses to auditory motion were weak in sighted individuals, whereas in early blind individuals, the pattern of responses was similar to that in PT. An encoding model analysis revealed that these results were best explained by a model based on spatially localized filters tuned for stimulus onsets and offsets. CONCLUSIONS: The recruitment of hMT+ for auditory motion in early blind individuals does not rely on mechanisms analogous to

those used for visual motion in the sighted. Instead, hMT+ adopts novel computations, similar to those found in PT, that are optimized for extracting auditory motion.

NEI R00 EY034546 and Georgia Tech Smithgall-Watts Early Career Award to WP; NEI R01EY014645 to IF

53.433 RELATIONSHIPS BETWEEN SURROUND SUPPRESSION, CROSS-SCALE INTERACTIONS, AND FIGURE-GROUND MOTION SEGREGATION

Ignacio Serrano-Pedraza¹ (iserrano@ucm.es), Raul Luna¹, Marcos Padilla-Ruiz¹, Milagros Tirado-Perez¹, Omar Bachtoula¹; ¹Universidad Complutense de Madrid

There are two well-known inhibitory mechanisms in motion perception. The first is surround suppression, where duration thresholds for motion discrimination are higher for large, high-contrast stimuli compared to smaller ones. The second is the cross-scale interaction, in which the duration threshold for detecting motion of a high spatial frequency pattern increases when a static low-frequency component is added. Previous work has suggested a link between surround suppression and motion-segregation efficiency, but it remains unclear whether motion segregation is linked to cross-scale interactions. To investigate this, we performed three motion-discrimination experiments. In Experiment 1, we measured duration thresholds using binary noise windowed by raised-cosine apertures of 5deg and 0.75deg. In Experiment 2, we measured duration thresholds for discriminating motion of a 3cpd drifting Gabor (4 deg diameter), presented either alone or combined with a static anisotropic noise centred at 1cpd. In Experiment 3, we measured motion-segregation duration thresholds for an oriented ellipse drifting at 4.94deg/s, composed of binary-noise and surrounded by binary noise moving in the opposite direction at the same speed; participants reported the orientation of the ellipse. From Experiments 1 and 2 we computed two indexes: a suppression index (SI), defined as the log-threshold difference between the large and small stimuli; and an interaction index (II), defined as the log-threshold difference between the compound and simple stimuli. SI did not correlate with II and did not correlate with motion-segregation thresholds. In contrast, II showed a significant and positive correlation with motion-segregation thresholds. These findings suggest that surround suppression and motion segregation are independent processes, whereas, for the first time, we show a strong relationship between cross-scale interaction and motion segregation. Therefore, motion segregation does not appear to depend on how strongly observers suppress large stimuli, but rather on how the visual system integrates motion information across spatial scales.

Supported by grant PID2024-155243NB-I00 from Ministerio de Ciencia, Innovación y Universidades (Spain) to ISP.

53.434 PERCEPTION OF CAUSALITY DEPENDS ON SENSORY UNCERTAINTIES AND BIASES IN COLLISION EVENTS

Lukas Maninger¹, Lina Eicke-Kanani¹, Anna-Lena Eckert², Constantin A. Rothkopf^{1,3}, Thomas S. A. Wallis^{1,3}; ¹Technical University of Darmstadt, ²Marburg University, ³Center for Mind, Brain and Behavior (CMBB), Universities of Marburg, Giessen, and Darmstadt

When observing collision-like interactions between two objects, humans can automatically attribute causality, determining whether one object (A) caused another (B) to move. For such launching displays, participants report weaker causal impressions when B's outgoing trajectory deviates more from Newtonian dynamics; however, there is little quantitative work on subjective causality perception. Here, we hypothesize that in the face of uncertainty, participants may construct a counterfactual initial configuration that is in higher agreement with the observed outcome. We collected data from 32 participants, each performing three tasks: 1. predicting B's outgoing trajectory after seeing A's movement; 2. estimating A's start position after viewing a launching display where B's movement direction had different degrees of plausibility given A's trajectory; and 3. rating causality subjectively for similar launching displays as in 2. We developed a Bayesian model to jointly account for behavior across all three tasks by incorporating uncertainties over observations, biases in orientation perception, and decision noise. Reported angles in 1 and 2 were generally accurate but showed repulsion away from the cardinals. Causality ratings tended to decrease with collision implausibility for all but two participants. Our probabilistic model of trajectory prediction and causality rating involving sensory uncertainty was able to capture behavioral data on a subject-by-subject level. Our data shows that causality perception is based at least in part on physical plausibility. Such a mechanism would be consistent with the notion that causality is attributed based on the compatibility of an observation with an internal prediction. However, we note that we found a small tendency to base reports of A's start position on B's implausible movement direction, indicating the possibility of a retrospective component to reconcile unexpected outcomes. While this bias was minor, we speculate that counterfactual simulations could also contribute to causality judgments in real-life scenarios under uncertainty.

This work was supported by the European Research Council (ERC; Consolidator Award 'ACTOR'-project number ERC-CoG-101045783) and the Deutsche Forschungsgemeinschaft (German Research Foundation, DFG) under Germany's Excellence Strategy (EXC 3066/1 "The Adaptive Mind", Project No. 533717223).

53.435 NEURAL NETWORKS REVEAL CANDIDATE COMPUTATIONAL MECHANISMS UNDERLYING ANOMALOUS MOTION ILLUSIONS

Fan L. Cheng¹, Isabella E. Rosario¹, Zitang Sun², Nikolaus Kriegeskorte¹; ¹Columbia University, ²Independent Researcher

Anomalous motion illusions—including peripheral drift illusion (e.g., Rotating Snakes), central drift illusion, and the Ouchi illusion—offer a powerful probe into the visual computations that transform small image shifts on the retina into vivid motion percepts. For the Rotating Snakes

illusion, no single explanation has gained consensus. Proposed neural accounts include contrast-dependent differences in response timing within V1 and MT, nonlinear saturation of motion detectors, and biased local motion estimates arising from early visual processing. These mechanisms suggest that multiple neural computations may jointly contribute to the illusory percept. To examine this possibility, we tested whether modern neural network models can reproduce motion illusions and reveal their underlying computational ingredients. We evaluated ten motion-estimation models—spanning standard deep architectures and biologically inspired models—under simulated viewing conditions (static, onset, microsaccade, and saccade). Across models, the bio-inspired Dual model showed the closest—though still partial—alignment with human perception of Rotating Snakes, and its responses were strongest during microsaccade-like image shifts. In contrast, the Ouchi illusion is well accounted for by models with first-order motion-energy mechanisms, whereas central drift stimuli remain poorly captured by all models. Ablation analyses of the Dual model reveal that first-order mechanisms alone cannot reproduce Rotating Snakes: higher-order motion features increase alignment with expected rotational flow, and recurrence is essential for integrating local cues into a coherent global pattern. Thus, the candidate mechanism identified for Rotating Snakes goes beyond classic first-order latency accounts by requiring additional higher-order and recurrent processing. Our results show that video-computable neural networks can generate testable hypotheses about mechanisms of anomalous motion illusions. The framework highlights gaps between human and machine motion perception and suggests new directions for building models that more closely mirror the computations of the human visual system.

53.436 DOES CONTRAST MODULATE ESTIMATION NOISE IN VISUAL SPEED PERCEPTION?

Daniel Yu¹ (daniel_yu@brown.edu), Jovan Kemp², Fulvio Domini¹; ¹Brown University, ²New York University Abu Dhabi

We depend on our visual speed perception to guide actions in dynamic environments and interpret how physical events unfold. Although we might assume that we can accurately perceive the speeds of moving objects, our speed estimates are often biased by other stimulus properties such as contrast: notably, low-contrast stimuli appear slower than high-contrast ones, especially at lower simulated speeds. Under probabilistic models of speed perception, low-contrast stimuli produce noisier estimates that yield biased responses when combined with a slow-speed prior. However, it is plausible that low-contrast stimuli may directly induce smaller estimates without being correlated with increased noise. To test this hypothesis, we conducted an experiment measuring speed discriminability for gratings with different contrasts. Two gratings were presented sequentially in each trial, and the participants judged which appeared faster. For a given run, the speed of one stimulus was held constant while the other was adjusted by an adaptive staircase procedure; we refer to the fixed-speed stimulus as the reference and the adjusted-speed stimulus as the

comparison. If contrast does modulate estimation noise, then we expect that the discrimination thresholds for stimuli pairs where one is high-contrast and the other is low-contrast should not significantly differ. However, we found that the thresholds depended on the contrast of the comparison, regardless of the reference: high-contrast comparisons yielded smaller thresholds, while low-contrast comparisons yielded larger ones. These results suggest that estimation biases are directly related to visual stimulus quality, given by grating contrast, which may modulate the slope of the response function rather than the noise of visual estimates.

This research is supported by the National Science Foundation under Grant #2522066.

53.437 TASK DEMANDS ALTER WHETHER SPATIAL SUPPRESSION OR SUMMATION IS OBSERVED IN BOTH YOUNGER AND OLDER OBSERVERS

Christopher Taylor¹ (taylorc@neco.edu), Yaffa Kizel; ¹New England College of Optometry

Spatial suppression and summation differ between young and old observers. We attempted to determine whether they could be altered suppression and summation in a simple task using Gabor stimuli. Duration thresholds were measured for discriminating motion (left/right) and/or orientation ($\pm 2^\circ$ from vertical) in 14 younger (19-27 yrs) and 14 older (65-89 yrs) observers. We used Gabor stimuli (1 cycle/deg) with sizes ($2\sigma = 0.7$ or 5 degrees) and contrasts of 2% or 22.5%. The tasks were: (1) two baseline demand tasks, single-dimension stimuli and left-right motion (replication) or orientation (2) two stimulus check tasks with novel dual-dimension stimuli that varied in both motion direction and orientation, with certain responses to either motion or orientation within each task. (3) A dual-dimension stimulus with uncertain single-responses (motion/orientation) that were post-cued randomly on each trial. That is, the observer responded to either motion or orientation, depending on randomly selected post-stimulus cue (4) a dual-task where both motion and orientation responses were collected in randomized order. Seven duration thresholds were measured using a staircase procedure, with each task repeated twice and counterbalanced across two days. In the baseline and stimulus-check tasks for motion, younger observers showed reduced suppression. Orientation showed summation for both groups. When task demands were increased in the uncertain single-response tasks, dual-task summation was observed in both age groups. There was greater summation for older observers in the uncertain-response dual-task orientation tasks. Consistent with previous work, observers exhibited spatial suppression when discriminating motion direction with response certainty. However, when either response uncertainty (single response) or dual-task was used, both younger and older observers showed spatial summation. Task demands can alter whether spatial suppression and summation are observed even in simple spatial vision tasks with identical stimuli. While we used aged observers, this effect could extend to other special populations.

Support: Meta Reality Labs, NECO internal funding, NIH T-35-007149

Action: Pointing, tracking

**TUESDAY, MAY 19, 8:30 AM – 12:30 PM,
PAVILION**

53.438 HUMANS OPTIMALLY INTEGRATE VISION AND PROPRIOCEPTION DURING CONTINUOUS MOVEMENT

Jonathan Tsay¹, Pam Villavicencio², Mabel Ziman¹, Dominik Straub³, Matthias Will⁴, Cris de la Malla²; ¹Carnegie Mellon University, ²University of Barcelona, ³University of Cambridge, ⁴University of California Berkeley

Imagine swatting an annoying fly—tracking its erratic visual motion while continuously sensing your arm's movement through proprioception. Success in such situations hinges on effective multisensory integration of vision and proprioception. Decades of work using static position judgments show that these cues are combined in a statistically optimal manner. However, whether this optimality principle holds during dynamic, continuous movements remains untested. To address this gap, we employed a continuous psychophysics paradigm in which participants used their left hand to continuously track, for 20-seconds along a constrained forward-backward axis, either a randomly (Brownian motion) moving visual cursor, a proprioceptive signal from a passively moved right hand controlled by a motorized manipulandum, or both signals concurrently. We manipulated visual uncertainty (clear vs. Gaussian-blurred cursor) and visuo-proprioceptive alignment (0-cm vs. 5-cm translational shift). Strikingly, the results closely matched the predictions of optimality in several ways. First, visual tracking degraded under high visual uncertainty. Second, visual tracking was worse than proprioceptive tracking, consistent with greater uncertainty in vision arising from slower neural transmission during movement. Third, when cues were aligned (0-cm shift), multisensory tracking exceeded both visual and proprioceptive unimodal tracking performance, demonstrating optimal cue integration for continuous movements. Fourth, when a visuo-proprioceptive discrepancy (5-cm shift) was introduced, multisensory tracking landed between the two cues but was systematically biased away from vision (the more uncertain cue) and toward proprioception (the less uncertain cue). Finally, after exposure to the discrepancy, vision shifted strongly toward proprioception, whereas proprioception showed minimal shift toward vision—revealing that recalibration, like integration, followed optimality principles. Together, these behavioral findings, supported by preliminary Bayesian modeling, show that vision and proprioception are integrated and recalibrated optimally during continuous movement, extending multisensory optimality from static positional judgments to continuous real-world movements.

CM is funded by grant PID2023-150883NB-I00 from the MCIN/AEI/10.13039/501100011033. PV is supported by grant

53.439 INFLUENCE OF VISUAL FEEDBACK NOISE ON SENSORIMOTOR ADAPTATION

Jacob L. Lader¹, Sean O'Bryan^{1,2}, Joo-Hyun Song¹; ¹Brown University, ²Bard College

Successful motor behavior depends on our ability to adjust movements when the environment changes or when outcomes deviate from expectations. A critical factor in this process is the quality of sensory information. Visual feedback is frequently noisy rather than stable, raising questions about how such noise influences the mechanisms that support sensorimotor adaptation. Explicit and implicit learning contribute to this adaptation in different ways, with explicit learning relying on strategic, cognitive resource-dependent corrections and implicit learning reflecting gradual, automatic adjustments based on sensory prediction errors (SPEs). We evaluated how noisy rotated visual feedback affects these learning processes using two visuomotor adaptation (VMA) tasks designed to emphasize explicit or implicit contributions to reach adaptation. In Exp. 1 (N=43), the delayed-feedback VMA emphasized explicit, strategy-based correction by postponing visual feedback to final reach positions. In Exp. 2 (N=40) the clamped-feedback VMA emphasized implicit learning by presenting cursor motion that followed a fixed trajectory independent of the participant's reach. In each task, a noisy visual feedback condition was introduced in which rotated feedback varied randomly across trials while maintaining a 45° mean. Since these tasks differ in the cognitive resources they engage, changes in cognitive effort may reveal how participants respond when error information is unreliable. Pupil diameter was recorded continuously to index these demands during adaptation. Visual feedback noise influenced behavior and pupil dilation differently across the two learning contexts. In the delayed-feedback (explicit) task, noise impaired motor adaptation but did not alter pupillary responses, suggesting that noise disrupted strategic correction without increasing cognitive effort. In contrast, in the clamped-feedback (implicit) task, noise did not affect behavior but increased pupillary responses, indicating additional effort to suppress irrelevant feedback. These results demonstrate that visual feedback noise modulates both sensorimotor adaptation and pupil-indexed cognitive effort, with effects that depend on the learning process engaged.

53.440 INVESTIGATING POINTING ACCURACY IN A TOUCHLESS INTERFACE

Sreehari Vattakkandy¹, Balagopal Raveendranath¹, John L. Russell², Dmitrii Paniukov²; ¹Texas Tech University, ²Hoververse AI, USA

People often share and interact with touchscreens at public places, like kiosks. Using touchless technology can minimize the spread of germs in such environments. In a touchless interface developed by Hoververse AI, users can interact with a virtual screen projected on tables, walls, etc. based on hand gestures. While using this interface,

users get feedback from their body and environment, to scale their actions to improve their performance. In the current study, right-handed users controlled the movement of a cursor on a projected screen. We investigated participants' accuracy of pointing the cursor precisely at targets present on the screen. On each trial, participants were instructed to follow a 40 BPM metronome to point six times alternating between two target points placed on a vertical, horizontal or diagonal line. The targets were placed either above or below the participants' shoulder level. Some targets were closer to where the participants stood in front of the screen and others were farther to their right. Participants completed a pre-test, calibration, and post-test phase. In the pretest and post-test, participants did not see the cursor, but it was visible during the calibration phase. The results show that pointing accuracy improved in the post-test, compared to the pre-test. This indicates that participants recalibrated their pointing behavior based on visual feedback from the cursor visible during the intervening calibration phase. Accuracy was also found to be better when the targets were closer on the screen and below the shoulder level, as compared to when they were above the shoulder level and farther away. Additionally, when the two targets were placed horizontally, the pointing accuracy was found to be better, as compared to when they were placed vertically. These results have implications for designing touchless interfaces in environments that need precise interactions, like clinics and hospitals.

53.441 SENSING DISTANCE: IMPACTS OF MOTOR LEARNING ON SENSORY DISTANCE PERCEPTION

Kess Folco¹, Hannah Block¹; ¹Indiana University Bloomington

Perceiving distance is a vital component of everyday behavior, involved in actions such as reaching for a water bottle or crossing a busy road. The final neural construct of distance for actions is often an integrated multisensory representation including information from visual, proprioceptive, and motor (e.g. efference copy) systems, weighted by factors such as task goals and information reliability. These systems are also influenced by learning and experience. To test how inter-system transfer and learning changes sensory representation, participants performed target-directed reaches using a KINARM EndPoint robotic manipulandum while tracking eye movements. The task was to reproduce distances provided either visually or proprioceptively before and after adapting to a force perturbation that reduced the amount of force needed to reach the target. Preliminary results (n=7) suggest that assistive force (proprioceptive perturbation) during reaches caused a change in distance representations for both proprioceptive and visual systems. Specifically, proprioceptive distances were perceived as shorter and visual distances were perceived as longer, indicating the multisensory implications of unisensory perturbations. This effect was greater for inter-system distances, where the distance was provided to the hand but reproduced with the eye, or vice versa. In addition, continuous kinematic data from gaze and hand behavior during adaptation suggested this shift in distance perception may be related changes in

gaze-hand coupling. Specifically, gaze typically leads reach trajectories, but following assistive force perturbations, gaze leads reach less. There was also a trend suggesting that people who shifted gaze-hand coupling the most saw more universal reductions in distance perception, regardless of sensory modality, across all types of reproductions. Together, these results emphasize the necessity of considering the impact that inter-system transfer may have on perception, particularly when sensory paradigms require an output behavior (e.g. button press or eye movement), which is then used to interpret the internal system.

53.442 EEG EVIDENCE FOR INCREASED COGNITIVE DEMAND FROM DEFOCUS BLUR

Yannick Sauer^{1,2}, Rajat Agarwala¹, Jenny Loren Witten^{1,2}, Antonia Roth^{1,2}, Patrick Michel², Judith Ungewiss², Siegfried Wahl^{1,2}; ¹University of Tübingen, ²Carl Zeiss Vision International GmbH

Blurry vision from uncorrected refractive errors, like presbyopia, affects the daily lives of billions of people worldwide. Many presbyopes report increased effort when performing visually guided tasks, suggesting an impact of reduced visual performance on cognitive demand. To study the functional impact of optical blur, we investigated how defocus blur affects visuomotor performance, subjective cognitive load, and EEG activity. Twenty-nine participants performed a pegboard task with tweezers under three different conditions: vision was corrected or intentionally blurred with two different levels of defocus lenses. The three-minute task was repeated three times per condition in random order. Cognitive load was assessed both subjectively using the NASA-TLX questionnaire and objectively using EEG. Results show a significant decrease in task performance and a significant increase in subjective cognitive load with increasing levels of defocus blur. EEG analysis revealed condition-dependent changes, specifically a decrease in theta and alpha activity in the occipital lobe and an increase in theta band power in the prefrontal cortex. These findings are consistent with elevated cognitive and attentional demands under reduced visual quality. Our results support the use of cognitive load measures to quantify the functional impact of blurry vision and to evaluate optical correction strategies, for example, different spectacle lens designs.

53.443 TASK-DEPENDENT CORTICAL SENSITIVITY TO DEFOCUS BLUR: AN FNIRS STUDY

Rajat Agarwala¹ (rajat.agarwala@uni-tuebingen.de), Yannick Sauer^{1,2}, Elena Aggus-Vella³, Marco Bragaglia³, Klara Hoxha³, Annalisa Bosco^{3,4}, Patrizia Fattori^{3,4}, Siegfried Wahl^{1,2}; ¹ZEISS Vision Science Lab, Institute for Ophthalmic Research, University of Tübingen, Germany, ²Carl Zeiss Vision International GmbH, Germany, ³Department of Biomedical and Neuromotor Sciences, University of Bologna, Italy, ⁴Alma Mater Research Institute For Human-Centered Artificial Intelligence (Alma Human AI), University of Bologna, Italy

Visual blur degrades sensory input and may increase cognitive effort during tasks that require visuomotor integration, yet the minimum level

of optical blur that measurably alters cortical haemodynamics remains unclear. Functional near-infrared spectroscopy (fNIRS) provides a non-invasive method to quantify changes in cortical oxygenation. This study investigated how induced optical blur modulates cortical activation across various brain regions during visuomotor and control tasks. Six participants performed four task conditions: (1) Visuomotor (VM): picking and placing beads according to a visual pattern; (2) Motor-only: bead placement with minimal visual demand; (3) Visual-only (VS): pattern observation without motor action; and (4) Baseline (BS): central fixation. Blur was induced using electronically tunable lenses set to 0, 1, and 2 diopters. Each trial included lens adjustment (~1 s), a 3-s cue, 10 s of task execution, and a 5-s recovery period. Cortical haemodynamics were recorded using the NIRSport fNIRS system, with optodes covering frontopolar, dorsolateral prefrontal, visual, and parieto-temporal cortices. Data were analyzed using a group-level random-effects GLM. Across task contrasts, VM vs BS produced the strongest and most widespread haemodynamic effects across frontal, parietal, and temporal regions, while VS vs BS showed smaller but reliable effects over left frontal and right parietal areas. VM vs VS yielded minimal differences, indicating highly similar haemodynamic responses between visuomotor and visual-only processing. Blur-response analysis revealed that both VM and VS exhibited significant haemodynamic changes already at 0.5D across Oxy-, Deoxy-, and Total-Hb signals, driven primarily by occipito-parietal visual association cortex, with additional fronto-parietal recruitment in VM. In contrast, baseline fixation showed significant Total-Hb effects only at 2D, indicating greater tolerance to blur. These findings demonstrate that task engagement dramatically lowers the neural blur threshold, with even mild optical degradation rapidly impacting cortical processing during active visual and visuomotor behavior.

This research is supported by the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant PIACES: PLasticity of perception in real and virtual spaCES, agreement No. 101086206.

Action: Perception, recognition

TUESDAY, MAY 19, 8:30 AM – 12:30 PM,
PAVILION

53.444 INFLUENCE OF FORM VS. OPTIC FLOW ON THE RECOGNITION OF NATURALISTIC BODY ACTIONS

Prerana Kumar^{1,2} (prerana.kumar@uni-tuebingen.de), Martin A. Giese¹; ¹Hertie Institute for Clinical Brain Research (HIH), University of Tuebingen, ²International Max Planck Research School for Intelligent Systems, Tuebingen

INTRODUCTION: The computational mechanisms of robust action recognition are not well understood. While it is known that humans can

recognize actions from videos with minimal shape information in individual frames, previous work used highly simplified stimuli. Recent work in computer vision used appearance-free action videos derived from variable, real-world videos by removing shape information from each frame, and showed that humans can categorize these actions when trained on the transformed videos. It remains unclear how well humans can generalize zero-shot to such appearance-free videos and what kind of model can reproduce this generalization. **METHODS:** In an in-lab psychophysical experiment, we investigated zero-shot generalization to appearance-free videos derived from real-world action videos. Participants (N=22) were trained to classify five action categories using naturalistic videos (UCF5 dataset) and tested on two types of appearance-free transformations of the videos: (i) dense-noise motion videos from an existing appearance-free dataset (AFD5), and (ii) random-dot videos with background flickering, generated by us. We modeled participants' behavior with a 3D convolutional neural network architecture with separate form (RGB) and motion/optic-flow pathways (both implemented with X3D backbones). The motion pathway includes an optic-flow estimator and a coherence gate. Our model was trained only on naturalistic UCF5 videos and tested on the appearance-free videos. **RESULTS:** Participants recognized actions in both appearance-free conditions well above chance (>78% accuracy, chance level: 20%), but less accurately than in naturalistic videos. Our model's test accuracy on both appearance-free stimulus types was >60%, outperforming recent models and reducing the gap to human performance. Ablation studies showed that the motion pathway was critical for this generalization, while the form pathway improved performance on naturalistic videos. **CONCLUSION:** Our findings highlight the importance of motion-based representations for explaining the robust generalization to appearance-free videos observed in humans and support multi-stream architectures as a model of action processing.

This work was funded by ERC 2019-SyG-RELEVANCE-856495.

53.445 PREDICTING ACTIONS OF A FEINTING OPPONENT

Anna Filina¹, Maryam Vaziri-Pashkam¹; ¹University of Delaware

In daily interactions, we often have to read others' body movements to predict their goals. In competitive contexts, these bodily cues can be misleading, as opponents may feint to deceive each other. To understand how humans predict actions when deceptive movements are possible, we designed a competitive reaching task. Two participants sat across from each other at a table, and each had a starting point and two targets. One participant (the Attacker) was instructed to hit either the left or right target with their finger, after which their opponent (the Blocker) was instructed to hit the same target as quickly as possible. Motion sensors were attached to participants' index fingers to measure reaction time, accuracy, and movement trajectories. In Deception blocks, Attackers were allowed to deceive the Blockers. In Control blocks, Attackers were instructed to go directly to the target on all trials. Reaction times were faster in Control

compared to Deception blocks, indicating that Attackers successfully misdirected the Blockers. Blockers' accuracies, however, did not differ between conditions, suggesting that Blockers maintained high accuracy despite the misdirection. Next, to investigate how participants read each other's bodily cues in this interactive game, we also measured reaction times when Blockers reacted to a video of a dot representing the Attacker's finger position on a computer screen. Reaction times were faster when participants had access to full-body movements compared to when they responded to the moving dot on the screen, with no difference in the size of this advantage in control compared to deception blocks. These results demonstrate that access to full-body movements is critical for predicting others' goals, regardless of the presence or absence of deception. They provide insights into how humans interpret and predict actions in settings where deceptive movements can occur.

53.446 NEURAL REPRESENTATIONS OF MOVEMENT DIRECTION OF OBSERVED NATURALISTIC ACTIONS: AN FMRI STUDY

Zelal Eltas¹, Murat B. Tunca^{1,2}, Burcu A. Urgan¹; ¹Bilkent University, ²Justus Liebig University Giessen

Perceiving the movement direction of observed actions is essential for interpreting intentions and guiding social interactions. Direction selectivity has been widely studied in motion perception literature using simple stimuli like dots, gratings, or point-light displays (PLDs). However, little is known about how the brain encodes direction during the natural, repetitive actions seen in daily life. To investigate this, a set of novel video stimuli was created using four classes of action: (i) object manipulation (stretching, shaking), (ii) object manipulation using a tool (wiping, painting), (iii) self-directed actions (scratching, rubbing), and (iv) self-directed actions using a tool (brushing, drying). Each action was recorded in three spatial directions (left-right, up-down, and front-back) with two versions and two actors, resulting in a total of 96 naturalistic action videos. 25 participants took part in an event-related fMRI experiment while they observed these action videos and made a judgment about movement direction. The brain activity was analyzed using multivariate pattern analysis (MVPA) and multiple regression representational similarity analysis (RSA). MVPA results suggested that action direction could be classified above chance levels in occipital, parietal, and motor cortices. Specifically, the highest decoding accuracy was observed in occipital, primary motor, and somatosensory regions. Importantly, RSA results demonstrated that direction information is still represented in early visual, occipito-temporal, parietal, and motor regions after controlling for low-level visual and motor features. These findings indicate that the movement direction of actions is represented at multiple levels of the action observation network (AON), extending from early sensory regions to higher-order parietal and frontal areas. More importantly, this study suggests that direction selectivity is a core feature of the action observation network, linking the processing of basic motion with the understanding of everyday actions.

This project is funded by Turkish Health Institutes Presidency (TUSEB) A2 Project Supporting Program (Project No: 37734).

53.447 MODELING TYPICALITY IN HUMAN ACTION PERCEPTION WITH CLIP REPRESENTATIONS

Filip Durovic¹ (fld23snj@bangor.ac.uk), Marieke Mur², Angelika Lingnau³, Paul E. Downing¹; ¹Bangor University, ²Western University, ³University of Regensburg

Large-scale multimodal models such as CLIP show strong alignment with high-level visual cortex, but it remains unclear whether their representational spaces also capture behavioral aspects of human vision. Here we ask whether CLIP embeddings can serve as a model of typicality in human action categorization. We measured human reaction times in an action categorization task: participants viewed an image depicting one of five actions (biking, cooking, dressing, eating, golf) for 300 ms followed by a 200 ms mask, then judged whether a concurrently presented word label (e.g., “biking”) matched the image. We extracted representations from CLIP’s final 1024-dimensional embedding layer (ResNet-50; trained on 400M image–text pairs) and computed distance-to-norm measures in this space, defined as the distance of each image embedding to either the mean image embedding of an action category (image-norm) or the single embedding of a category label (label-norm). Both measures reliably predicted response times: images closer to their category norm yielded faster “same” judgments, and images closer to norms of nonmatching categories yielded slower “different” judgments. Label-norm distances better explained “same” trials, whereas image-norm distances better explained “different” trials, suggesting distinct contributions of semantic and visual similarity. Comparing model variants revealed that a smaller CLIP model and an image-only SimCLR model, sharing the same visual backbone and trained on the same 15M-image subset (CLIP with image–text contrastive learning, SimCLR with image-only contrastive learning), both captured trial-level structure, but their predictive power was weaker. These findings indicate that CLIP’s joint vision–language representations provide a robust account of exemplar typicality in human action perception, linking large-scale visual-language models to fine-grained human behavioral variability.

This work was supported by the Economic and Social Research Council (ESRC), UK Research and Innovation (UKRI).

53.448 MOTOR-VISUAL PRIMING: DOES TRACKING A TRAJECTORY BY HAND FACILITATE DISCERNING IT BY EYE?

Zekun Sun¹ (zekun.sun@yale.edu), Samuel D. McDougle^{1,2}; ¹Department of Psychology, Yale University, ²Wu Tsai Institute, Yale University

Perception often precedes action — we see, we think, and then we act. Yet the reverse can also be true: Action may enhance what we perceive. For example, children learn about the visual world by moving their bodies and interacting with their surroundings; and, when we

touch an object, haptic input can activate a related visual image. If action scaffolds perception in a more direct way, then performing a dynamic movement might activate the visual structure it implies. Here, we ask whether a temporally unfolding motor trajectory can activate a corresponding spatial representation in vision, thereby improving recognition of the motor-implied pattern. In our experiment, participants first moved their hand to track a disc that moves along the trajectory of a novel character. Immediately afterward, they viewed a static noise image and indicated via key press whether it contained only noise or a hidden character. Tracking performance was quantified using the Procrustes distance between participants’ hand movements and the target’s trajectory. Across 43 participants, tracking performance significantly predicted recognition accuracy ($p < .01$): Participants who more precisely reproduced the trajectory were more likely to detect the character within noise. This relationship appeared to vanish when participants tracked one character but attempted to recognize a different character, ruling out nonspecific performance factors. These findings point to a novel “motor-visual priming” effect, in which dynamic movements may facilitate subsequent visual detection by activating a shared, cross-modal spatial representation.

This work was supported by grant R01 NS134754 (S.D.M.) from the National Institutes of Health

53.449 POFFENBERGER WITH AN UNEXPECTED TWIST

Betül Beyza Cengil^{1,2}, Huseyin Boyaci^{1,3}; ¹Bilkent University, ²Social Sciences University of Ankara, ³Justus Liebig University Giessen

In this study, we employ the century-old Poffenberger paradigm to investigate the effects of expectation on visuospatial interactions. In the Poffenberger paradigm, participants respond to a constant and lateralized visual stimulus using either their ipsilateral or contralateral hand, and their response times (RTs) are recorded. The differences in RTs between ipsilateral and contralateral trials are attributed to the cost of interhemispheric transfer based on the shared contralaterality principle in the visual and motor systems: ipsilateral responses require uncrossed visuospatial interactions, whereas contralateral responses require crossed visuospatial interactions. In our design, participants ($n = 20$) responded to briefly presented lateralized visual stimuli with their ipsilateral or contralateral hand depending on the stimulus characteristics (e.g., an X or an I). Furthermore, prior to the target stimulus, we provided a centrally presented cue indicating which visual stimulus was more likely to appear in that trial with a certain level of congruency (e.g., 100%, 80% or 50%). Whereas the side of the presentation was randomized. Thus, we manipulated the expectations about ipsilateral and contralateral responses irrespective of the visual hemifield and hand. Results showed a main effect of expectation. Post hoc analyses revealed that participants were statistically significantly faster in congruent trials (RT = 529 ms) compared to incongruent trials (RT = 579 ms) and no-expectation trials (RT = 547 ms), and slower in incongruent trials compared to no-expectation trials. There was also an interaction between congruency (congruent, incongruent) and response type (ipsilateral, contralateral): violations of contralateral

response expectations increased the RTs more than violations of ipsilateral response expectations. Control experiments demonstrated that the effect could not be attributed solely to visual processing of the stimulus. These results show that expectations can affect the temporal dynamics of visuomotor interactions in speeded tasks.

53.450 PRIMED FOR ACTION: HOW HUMANS DECODE ACTIONS FROM ALMOST NOTHING

Filip Rybansky¹ (filip@rybansky.com), Sadegh Rahmani², Andrew Gilbert², Frank Guerin², Anya Hurlbert¹, Quoc Vuong¹; ¹Newcastle University, ²University of Surrey

Why do humans find actions easy to recognise compared to state-of-the-art machines? While machines struggle with naturalistic action videos, people can recognize them from minimal amounts of spatial and spatiotemporal features. We previously identified Minimal Recognisable Configurations (MIRCs) in Epic-Kitchens-100 videos which were correctly (Easy) or incorrectly (Hard) categorised by our computer-vision network. High-level (e.g., active object, active hand) and low-level (e.g., orientation, motion) critical features for action recognition were determined from these MIRCs. Here we used eye-tracking to converge on critical features and establish whether people's eye-movements can explain why machines found some actions hard. We further tested whether contextual top-down information can affect people's performance and eye movements, as machines predominantly use bottom-up information. Participants (N=36) were briefly presented with a static prime image (400ms), then viewed the original or MIRC video and identified the action while eye movements were recorded. Primes were the middle frame of the original video. High-level features were further removed, or the prime was phase-scrambled to progressively reduce contextual information. Normalised Scanpath Saliency (NSS) was used to compare which features best predicted where participants looked. Notably, active objects performed significantly better than active hands (NSS=1.60vs0.75) and, for low-level features, orientation performed better than motion (NSS=1.16vs0.53). The predictive ability was higher for all features in Hard than Easy videos (NSS=0.69vs0.51). For priming, unedited primes produced higher accuracy and shorter First Fixation Durations than phase-scrambled primes when viewing MIRCs (ACC=0.78vs0.57, FFD=330vs299ms). Our results suggest that egocentric action recognition is dominated by the active object, and by spatial over motion features. Overall, participants were more likely to fixate on the critical features for Hard videos but could recognise Easy videos even by fixating regions with lower feature activation. Human strategies such as increased attention to critical features can help improve action recognition by machine in difficult conditions.

Leverhulme Trust

53.451 SPATIAL WORKING MEMORY SUPPORTS INTEGRATION OF TIMING INFORMATION ACROSS ACTION AND PERCEPTION

Tri Nguyen¹, W. Ryan Waite¹, Joo-Hyun Song¹; ¹Brown University

Our recent work (Guo et al., 2025) showed that participants who practiced a virtual ball-throwing task developed a stable motor timing that selectively improved sensitivity to matching auditory temporal intervals. This motor-perceptual link appears to emerge from learned temporal structure, but the cognitive resources supporting it remain unexplored. Based on evidence showing that individual differences in sensorimotor skill acquisition are related to visuospatial working memory capacity, the present study examined whether spatial working memory contributes to this motor-perceptual link. Over two days, participants practiced the virtual ball-throwing task and, between throws, performed one of three interleaved tasks: an auditory interval discrimination task in which the intervals matched the timing structure of the motor task (N = 17), an auditory pitch discrimination task that introduced motor-irrelevant information (N = 17), and a control task with minimal cognitive load (N = 35). Replicating prior work, improvement in temporal sensitivity was greatest in the interval group, where perceptual intervals matched the timing characteristics of the practiced movement. Crucially, to examine the role of working memory, we used participants' spatial working memory capacity to predict motor and perceptual performance. Working memory capacity was most predictive of both motor and perceptual performance in the interval group, especially during the early learning phase. The influence of working memory was less pronounced in the pitch group, where the perceptual information was irrelevant to participants' emerging motor timing, and in the control group, where cognitive load was low. These findings suggest that spatial working memory supports the maintenance and integration of spatiotemporal information across motor and perceptual systems during motor learning, contributing to changes in perceptual sensitivity that reflect the temporal structure of practiced actions.

53.452 WHAT MAKES GAMEPLAY VIDEOS ENJOYABLE TO WATCH?

Kristine Zheng¹ (zhengkristine@gmail.com), Jean-Peic Chou¹, Junyi Chu¹, Judith Fan¹; ¹Stanford University

Many people choose to spend time watching other people play video games. What makes some gameplay more enjoyable to watch than other gameplay, even in the context of the same game? Here we explored features of gameplay that might influence observers' self-reported enjoyment, especially factors related to the apparent skill level of the game player. We developed a variant of the Flappy Bird game, in which a player traverses an environment from left to right while trying to minimize the number of collisions with vertical pipes of variable length extending from the top and bottom of the screen. We trained four artificial agents that exhibited different "playstyles," varying in how close these player agents passed by these pipes and the velocity with which they approached them. One agent consistently traveled just beneath the upper pipes ("top-skimming"); one traveled just above the lower pipes ("bottom-skimming"); one avoided both sets of pipes with a large margin ("centered"); and one agent followed an undulating, swooping trajectory that made it approach obstacles at

high velocity ("swing"). Participants (n=80) watched one video clip of each agent navigating a single course, then rated how enjoyable it was to watch. The number of collisions sustained by agents was variable (i.e., between 0-3 per course), but collision frequency was held constant across clips viewed by each participant to isolate the effect of playstyle. We found that enjoyment ratings assigned to these agents were reliably different ($\chi^2(3)=16.55$, $p<.001$; $d\approx 0.05-0.29$), with "swing" being most enjoyable, then "centered", "top-skimming", and "bottom-skimming." These preliminary findings suggest that what observers find enjoyable in gameplay goes beyond the outcome in each clip (e.g., number of collisions), but also encompasses the way in which a player achieved that outcome.

This work was supported by NSF CAREER Award #2436199, NSF DRL #2400471, and awards from the Stanford Human-Centered AI Institute (HAI), Stanford Accelerator for Learning, and the Brown Institute for Media Innovation.

Multisensory Processing: Cross-modal interactions

TUESDAY, MAY 19, 8:30 AM – 12:30 PM,

PAVILION

53.453 CONFIDENCE SHAPES MULTISENSORY INTEGRATION

Jason Samaha¹, Wei Dou¹, April Pilipenko¹; ¹University of California, Santa Cruz

The integration of cues from different sensory modalities is thought to closely follow the principles of reliability-weighting, whereby multisensory estimates reflect a combination of unisensory estimates weighted in proportion to their sensory reliabilities. Traditionally, the reliability of a sensory modality is estimated from unisensory discrimination performance, reflecting the assumption that the relevant representation of reliability is that which governs objective performance. However, recent work suggests that subjective confidence may influence reliability weighting beyond performance. Experiment 1 was a purely visual experiment in which participants discriminated the location (left or right) of the mean of a set of dots generated from Gaussian distributions with variable means and (horizontal) standard deviations (SD). We found that, despite matching performance between high and low SD stimuli, participants reported higher confidence when localizing high SD stimuli, consistent with previous findings of overconfidence for high variance stimuli and demonstrating that visual localization confidence can be dissociated from objective performance. Experiment 2 measured unisensory and multisensory confidence and performance using the same dot stimuli and white noise auditory bursts presented from varying virtual sound locations. We found that when auditory and visual unisensory performance was matched, participants reported higher confidence in

visual localization and weighted vision more than objective reliability weighting (i.e., maximum-likelihood estimation; MLE) would predict during multisensory localization. Additionally, high and low SD visual stimuli were not differentially weighted as MLE would predict, reflecting a possible over- or under-weighting of high and low SD stimuli, respectively, in line with confidence. Experiment 3 minimized the audiovisual spatial disparity to maximize integration but still led to an over-weighting of vision in general and of the high SD stimuli, both of which were associated with higher confidence. Collectively, our studies suggest that subjective confidence contributes to the estimate of reliability which is used to weight audiovisual integration.

53.454 AUDIOVISUAL INTEGRATION DURING MOVIE WATCHING IN THE INFANT AND ADULT BRAIN

Yuna Kwak¹ (yuna.kwak@yale.edu), Lillian Behm¹, Juliana E. Trach¹, David J. Lewkowicz¹, Nicholas B. Turk-Browne¹; ¹Yale University

Adults show clear behavioral facilitation for multisensory stimuli (e.g., faster and more accurate detection) relative to unisensory stimuli, supported by brain regions such as the superior temporal sulcus (STS). Although similar behavior emerges in infancy, how multisensory integration is supported across the infant brain is unknown. In an ongoing study, we are using fMRI to characterize audiovisual binding in awake infants (3-24 months) during a naturalistic movie paradigm and to compare their neural responses to those of adults. We played three versions of the same movie clip to each participant in both age groups: audiovisual, video-only, and audio-only. Our main analysis relies on intersubject correlation (ISC), which identifies brain regions with reliable neural responses to the stimulus in each condition by quantifying the similarity of fMRI time-series across individuals. In adults, ISC was higher (1) for video-only and audio-only conditions in early visual and auditory cortex, respectively, relative to other brain regions and (2) in audiovisual versus unisensory conditions — evidence of multisensory facilitation — in known multisensory areas such as STS and the temporo-parieto-occipital junction, as well as in the medial temporal lobe. Decoding analyses support these findings and further suggest that the observed facilitation effect reflects the fusion of unisensory inputs rather than a summation of independent signals. Preliminary results from infants show initial evidence of (1) unisensory coding in early visual and auditory cortex and (2) audiovisual facilitation across the neural hierarchy, as in adults. The full sample will be needed to identify the specific regions involved and to understand how closely infant processing aligns with adults. Overall, these findings indicate that a wide network of regions supports multisensory integration in adults and suggest that this network begins emerging early in development.

53.455 TITRATED AUDIOVISUAL CROSSMODAL CONGRUENCIES

Karen L. Gunther¹, Brandon E. Hammer¹; ¹Wabash College

We have conducted a titrated percent congruent speeded classification task with visual and auditory stimuli to test the percent

congruency at which the congruency effect flips to faster response times on incongruent stimuli [Chiou & Rich (2012) found this flip with 20% of the trials in a block being congruent, 80% incongruent], and if the strength of the congruency effect is higher when subjects respond to auditory stimuli, consistent with visual dominance (the effect of vision having a particularly strong influence on other senses; e.g., Posner, Nissen & Klein, 1976). We presented subjects with white visual dots on a black background, appearing either above or below fixation, simultaneously with either high or low auditory pitches (1500 and 300 Hz, respectively). High visual dots and pitches, or low of both, are congruent trials; high of one paired with low of the other are incongruent. We presented blocks of trials with 96% congruency, down to 4% congruency, and asked subjects to respond to either the visual stimuli or the auditory stimuli in different blocks of trials. The initial six subjects (of 20 planned) mostly support our hypotheses. At the lower percentages of congruent trials, the congruence score [(incongruent reaction time) – (congruent reaction time)] is lower, reflecting relatively slower reaction times on the congruent trials and thus a weaker congruency effect, although we are not currently observing a strong flip where reaction times to incongruent stimuli pairings are faster in majority-incongruent blocks. Subjects do have higher congruence scores (longer reaction times on incongruent stimuli) when responding to the auditory stimuli, supporting the presence of visual dominance. These results suggest that the proportion of congruent trials can mediate the magnitude and direction of crossmodal congruency effects.

The Eldon Parks Memorial Fund supported the second author's summer internship.

53.456 CENTER-OF-MASS MEDIATES SEX DIFFERENCES IN THE BODY TILT ILLUSION

Sophia R. Baia¹ (srbaia@asu.edu), Ashlyn Overby¹, Michael K. McBeath^{1,2}; ¹Arizona State University, ²Max Planck Institute for Empirical Aesthetics

Introduction: The body tilt illusion is the exaggerated perception of body tilt angle. People tend to overestimate that they are tilted at 45° when only tilted about 30°. Past research confirms this illusion increases with lack of visual and auditory cues and interacts with gender. The present study examines distortion in body tilt perception as a mediating function of people's center-of-mass to see if it might account for the sex difference. This follows the finding of women generally having a lower center-of-mass compared to men (Eler, 2005). **Method:** We constructed a see-saw apparatus to measure body center-of-mass point as well as height, weight, and naval location. We used a human-sized 3D gimbal to measure the body tilt illusion, tilting forward, backward, and to the sides, both with eyes open and closed. 10 men and 14 women were tilted in each condition, estimating when they reached 45°, without feedback. **Results:** The results replicated previous findings with a large overall body tilt illusion for both sexes, a larger tilt illusion to the side compared to forward-backward, and a large increase with eyes closed compared to open. Additionally, we

found a non-significant trend for women to exhibit a larger tilt illusion, which was better accounted for by the mediating variable of center-of-mass. **Discussion:** The findings further support that perception of body tilt is directional, multisensory and consistent with the illusion being functional for helping people keep their balance. In addition, the lack of a gender effect, but trending mediating effect of center-of-mass supports that this is an underlying mediating physical characteristic. The overall pattern of findings supports that the body tilt illusion is relatively universal, likely functional, and that past differences attributed to gender may be more parsimoniously accounted for by individual body shape differences related to physical balance like center-of-mass.

53.457 CAMERA MOTION IMPACTS CYBERSICKNESS IN AUGMENTED REALITY

Henry Kam¹ (henry.j.kam@gmail.com), Niall Williams¹, Jeffrey Lubin², James Bergen², Mikhail Sizintsev², Qi Sun¹, Colin Groth¹; ¹New York, ²SRI International

Cybersickness is a challenge that persists in wide field of view, near-eye displays, particularly when visual motion conflicts with vestibular signals. We analyzed how rotational camera movements contribute to cybersickness in optical see-through augmented reality (AR). Participants wore a Magic Leap 2 headset while viewing a head-locked rectangular viewport showing a virtual box-room with paintings on the walls. The virtual camera oscillated side to side while participants remained physically stationary, creating strong visual-vestibular conflict. Across three within-participant sessions, we tested the effect of different oscillatory rotation speeds (20 degrees, 30 degrees, 40 degrees per second) on cybersickness in AR. The highest speed was chosen as the average maximum level at which an adult is still able to perform smooth gaze pursuit. For each 45-second trial, participants completed a gaze-maintenance task to ensure consistent exposure to the optical flow: up to five red and blue paintings occasionally appeared and disappeared on the room walls. Participants kept their heads still while tracking the number of each colored painting over time. After each trial, users reported a Fast Motion Sickness (FMS) rating, and Simulator Sickness Questionnaire (SSQ) scores were collected before and after each 15-minute session to quantify cybersickness change. A linear fit between rotational speed and SSQ change captured a weak linear dependence of cybersickness on visual motion parameters. As camera speed increased, more visual-vestibular conflict was induced ($r^2 = 0.50$). These results provide a compact perceptual model linking measurable properties of camera movement to cybersickness in AR. As a result, AR scene design can be quantified and monitored to create visually stable, and comfort-optimized AR experiences.

DARPA ICS: Intrinsic Cognitive Security

53.458 PUPILLOMETRY REVEALS A PROCESSING EFFICIENCY ADVANTAGE FOR MEMORABLE VOICES

Cambria Revsine¹, Wilma A. Bainbridge¹; ¹University of Chicago

Pupillometry can provide insight into participants' underlying cognitive processes, such as effort and attention, while performing a task. However, when used in visual tasks, inherent confounds such as luminance levels complicate conclusions drawn with this measure. Here, we used pupillometry to measure cognitive effort in a purely auditory task, to test whether there is a processing advantage for memorable voices, as has been previously shown in visual stimuli. Participants verbally recalled sentences spoken by speakers with a range of intrinsic memorability scores, as determined in a prior study (Revsine et al., 2025). We found that more memorable speakers elicited smaller pupillary dilations during the stimulus and recall periods, suggesting that they are easier to process. Speaker memorability was also positively predictive of recall accuracy of the sentences. To further investigate these effects, we ran a follow-up experiment containing stimuli of speakers from the extremes of the memorability distribution. While the main effect of voice memorability on memory accuracy failed to replicate, we found an interactive effect between memorability and participants' overall performance—in other words, speaker memorability boosted recall for participants with poorer memory. Memorability was again negatively associated with pupil size, in the worst performers. This result suggests that in these participants, more efficient processing of memorable voices may free up cognitive resources for remembering content, yet this advantage is less apparent in better performers whose memory abilities may already be at ceiling. Our findings add evidence to a growing account of stimulus memorability as reflecting perceptual processing efficiency, across sensory modalities.

53.459 EXTENDING MEMORABILITY BEYOND VISION: EXAMINING ODOR MEMORABILITY AND THE ROLES OF PLEASANTNESS, INTENSITY, AND INDIVIDUAL DIFFERENCES

Vicky Zhang¹, Eli Wakefield¹, Leslie Kay¹, Wilma Bainbridge¹;
¹University of Chicago

Memorability, or the likelihood that a stimulus will be remembered, is an intrinsic stimulus property that shows strong consistency across observers. This consistency has been demonstrated across diverse stimulus types, with people reliably remembering and forgetting the same faces (Bainbridge et al., 2013), objects (Kramer et al., 2023), words (Xie et al., 2020), and voices (Revsine et al., 2025). However, memorability research has focused almost exclusively on visual and auditory modalities, leaving it unclear whether such intrinsic consistency extends to olfaction, whose direct pathway from the olfactory bulb to hippocampal memory circuits may offer unique insight into the broader mechanisms underlying memorability across modalities (Aggleton & Mishkin, 1986; Herz & Schooler, 2004). Using twelve intensity-matched monomolecular odorants, we investigated whether people reliably remember and forget the same odors and whether perceptual factors like pleasantness influence odor memorability. Adult participants smelled each odor during incidental encoding, completed a continuous recognition task, and rated

pleasantness, intensity, and confidence. Correct recognition scores varied widely across items (approximately 0.18–0.91), with some odors reliably remembered and others consistently forgotten, revealing robust stimulus specific and observer consistent odor memorability that parallels intrinsic memorability effects documented in vision and audition. We next examined perceptual predictors of these item-level differences. Pleasantness showed negative but non significant relationships with both corrected recognition and confidence. Intensity, while only weakly related to corrected recognition, showed a significant positive association with recognition confidence, suggesting that perceptual salience increases metacognitive certainty without proportionally improving objective memory. Together, these findings demonstrate the intrinsic memorability of odors and show that neither pleasantness nor intensity alone accounts for these effects. This work extends the concept of memorability to olfaction, revealing both cross-modal similarities and differences, and mirroring visual findings showing that stimulus aesthetics do not reliably predict memorability (Isola et al., 2011).

Face and Body Perception: Development, clinical

TUESDAY, MAY 19, 8:30 AM – 12:30 PM,
PAVILION

53.460 UNCOVERING NEURAL MECHANISMS OF FACE PROCESSING IN AUTISM USING EEG AND "BUBBLES"

Allyson Bastien¹, Channele Demeule¹, Anthony Proulx¹, Caroline Blais¹, Daniel Fiset¹; ¹Université du Québec en Outaouais

Autism is a neurodevelopmental condition marked by difficulties in social communication and interaction (American Psychiatric Association, 2022). Face processing, in particular, is often impaired in Autistic individuals (Griffin et al., 2021). For instance, Autistic adults make limited use of the eye region and rely more on the mouth than neurotypical adults (Spezio et al., 2007). However, the neural mechanisms underlying these differences remain unclear. The present study sought to compare the brain's electrophysiological response to faces in Autistic (n = 2 men) and neurotypical (n = 29; 12 men) adults. Brain activity was recorded using EEG while participants performed 2000 trials of a categorisation task (smiling or not smiling). Faces were partially revealed using the "Bubbles" method (randomly located Gaussian windows; Gosselin & Schyns, 2001). Classification images were used to reveal the pixels associated with EEG activity at electrode PO8 from -200 to 600 ms relative to stimulus presentation. Autistic and neurotypical subjects showed similar response patterns, both relying on the mouth to complete the task. In neurotypical subjects, a pixel test ($p < .05$; Stat4Ci Toolbox; Chauvin et al., 2005) revealed a significant association between N170 amplitude and the contralateral eye, the ipsilateral eye and the mouth. Findings revealed a greater complexity among Autistic participants. Participant A

demonstrated a weak reaction to the eye region, despite a N170 latency similar to neurotypical subjects. In contrast, participant B exhibited a strong and particularly sustained reaction to the eyes and mouth, while showing a substantially delayed N170 relative to neurotypical participants. Increasing sample size would enable the identification of patterns, thereby facilitating a more precise understanding of the neural mechanisms underlying social difficulties in autism. These results demonstrate how the integration of two rigorous methods, EEG and "Bubbles", allows a deeper understanding of face processing specificities in autism.

The present study was supported by the Natural Sciences and Engineering Research Council of Canada to Daniel Fiset (#RGPIN-2022-04350)

53.461 PREDICTING FACE- AND OBJECT-SELECTIVE TOPOGRAPHIES IN AUTISM USING HYPERALIGNMENT

Ian Abenes¹, Yue Wang², Runnan Cao², Anila D'Mello^{1,3,4}, Shuo Wang², Jiahui Guo¹; ¹The University of Texas at Dallas, ²Washington University in St. Louis, ³The University of Texas Southwestern Medical Center, ⁴Peter O'Donnell Jr Brain Institute

Autism is a neurodevelopmental disorder associated with impairments in social interaction and communication; however, it is also accompanied by deficits in various perceptual processes, concomitant with deficits in category-selective cortex (Humphreys et al., 2008, Weisberg et al., 2012). Traditional functional localizers suffer from inefficiency, limited ecological validity, and are not patient- or kid-friendly enough. To explore advanced alternative localization paradigms, we used the hyperalignment algorithm, which functionally aligns individuals in a high-dimensional information space, to investigate whether their individualized category-selective topography can be estimated with an independent scan and from an independent group of participants. Two runs of dynamic functional localizer data from 24 autistic individuals and 24 controls were included, which incorporated stimuli of faces, objects, and scrambled objects (Wang et al., 2024). We hyperaligned participants using their connectivity (connectivity hyperalignment, CHA) or response profiles (response hyperalignment, RHA) with one run of the localizer scan and applied the derived transformation matrices to the other run. Analysis was completed using a cross-validation procedure. We estimated functional topographies using the hyperalignment predicted localizer runs (RHA/CHA) and the anatomically aligned localizer runs (AA), and correlated them with topographies based on their own localizer runs to assess prediction performance. For face- and object-selective topographies, both RHA and CHA outperformed AA. In a searchlight analysis with 15 mm radii, we found a general improvement of RHA/CHA over AA across the whole cortex, with the most noticeable improvement in face- and object-selective cortex for their respective category. Importantly, we demonstrated that cross-group predictions (i.e., using control participants' data to predict autism patients' topographies, and vice versa) produced estimates similar to within-group predictions. We demonstrate that high-fidelity estimation of

individualized functional topographies can be achieved using hyperalignment in autism, offering a new approach for individualized functional localization in this clinical population.

53.462 COMPARATIVE SEVERITY OF FACIAL RECOGNITION DEFICITS IN NON-PROLIFERATIVE DIABETIC RETINOPATHY

E. K. Stacy¹ (emma.k.stacy-1@ou.edu), S. F. Newbolds¹, V. A. Shah², I.M. Sluch³, M. J. Wenger¹; ¹University of Oklahoma, ²Retina Consultants of Oklahoma, ³US Veterans Administration

Diabetic retinopathy (DR) is a disease common in patients with Type 2 diabetes that affects retinal vasculature. Patients with DR report feeling socially isolated due to struggles with several vision-related activities, including recognizing faces. Despite this, there is little known about the prevalence and severity of facial recognition deficits in DR. The current study investigated objective measures of facial perception and memory in patients with moderate to severe non-proliferative diabetic retinopathy, with and without macular edema, along with control participants with healthy vision. Participants were recruited from the Oklahoma City metropolitan area and divided into three groups: two patient groups with moderate to severe non-proliferative diabetic retinopathy—one with macular edema and the other without—and a control group composed of age, race, and sex-matched individuals with normal or corrected-to-normal vision. Each participant completed a contrast sensitivity assessment, the Cambridge Face Memory Test (CFMT), and the Vanderbilt Holistic Face Processing Test (VHPT). Results indicated that average contrast detection thresholds were higher in participants with DR relative to controls. Additionally, the DR groups demonstrated lower CFMT and VHPT accuracy compared to healthy controls, while demonstrating higher accuracy in congruent relative to incongruent trials. Although contrast detection thresholds were negatively related to accuracy in both the CFMT and the VHPT, neither relationship was significant. This project represents a unique attempt to characterize the prevalence and severity of facial recognition deficits in DR, to further inform potential behavioral and medical treatment strategies.

53.463 USING SYSTEMS FACTORIAL THEORY AND GENERAL RECOGNITION THEORY TO EXAMINE FACIAL PERCEPTION IN DIABETIC RETINOPATHY

Sarah F. Newbolds¹, Vinay A. Shah², Ilya M. Sluch³, Michael J. Wenger¹; ¹University of Oklahoma, ²Retina Consultants of Oklahoma, ³US Veterans Administration

Diabetic retinopathy (DR) is a complication of diabetes caused by chronic hyperglycemia and is a leading cause of blindness in the working population. In 2020, DR was estimated to affect over 103 million adults worldwide, and the prevalence is expected to increase. DR is associated with vision issues that can seriously affect daily life, including problems with reading, working, and driving. Although not yet studied quantitatively, many subjective reports have indicated that people with DR experience serious difficulty with facial recognition.

Due to the burden that this symptom presents for those with DR, quantitative investigation of this topic is of particular importance. In the present study, we apply general recognition theory (GRT) and systems factorial theory (SFT)—mathematically rigorous frameworks for analyzing perceptual representation and cognitive processing—to performance on facial recognition tasks involving the canonical dimensions of featural and configural information. Using data from affected and control participants drawn from a large-scale study on DR, we examine how these two dimensions are perceptually encoded and processed in DR. Results from the GRT data indicate violations of perceptual separability and demonstrate individual differences in perceptual independence and decisional separability in those with DR. Results from the SFT data indicate serial exhaustive limited capacity processing in individuals with DR. This study is the first to apply formal perceptual and processing models to facial processing in DR. As such, this study allows unique and novel insights into the perceptual and cognitive impact of DR on a function that is foundational to a range of cognitive tasks and social interactions.

53.464 FACES OF PROSOPOMETAMORPHOPSIA: A COLLECTION OF VISUALIZATIONS OF FACE DISTORTIONS

Brad Duchaine¹, Antônio Mello¹, Sydney Fortner², Alexis Kidder^{1,3}; ¹Dartmouth College, ²University of California, Santa Barbara, ³National Institute of Mental Health

Prosopometamorphopsia (PMO) is a visual perceptual disorder in which faces appear distorted. Fewer than 100 cases of PMO have been reported in the literature, and the great majority are brief, single-case reports. However, because our laboratory has been contacted by more than 250 people who have experienced face distortions, we have conducted in-depth investigations of PMO as a means to better understand typical face perception and have also evaluated interventions that reduce distortions. Here, we present a collection of visualizations that illustrate the face distortions perceived by 15 individuals with PMO, along with synopses of their case histories. Our aim in displaying these visualizations is to provide vision scientists and healthcare professionals with illustrations that convey the nature and severity of the distortions present in PMO. Distortions affecting face shape have received the most attention in the literature, but the distortions in our participants often alter other aspects of facial appearance as well, including the texture, color, and depth of facial surfaces. In some cases, facial appearance is transformed so intensely that participants have great difficulty making judgments about fundamental characteristics of faces such as age, sex, gaze direction, and race. While the variation between cases means that tasks must often be tailored to particular participants, these distinct forms create opportunities to investigate numerous theoretical questions from visual neuroscience. Reports provided by many participants indicate that their verbal descriptions of face distortions were dismissed or misinterpreted in clinical settings, with some individuals even being prescribed antipsychotic medication under the

assumption that their perceptual experiences reflected broader psychiatric pathology. These difficulties highlight the need for greater clinical awareness of PMO, and we hope that disseminating these visualizations will contribute to more accurate diagnosis and improved care.

Hitchcock Foundation award to BD

53.465 NEURAL REPRESENTATIONAL ALIGNMENT IN DEVELOPMENTAL PROSOPAGNOSIA DIFFERS BEYOND VISUAL REGIONS DURING NATURALISTIC VISUAL PROCESSING

Alexis Kidder^{1,2} (alexis.kidder.gr@dartmouth.edu), Clara Sava-Segal¹, Emily Finn¹, Brad Duchaine¹, Chris Baker²; ¹Dartmouth College, ²National Institute of Mental Health

Although developmental prosopagnosia (DP) is defined by difficulty recognizing faces, the severity and specificity of this deficit varies widely, with some DPs also reporting atypical perception to other stimuli. The physiological correlates underlying this variability are largely unknown and are typically investigated by averaging univariate responses to stimulus categories within visual regions. To address these methodological constraints, we compared whole-brain neural alignment during movie-watching using pairwise inter-subject correlations (ISC). We compared inter- and intra-group neural variability in DPs (N=28) and controls (N=45) using an open fMRI dataset where participants watched *Game of Thrones* (GoT; Noad et al., 2024). We found that controls were more aligned than DPs in several regions within and outside visual regions, such as right early visual areas and bilateral superior temporal sulcus (STS). DPs had stronger alignment in left early visual areas. When separating participants within each group who were familiar and unfamiliar with GoT, we replicated that familiar controls had greater alignment in right STS compared to familiar DPs (Noad et al., 2024), and showed familiar DPs had greater alignment in bilateral early visual areas. Familiar controls (N=22) had the most whole-brain alignment, followed by unfamiliar controls (N=23), familiar DPs (N=13) and unfamiliar DPs (N=15). To expand on these findings, we used pairwise pattern ISC to capture changes in multivariate patterns of neural activity at every timepoint. Neural alignment dynamically changed during perception, and results indicate peak differences between groups were driven by faces conveying social information. Altogether, DPs had greater whole-brain neural variability during visual processing compared to controls. Within DPs, low-level visual regions were most similar, suggesting that increased neural variability may occur in later stages of visual processing. Future work will examine the relationship between ISC and behavioral performance to link neural variability with DP perceptual profiles.

53.466 THE TEXT IS ALL IN THEIR HANDS: A NEUROFUNCTIONAL MODEL FOR LIMB-TO-TEXT (VWFA) CORTICAL RECYCLING

Sharon Gilaie-Dotan¹ (shagido@gmail.com); ¹Bar Ilan University

In recent years Dehaene, Cohen and colleagues proposed that the VWFA recycled evolutionarily pre-existing neural circuits in ventral visual cortex, likely those with visual preferences for T or Y junctions useful for letter recognition. While additional propositions were put forward, it was recently shown by Nordt and colleagues (2021) that the VWFA is built upon recycled prereading limb-selective neural substrates. Here I propose that prereading foveal visual expertise for limb movements (likely essential for motor-action expertise development) substantiates a representation sufficient to allow for efficient reading to successfully develop. The prereading neural limb representation consists of 3 distinct (yet likely connected) components that are activated when we see another person: the hand or whole-body stick figure (HSF/WBSF), its “outfit” (skin, cloths, etc.), and its kinematics. The HSF/WBSF aspect is the only one clearly coded in the VWFA and represents the instantaneous structure/configuration of the limb segments at that given timepoint. Its development is not necessarily dependent on vision (may develop from somatosensory/motor/vestibular inputs). The stick figure is composed of a coarse Gabor-like configuration (mixture of frequencies and sizes corresponding to limb sizes, distances etc.) each representing one segment in the HSF/WBSF. Importantly, the activated HSF/WBSF representation allows viewers to successfully decipher all limb locations overcoming occlusion (by cloths, other body or hand parts). Critically, the representational space these stick figures span are the building blocks of written languages. Specifically, finger segments’ structure and spatial relations allow to create any written letter combination (with sufficient number of hands/limbs). This recycled representation is likely hierarchical in its transition from the physics of vision to the semantics of language, being shaped by top-down language specific statistics and information. Future studies are needed to test the model’s validity and the multiplicity of predictions stemming from it.

This study was funded by ISF Individual Research Grants 1462/23 to SGD

53.467 DO YOU SEE A MOMMY? AUDITORY GENDERED LABELS INFLUENCE ON INFANT FACE PROCESSING

Charisse B Pickron¹ (cpickron@umn.edu), Vandita Gupta¹, Tamanna Dharamsey², Benjamin R Eisenreich¹, Melissa J Polonenko¹; ¹University of Minnesota, ²Ashoka University

Infants develop face processing expertise and build categories about people throughout the first 24 months of life. But to what extent is early face perception and attention influenced by auditory labels? By 3 months of age infants perceptually form categories of male and female faces, yet it is unclear if any gendered auditory labels are associated with such faces. The words ‘mommy’ and ‘daddy’ are commonly used words in the U.S. and have heteronormative associations with female and male faces respectively. In this pre-registered on-going study, we are investigating the way face perception and attention are shaped by auditory labels that contain gendered information. We are recording electroencephalography (EEG) with a 64-channel high-density

hydrocel ANT Neuro cap from 9-, 12-, 18-, and 24-month-old infants while they complete an audio-visual face perception task. The aim is to evaluate time-locked neural responses (ERPs) to faces and novel objects which have been primed with a verbal prompt of ‘mommy’ or ‘daddy.’ Three conditions are randomly presented to infants: 1) Congruent: hearing the word ‘mommy/daddy’ and seeing heteronormative gender-matched faces, 2) Incongruent: hearing these labels and seeing mismatched gendered faces, and 3) Object: seeing a novel non-human object after the heard label. We will analyze amplitude differences for face-sensitive components N290 and P400, and an attention allocation component Nc (Guy et al., 2016) using Mixed Measures ANOVA with Condition and electrode region as within-subjects factors and infant age as a between-subjects factor. Preliminary data (n=4) display N290 & P400 averaged waveforms with similar amplitudes across the three conditions, however the Nc amplitude is larger for Object and Incongruent conditions relative to Congruent. We anticipate data from approximately 30 participants by May. Results will inform our understanding of early antecedents to gender stereotype formation. The authors declare no conflicts of interest.

53.468 DETERMINING THE OPTIMUM FRAME RATE REQUIRED FOR ACCURATE FACE-DIET ESTIMATION FOR WEARABLE EYE-TRACKING

Parsa Delavari^{1,2}, Ipek Oruc^{1,2}; ¹Graduate Program in Neuroscience, University of British Columbia, ²Department of Ophthalmology and Visual Sciences, University of British Columbia

Understanding how often people see faces and attend to them is central to theories of face perception and experience-driven visual development. Wearable eye-tracking devices with egocentric cameras offer a powerful way to quantify this “face diet” in natural environments, but the temporal resolution needed to recover these measures reliably is not well established. The minimum required frame rate determines both measurement fidelity and computational burden. Here we examine the minimum frame rate needed to obtain estimates of daily-life face exposure, including a new measure of gaze-confirmed viewing of attended faces, with acceptable accuracy. Five participants wore Tobii Pro Glasses 3 for two hours of everyday activity. Faces were detected using InsightFace, and gaze data were used to identify attended faces in the scene. Based on full-resolution 25 fps recordings, participants were exposed to 15.55 ± 8.12 minutes/ hour of faces, consistent with prior work. Gaze-confirmed viewing of attended faces was 1.93 ± 1.76 minutes/hour, suggesting that direct attention to faces occupies only a small proportion of time. These 25 fps estimates were used as a benchmark to compare against various levels of downsampling at 12, 6, 3, 2, and 1 fps, and at 1 frame every 2, 4, 8, 15, 30, 60, and 120 seconds. For each rate, multiple subsamples were generated, and 95% confidence intervals were compared against a 5% error tolerance. For all faces, the confidence intervals for three of five participants remained within tolerance at 0.25 fps, and one at 0.07 fps; the most conservative case required 1 fps. In

contrast, attended-face exposure required substantially higher temporal resolution, with stable estimates ranging from 1 fps to 12 fps across participants. Identifying frame rates that balance practical feasibility and reliable estimation is essential, and these findings highlight that minimum frame-rate requirements may vary in a task-dependent manner.

This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant RGPIN-2025-05239 (IO).

Development

**TUESDAY, MAY 19, 8:30 AM – 12:30 PM,
PAVILION**

53.469 ASSOCIATIONS BETWEEN GLOBAL FORM PERCEPTION, MOTION PERCEPTION, AND AUTISM QUOTIENT: AN ONLINE STUDY

Chloe Ding¹ (chloe.ding@uwaterloo.ca), Benjamin Thompson^{1,3}, Andrew E. Silva^{1,2}; ¹University of Waterloo, ²Idaho State University, ³Centre for Eye and Vision Research

We previously observed a relationship between vision and autism quotient (AQ) scores in a group of 9-year old children born at risk of neonatal hypoglycemia (Silva et al., 2023). Both dorsal and ventral stream function were associated with AQ score in boys, but only dorsal stream function was associated with AQ score in girls. To further explore this potential association between vision, AQ score, and biological sex, we conducted an online study in adults. Our hypothesis was that the associations that we found in our cohort of 9-year-old children would be replicated in the adult population. 140 adults (101 neurotypical, 39 diagnosed with ASD, 65 female) completed an AQ questionnaire using the online platform Prolific. Of these, 87 participants completed an online motion discrimination task (dorsal stream), and 123 completed an online form discrimination task (ventral stream). In the motion task, participants indicated whether a stimulus composed of 100 dots moved left or right. In the form task, participants indicated whether a stimulus composed of 200 dots formed a circular or a rectangular Glass pattern. Coherence thresholds were assessed using a 2-down, 1-up staircase. We examined the relationship between AQ score and the two visual tasks with separate linear regressions, including terms for the interaction effect of sex. There was no effect of biological sex and no significant associations for form task thresholds. Lower autistic traits predicted better motion task performance ($p = 0.041$, $r = 0.221$). This correlation was driven by the AQ test's "communication" subscore, ($p = 0.026$, $r = 0.240$). The associations between AQ, vision and biological sex that we observed in at-risk children were not replicated in our adult sample. However, an association between motion perception and AQ score was observed,

indicating a link between visual processing and the expression of autistic traits.

NSERC Discovery Grant

53.470 FLICKERING LENSES ENHANCE READING PERFORMANCE DUE TO A PLACEBO EFFECT

Simone Gori¹, Sandro Franceschini², Giovanna Puccio², Sara Bertoni¹, Sara Mascheretti³, Andrea Cappellini⁴, Andrea Facoetti²; ¹University of Bergamo, ²University of Padua, ³University of Pavia, ⁴Scuola Internazionale di Ottica e Optometria, Firenze

Developmental dyslexia (DD) is the most frequent neurodevelopmental disorder. Traditional remediation programs for DD are rarely controlled for the placebo effect, raising the hypothesis that positive expectations might be responsible for their efficacy. Wearing expensive flickering glasses has been associated with extraordinary improvements in reading skills. The placebo effect and efficacy of these glasses on reading performance were tested. A double blind within-subject experimental design was used in children with DD ($n = 49$; Experiment 1) and unselected young adults ($n = 48$; Experiment 2). Positive expectancy (placebo effect) improved word reading accuracy in young children with DD, with an effect size larger than those reported for gold-standard training programs. This improvement in reading accuracy was replicated in adult poor readers; whereas typical readers improved only in pseudoword decoding speed. Individually-tuned flickering glasses decreased the advantage of word reading over pseudoword reading (the lexicality effect) and predicted pseudoword decoding speed in children with DD. These findings cast shadows on the real efficacy of dyslexia standard training and highlight how the role of placebo effect in training for DD could be dramatically underestimated

This work was supported by: -Ministero dell'Università e della Ricerca PRIN Project CUP: F53D23004610006—ID MUR: 2022772HTJ_01 to S.G.; Ministero dell'Università e della Ricerca (MUR 2022FWME9W, Bando PRIN 2022 - DD 104 del 02/02/2022, DD 1401 del 18/09/2024) CUP C53C24001180006, to A.F.

53.471 INVESTIGATING THE ROLE OF VISUAL EXPERIENCE IN THE DEVELOPMENT OF DIFFERENT FUNCTIONS OF THE FERRET'S MOTION PATHWAY

Emmanuel Osikpa^{1,2}, Kristina Nielsen^{1,2}; ¹Department of Neuroscience, Johns Hopkins School of Medicine, ²Zanvyl Krieger Mind/Brain Institute, Johns Hopkins University

Visual experience has been shown to play a crucial role in the correct development of at least some visual functions. Direction selectivity in ferret primary visual cortex (V1), for example, requires visual experience (Li et al. 2006), as does direction selectivity in cat and ferret higher visual motion area PMLS (Spear et al, 1985; Khamiss 2025). However, not all functions are equally impacted by manipulations of visual experience: Orientation selectivity in ferret V1 does not require visual experience to develop normally, in contrast to the profound loss

of direction selectivity (Li et al. 2006). Furthermore, deficits induced by altered visual experience differ between brain regions: Direction selectivity appears to be much more strongly affected by lack of visual experience in ferret PMLS than V1 (Khamiss 2025). Clearly, there is a complex interplay between visual experience and different functions at different stages in the visual hierarchy. Identifying which functions, in which areas, can develop based on intrinsic factors alone, and which require visual experience, will be important not just for improving our understanding of visual development, but also for building better models for developmental disorders. Here, we begin to address this issue by testing the impact of visual experience on a range of tuning functions in ferret V1 and PMLS. More precisely, we use extracellular recordings in V1 and PMLS to quantify tuning for stimulus parameters like orientation, direction, contrast, spatial and temporal frequency with and without visual experience. Our results confirm that V1 and PMLS orientation selectivity can develop largely normally without visual experience, while both areas show large deficits in direction selectivity. Other tuning functions are also impacted, including a significantly lower contrast sensitivity without visual experience. Again, these deficits appear larger in PMLS than V1, confirming the importance of systematically probing changes in tuning functions across visual areas.

53.472 COMPARING EARLY FACE EXPERIENCES ACROSS TWO COMMUNITIES

Alexis R. Colwell¹, Linda B. Smith¹; ¹Indiana University Bloomington

The perception of human faces is central to identifying people, communication, and reading emotions. Human infants show experience-dependent rapid developmental advances in all these abilities over the first year of life post birth. Advancing research suggests strong developmentally changing constraints on those experiences, with early experiences characterized by high frequency frontal views of faces that are close to the perceiver, with frequency and proximity decreasing with age and variability increasing. However, these findings have been shown for infants in one US town, and the observed developmental changes in face inputs could be specific to these infants and not infants more generally. We used head cameras to capture the faces in view for infants (n = 44, 1 to 12 months) in a mid-size US town and a densely populated urban fishing village in India. Analyses of the 112,000 images indicated that US infants saw more faces overall compared to Indian infants, but Indian infants saw a greater number of unique faces. Despite this difference in number of unique faces, most faces infants saw in both countries belonged to the same few people. Despite these differences, similar developmental trends were observed: the frequency of faces and their proximity to the infant declined over the first year in both locations. Further, frontal views of faces dominated visual experiences in both locations. The results suggest similar developmental patterns in face experience during the first year of life in both locations despite marked differences in the visual environments.

53.473 INFANT EGOCENTRIC SCENE STATISTICS ARE SIMPLER THAN ADULT EGOCENTRIC AND CANONICAL NATURAL SCENES

Philip McAdams¹, Alexis Colwell¹, Minju Kim¹, Linda B Smith¹; ¹Indiana University

Vision emerges in a richly structured world. Classic work on natural scene statistics has quantified framed outdoor photographs, revealing scale-invariant amplitude spectra and characteristic local phase and contour structure, orientation anisotropies, and distinctive distributions of luminance and chromatic contrast. These statistics, part of our evolutionary experience, are thought to constrain early visual representations and approximate the input that shapes these representations during development. Yet infants and adults in industrialised societies spend ~90% of waking time indoors, and laboratory studies sample only a tiny fraction of this visual diet and rarely include naturalistic egocentric views. Building on recent work, we ask whether the everyday egocentric visual experiences of infants and adults share the same statistics as canonical natural scenes. We use head-mounted cameras to quantify the visual diet of 96 infants (1–24 months) and 20 adults in the US. We analyse ~100,000 primarily indoor (95%) egocentric images and compare them with two outdoor natural-scene corpora: adult egocentric views and natural-scene databases (e.g., McGill). We standardised images and computed image statistics linked to early visual processing (edge density/orientations, spectral slope, fractal dimension, luminance and colour), analysed with linear mixed-effects models and PCA-based comparisons across corpora. Infant and adult everyday egocentric visual experiences are simpler than egocentric outdoor and natural-scene databases. Simplicity is most pronounced in the first year, with scenes showing low edge density, edge-orientation entropy and fractal dimension, steeper spectral slopes ($\beta = -1.6$), and constrained luminance and chromatic distributions. Complexity increases in 12–24-month-olds, who encountered the brightest scenes and maximal luminance/chromatic contrasts relative to infant or adult egocentric groups. Thus, the “visual curriculum” shaping early visual cortex departs from canonical natural-scene statistics: it starts simpler, becomes more adult-like by the second year, yet remains simpler than natural scenes. Theories of visual perception and development must reconcile evolutionary constraints with egocentric daily-life inputs.

53.474 LEARNING TO SEE THROUGH A BABY'S EYES: EARLY VISUAL DIETS ENABLE ROBUST VISUAL INTELLIGENCE IN HUMANS AND MACHINES

Yusen Cai¹, Qing Lin¹, Bhargava Satya Nunna², Mengmi Zhang¹; ¹Nanyang Technological University, ²Indian Institute Of Technology Madras

Newborns perceive the world with low-acuity, color-degraded, and temporally continuous vision, which gradually sharpens as infants develop. To explore the ecological advantages of such staged “visual diets”, we train self-supervised learning (SSL) models on object-

centric videos under constraints that simulate infant vision: grayscale-to-color (C), blur-to-sharp (A), and preserved temporal continuity (T)—collectively termed CATDiet. For evaluation, we establish a comprehensive benchmark across ten datasets, covering clean and corrupted image recognition, texture–shape cue conflict tests, silhouette recognition, depth-order classification, and the visual cliff paradigm. All CATDiet variants demonstrate enhanced robustness in object recognition, despite being trained solely on object-centric videos. Remarkably, models also exhibit biologically aligned developmental patterns, including neural plasticity changes mirroring synaptic density in macaque V1 and behaviors resembling infants' visual cliff responses. Building on these insights, CombDiet initializes SSL with CATDiet before standard training while preserving temporal continuity. Trained on object-centric or head-mounted infant videos, CombDiet outperforms standard SSL on both in-domain and out-of-domain object recognition and depth perception. Together, these results suggest that the developmental progression of early infant visual experience offers a powerful reverse-engineering framework for understanding the emergence of robust visual intelligence in machines. All code, data, and models will be publicly released.

This research is supported by the National Research Foundation, Singapore under its NRFF award NRF-NRFF15-2023-0001 and Mengmi Zhang's Startup Grant from Nanyang Technological University, Singapore

53.475 SHORT-TERM DEVELOPMENTAL TRAJECTORIES OF DORSAL-VENTRAL PATHWAYS AND THEIR RELATIONSHIPS WITH FIRST-GRADE LEARNING

Xueying Ren¹ (xueying.ren@vanderbilt.edu), James R. Booth¹, Gabriele Amoroso², Franco Pestilli², Sophia Vinci-Booher¹; ¹Vanderbilt University, ²The University of Texas at Austin

Visual processing regions in the ventral temporal cortex connect directly to the parietal cortex via the posterior vertical pathway (PVP), yet the development of its white matter tracts and their relationship with learning, particularly during periods of rapid change, is poorly understood. This pre-registered study (<https://osf.io/97ybe>) used a dense longitudinal neuroimaging design to capture rapid and potentially nonlinear changes in the microstructure of PVP tracts throughout the first grade, a year of foundational learning. We collected diffusion MRI and behavioral assessments of number processing and word processing from ten first-grade children every month for 12 months. First, regarding the behavioral data, we identified nonlinear changes in learning throughout the first year of schooling. Then, we used a time series approach to model the temporal relationships among brain structures at the start and ongoing brain and behavioral changes observed throughout the year. Contrary to our expectation, we did not find a relationship between the posterior Arcuate (pArc) and word processing. Instead, we found that month-to-month changes in fractional anisotropy (FA) in the right Middle Longitudinal Fasciculus (MDLF)—specifically the connections to the Angular Gyrus (MDLF-Ang) and Superior Parietal Lobe (MDLF-

SPL)—positively predicted month-to-month changes in number processing, but not word processing. This captures an ongoing coupling between brain change and learning. Also, this temporal relationship was underpinned by changes in orientation dispersion (ODI), indexing axonal organization. Our results demonstrated that rapid changes in at least one of the microstructure of PVP tracts occurred alongside rapid changes in number processing during a relatively short critical window of foundational learning, suggesting a dynamic and ongoing interplay between brain and behavioral change.

Funding: R01 HD114489

53.476 SLOW VISION AS A DEVELOPMENTAL FEATURE: TEMPORAL CONSTRAINTS SHAPE OBJECT CATEGORIZATION IN INFANCY

Jean-Rémy HOCHMANN^{1,2} (jr.hochmann@gmail.com), Céline Spriet^{1,2}, Liuba PAPEO^{1,2}; ¹CNRS UMR5229 – Institut des Sciences Cognitives Marc Jeannerod, 67 Bd Pinel 69675 Bron, France, ²Université Claude Bernard Lyon 1, France

The hierarchical neuronal and cognitive underpinning of visual perception are largely in place early in life, possibly at birth, but immature brain and fledgling myelination might slow down the transmission and integration of information into complex representations. In 5 experiments involving 160 infants, we used frequency-tagging electroencephalography combined with rapid serial visual presentation to show that the capacity to categorize objects at 4 and 9 months is respectively seven times and three times slower than in adults. Next, we tested adults with accelerated visual streams to simulate object perception in young infants. We used representational similarity analysis (RSA) to compare representational models extracted respectively from a multivariate pattern analysis of the adult EEG signal and from the eight layers of an artificial deep convolutional neural network (CNN) trained on object categorization (Alexnet). RSA revealed significant correlations between EEG-based representations and each layer of the CNN, for all speed conditions (6, 12 or 30 Hz). Cluster-mass permutation tests however revealed that increased speed of presentation from 12 and 30 Hz reduced correlations of neural representations with layers conv3, conv4 and conv5 around 205-314 ms. This suggests that, when visual streams accelerated, low complexity features, accessed early and coded in early layers modelling EVC, were maintained, whereas higher-complexity visual features akin to those computed in more anterior visual areas were partly lost. In other words, when streams accelerated, lower-level sensory information was mostly preserved, but the computations transforming that information into meaningful categories were impacted. Taking adults' visual perception at faster stimulus presentation as a simulation of young infants' perception, these results suggest that infants may have access to more limited feature maps, yielding representation of less detailed, broader categories. Slow vision may constitute a design feature of early development, highlighting the most stable and general features of the world.

L. P. was supported by a European Research Council Starting Grant (Project: THEMPO, Grant Agreement 758473). J.-R. H. was supported by the Fondation de France (Grant number: 00134704 / WB-2022-46104).

53.477 PREDICTING RETINOTOPY FROM ANATOMY IN >11,000 INDIVIDUALS REVEALS AGE-RELATED CHANGES IN PRIMARY VISUAL CORTEX

Fernanda L. Ribeiro^{1,2,3,4} (fernanda.l.ribeiro@uni-giessen.de), Felix Hoffstaedter^{5,6}, Martin N. Hebart^{1,3,4}; ¹Department of Medicine, Justus-Liebig University Giessen, Germany, ²School of Electrical Engineering and Computer Science, The University of Queensland, Australia, ³Vision and Computational Cognition Group, Max Planck Institute for Human Cognitive and Brain Sciences, Germany, ⁴Center for Mind, Brain and Behavior (CMBB), University of Marburg, Giessen and Darmstadt, Germany, ⁵Institute of Neuroscience and Medicine (INM-7: Brain and Behaviour), Research Center Jülich, Germany, ⁶Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University Düsseldorf, Germany

Understanding how visual cortex organization changes across the lifespan may help elucidate visual performance variation within and across individuals. Previous work has demonstrated that more V1 cortical surface area is dedicated to representing the horizontal versus the vertical meridian, termed horizontal-vertical anisotropy (HVA), and more surface area to representing the lower versus the upper vertical meridian, termed vertical-meridian asymmetry (VMA) (Benson et al., 2021). Moreover, these asymmetries in cortical representation of the horizontal and vertical meridians vary between children and adults (Himmelberg et al., 2023), mirroring asymmetries in visual task performance around the visual field (Carrasco et al., 2022). Here, we investigate whether these age-related differences can be predicted from cortical geometry by retrospectively applying the deepRetinotopy toolbox (Ribeiro et al., 2025) to 11,060 anatomical scans. Specifically, we applied deepRetinotopy to the ABCD baseline cohort ($n = 9,947$; age 9-11) and the HCP Young Adult dataset ($n = 1,113$, age 22-36) to predict individual-level retinotopic organization from cortical folds, derived from T1-weighted images. Then, we estimated cortical VMA and HVA, operationalized as the difference between the surface area dedicated to representing two different portions of the visual field (horizontal versus vertical meridian for HVA; lower vertical versus upper vertical meridian for VMA) divided by the mean combined surface area. We found group differences in cortical VMA (higher VMA in adults compared to children; Cohen's $d = 0.18$, $p < 0.001$) and HVA (higher HVA in adults compared to children; Cohen's $d = 0.09$, $p < 0.01$), consistent with prior work. Importantly, because our predictions derive from cortical geometry alone, our findings reveal that asymmetries in how the visual cortex samples different meridians can arise from local changes in cortical geometry. More broadly, these findings demonstrate how deepRetinotopy enables investigating visual cortex organization across development and individuals at an unprecedented scale.

FLR acknowledges support through the European Union's Horizon Europe research and innovation funding program under the Marie Skłodowska-Curie Actions project ID 101146996.

TUESDAY AFTERNOON POSTERS IN BANYAN BREEZEWAY

Visual Memory: Imagery

TUESDAY, MAY 19, 2:45 – 6:45 PM, BANYAN BREEZEWAY

56.301 MULTIPLICATIVE NOISE IN THE PERKY EFFECT; HOW IMAGERY AFFECTS VISION

Adam Reeves¹; ¹Northeastern University

We measured the 'Perky effect', the change in vision consequent on having a visual image (Perky, 1910), in a test of visual acuity. Subjects reported whether an upper line was left or right of a lower line (2AFC), or reported if one line was left, center, or right (3AFC). Accuracy was converted to d' . Maintaining an image of vertical lines around the vertical line stimuli increased d' at low levels ($d' < 0.6$) but lowered d' at high levels ($d' > 1$), relative to d' in no imagery. The change from 'reverse Perky' (helping) to the regular Perky effect (hindering) occurred in two separate experiments with 54 and 135 subjects. It was modeled by assuming imagery both (a) boosts the signal by 2 to 4 times (helping) and (b) adds increasing amounts of signal-dependent multiplicative noise, N_x , to the intrinsic noise, N_o (hindering). In imagery, N_x eventually dominates as signal strength increases, reducing sensitivity by up to 0.9 d' units relative to no imagery. Adding external noise reduces this effect to ~0.3 d' units (Dijksra et al, 2022), we think by partially swamping the multiplicative noise. The basic constraint that d' must increase with signal level can be met if N_o is Gaussian and N_x is exponential, the total noise being exGauss . The model was fit by least-squares with 2 parameters and this constraint in place; RMS errors were low, $0.08 < 0.17$. ExGauss noise implies that the ratio $\text{SD}(\text{Imagery})/\text{SD}(\text{No imagery})$ increases predictably with signal level, as indeed we found ($r > 0.93$).

Thanks to Severine Merand and Amedeo D'Angiulli

56.302 A MULTIDIMENSIONAL EXPERIENCE SAMPLING APPROACH TO MENTAL IMAGERY IN LEARNING AND MEMORY

Silvia Shiwei Zhou¹ (16sz48@queensu.ca), Hala Rahman¹, Keanna Rowchan¹, Tasha Ignatius¹, Jeffrey D. Wammes¹; ¹Queen's University

One's ability to generate mental imagery can influence a broad spectrum of cognitive processes. However, imagery is often measured as a single item or construct using trait-level questionnaires despite the multifaceted nature of imagery both within and across individuals.

This makes it difficult to capture whether not only the ability to generate imagery but also the particular properties of that imagery (e.g., sharpness, colorfulness, prototypicality) contribute to learning and memory. Multidimensional experience sampling (mDES) offers a promising method for capturing the nuanced qualities of ongoing thoughts (e.g., visual vs. verbal, detailedness). Inspired by mDES, we developed a novel set of 15 imagery-mDES items that ask participants to rate the experiential properties of their mental images on continuous scales. These items assess perceptual (e.g., colorfulness, sharpness), subjective (e.g., difficulty, immersive), and dynamic properties (e.g., broader context) of imagery. Participants (N = 82) encoded by viewing an object, generating a mental image for 4–8 seconds, making a judgment of learning (JOL), and then completing the imagery-mDES items. Visual memory was later tested using a difficult 4-alternative forced-choice remember/know/guess task (RKG). We identified three dominant principal components that describe the dimensions along which people's imagery varied: "vivid detailed," "dynamic contextual," and "difficult prototypical." These patterns were meaningfully related to memory: remembered objects were associated with more "vivid detailed" imagery, and less "difficult prototypical" imagery. Among correct trials, "vivid detailed" imagery was associated with a dramatic RKG gradient, whereas "dynamic contextual" imagery led to fewer guesses. Participants' JOLs paralleled these effects, suggesting metacognitive sensitivity to imagery properties. By characterizing imagery at the item level, our approach helps quantify how different imagery themes contribute to encoding and retention. This multidimensional approach opens new possibilities for understanding how visual cognition shapes learning, and revealing individual differences in how imagery is deployed.

56.303 HOW DO WE REMEMBER VISUAL INFORMATION WITHOUT VISUAL IMAGERY?

Ava G. Peruski¹ (agperusk@iu.edu), Karin H. James¹; ¹Indiana University, Department of Psychological and Brain Sciences

Utilizing visual imagery is a common occurrence in day-to-day life for most people, particularly in recalling visual memories. However, a portion of the population is unable to create these images – a condition known as aphantasia. This begs the question, how does one remember visual information without visual imagery? Past literature has found that aphantasics differ from typical imagers in the effectiveness of different retrieval strategies and that aphantasics self-report using written/verbal strategies to encode and retrieve visual information. With this in mind, we wondered whether the effectiveness of encoding support strategies would vary between groups of different imagery abilities, which could give us an insight into how visual imagery is tied to visual memory. We hypothesized that when given different encoding support strategies (thinking, writing, drawing, control) for memorizing visual information, writing (based on the self-reported use of writing for encoding and retrieval) would be more effective for aphantasics compared to a high imaging ability group. We measured both free recall and forced choice recognition of items

included in cartoon pictures and found no difference between our two groups, but we did find that thinking, followed by writing, were the most effective encoding strategies overall for memorizing visual information. It is possible that because writing and drawing required participants to look away from the picture, while thinking did not, this caused thinking to be the most effective strategy. Additionally, it is possible that drawing and writing would have been more effective for longer-term retention. To address these limitations, future work will examine these conditions as different rehearsal strategies and will examine multiple time points.

Research reported in this abstract was supported by the Eunice Kennedy Shriver National Institute Of Child Health & Human Development of the National Institutes of Health under Award Number T32HD007475. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

56.304 SPATIAL DEPENDENCE AND SPATIAL TOLERANCE IN IMAGERY AFTEREFFECTS FOR STATIC ORIENTATION AND DYNAMIC ROTATION

Yanna Mao¹, Rong Jiang¹, Ming Meng²; ¹South China Normal University, Guangzhou, China, ²University of Alabama at Birmingham, Birmingham, USA

Visual imagery can evoke perceptual-like representations in early visual cortex and bias subsequent conscious perception (i.e., imagery aftereffects). Unlike perception, which is constrained by retinal architecture, visual imagery, however, can subjectively extend beyond these retinal boundaries. This contrast raises the question of whether imagery aftereffects follow systematic spatial constraints. To address this issue, we conducted three psychophysical experiments examining the spatial characteristics of imagery aftereffects. Experiment 1 tested whether imagery aftereffects for static stimuli exhibit spatial specificity. Participants imagined a static grating of horizontal or vertical orientation at a specified location, and then detected a briefly presented probe embedded in noise. Results revealed a spatial dependence of imagery aftereffects. Specifically, imagery facilitated detection of orientation-matching probes relative to a no-imagery baseline only at the imagined location, with no facilitation elsewhere. Experiment 2 investigated how spatial location modulates imagery aftereffects for a dynamic rotating disc. Participants imagined a static disc initiating clockwise or counterclockwise rotation at varying distances, and then viewed a centrally presented bistable disc. The results revealed spatial tolerance in imagery aftereffects. Specifically, the perceptual bias toward the imagined direction decreased gradually with increasing distance, yet remained detectable even at the largest separation (12°). Experiment 3 tested whether imagery aftereffects exhibit central-peripheral asymmetry. Whereas Experiment 2 presented probes only at foveal fixation, Experiment 3 contrasted imagery at fixation with testing in the periphery and vice versa. Both conditions produced positive imagery aftereffects relative to the location-specific baseline perceptual bias, with comparable

magnitudes, suggesting an absence of central–peripheral asymmetry. Taken together, these results indicate that imagery aftereffects are influenced by spatial location and further modulated by stimulus properties. Visual imagery thus operates under systematic spatial principles, offering insights into how imagery interacts with early visual processing.

56.305 THE TIME COURSE OF VISUAL IMAGE GENERATION

Yueyang Zhang¹, David Melcher¹; ¹New York University Abu Dhabi

We investigated the time course of visual image generation: does imagery involve an instantaneous retrieval of a stored image or is it a gradual, constructive process? In order to map the time course of visual imagery vividness, we moved beyond traditional self-report vividness scales to directly compare imagery with the processing steps of diffusion-based image-generation models. In two behavioral experiments, participants viewed images of common objects (e.g., animals, tools, foods) that were created using Stable Diffusion 1.4 and CLIP ViT-B/32. Prompts followed the structure “a + color + object”. Each image was produced with 10 inference steps, and outputs from steps 1–10 served as our vividness scale. In Experiment 1 (N = 33), participants heard an auditory prompt, imagined the object for 1 s or 4 s, answered a catch question to ensure comprehension, then rated the vividness of their mental image by comparing it to the diffusion-step images, focusing solely on vividness. A linear mixed model controlling for familiarity showed that longer imagery time significantly increased vividness, and mean vividness was correlated with VVIQ scores. Experiment 2 (N = 38) extended the design to four imagery durations: 1, 2, 3, and 4 seconds (N = 38). Again, reported vividness increased with imagery time, with the largest gain between 1 and 2 seconds, with diminishing gains for longer durations. These findings support a gradual construction view of imagery, where representations develop from more unclear to more detailed images. This suggests that visual imagery is a stepwise, constructive process rather than an instantaneous retrieval.

56.306 A NOVEL APPROACH TO CHARACTERIZING THE VISUAL DIMENSIONS OF MENTAL IMAGERY VIVIDNESS USING GAUSSIAN PROCESS REGRESSION

Oğul Can Yurdakul¹ (oyurdaku@uci.edu), Xueyi Huang², Angela Shen¹, Emil Olsson¹, Ali Ekhlas¹, Michaela Klimová², Jorge Morales², Megan A. K. Peters^{1,3,4}; ¹University of California Irvine, ²Northeastern University, ³Canadian Institute for Advanced Research, ⁴University College London

Mental imagery research has long treated vividness as a central but poorly specified construct. However, questions remain: Is ‘vividness’ objectively definable? Does it reflect a single latent dimension or multiple interacting dimensions? How does it relate to perceptual reality monitoring? These questions have theoretical implications for higher-order theories of consciousness – which disagree about whether vividness constitutes a content-invariant reality-monitoring

signal or a separable higher-order dimension – and methodological implications for developing scalable, data-driven approaches for quantitatively characterizing vividness across individuals, contents, and perception versus imagination. Here, we introduce a novel computational approach for recovering the multidimensional structure underlying vividness judgments using data collected as pilot for a large, preregistered behavioral experiment (full experiment reported by Huang et al., this VSS). Participants (n = 88 in the pilot) generated mental images and rated vividness, then reproduced their imagery experience on abstract Voronoi images across three dimensions: sharpness, opacity, and color saturation. Rather than pre-specifying the form of the mapping to vividness, we use Gaussian process regression to flexibly model joint relationships between vividness and these dimensions while also capturing uncertainty in the fitted model arising from uneven sampling. Our approach estimates functions linking vividness to these dimensions to characterize nonlinear, heterogeneous mappings between reconstruction dimensions and vividness, and individual differences in how these dimensions contribute to vividness reports. Crucially, our approach avoids parametric assumptions of linear models and enables recovery of potentially nonlinear manifolds linking these interacting dimensions. Preliminary results reveal peaks in the joint space which suggest nonmonotonicity and individual differences in how these features define vividness. Our framework offers a generalizable, theory-flexible approach for quantifying the latent structure of imagery vividness and its relationship to perceptual dimensions, facilitating future psychophysical and neuroimaging experiments characterizing imagery’s relationship to perception and the neurocomputational processes underlying reality monitoring.

This work was supported by the Templeton World Charity Foundation, Inc. (funder DOI 501100011730) under the grant <https://doi.org/10.54224/22032> (to JM & MAK) and in part by the Canadian Institute for Advanced Research (CIFAR, Fellowship in Brain, Mind, & Consciousness; to MAK).

56.307 WHAT WE TALK ABOUT WHEN WE TALK ABOUT VIVIDNESS

Xueyi Huang¹, Angela Shen², Emil Olsson², Michaela Klimova¹, Ali Ekhlas², Oğul Can Yurdakul², Megan A. K. Peters^{2,3}, Jorge Morales¹; ¹Northeastern University, ²University of California, Irvine, ³University College London

The vividness of visual mental images is conventionally measured on a 1–5 scale. However, because mental imagery is inherently subjective, vividness alone provides limited information about the underlying mental images. Can vividness ratings be linked to specific visual properties on an individual basis? Are people accurate at reporting visual properties from imagination? Is there a “shared” vividness concept across people? We investigated these foundational questions across four experiments leveraging a novel method that captures trial-by-trial links between vividness and visual properties: Subjects imagine real-life objects as vividly as possible, rate their

vividness, and either reconstruct or classify their imagery's visual properties—sharpness, opacity, and saturation—on Voronoi tessellations. Experiments 1 & 2 (n=200 each, preregistered) revealed that while subjects associate specific visual-property combinations with particular vividness degrees, there are large individual differences. Experiment 3 (n=50) validated the reconstruction method: Subjects viewed realistic images altered with random property values and accurately reconstructed them on Voronoi. Experiment 4 (n=100) replicated Experiment 1 in a new sample. Subjects then rated the vividness of realistic images whose visual properties were either taken from the subject's own reconstructions or from the averages of all subjects during the imagery task. Consistent with Experiment 2 which used a double-pass procedure to show within-subject reliability, ratings of subjects' reconstructions were highly correlated with subjects' original vividness ratings. Surprisingly, subjects' responses were even more strongly correlated with the original ratings when using group-average visual properties, making them an even more stable index of vividness. Together, our findings demonstrate that despite their subjective nature, mental images and their experienced vividness can be described accurately and objectively. Crucially, our novel paradigm captures both individual differences and converging interpretations in how visual properties map onto vividness. It appears that we do know what we're talking about when we talk about vividness.

This work was supported by the Templeton World Charity Foundation, Inc. (funder DOI 501100011730) under the grant <https://doi.org/10.54224/22032> to (JM & MAKP) and in part by the Canadian Institute for Advanced Research (CIFAR, Fellowship in Brain, Mind, & Consciousness; to MAKP).

56.308 APHANTASIA: NOT THE PROBLEM OF NEURAL REPRESENTATION?

Sang Wook Hong^{1,2} (shong6@fau.edu), Yosun Yoon^{1,2}; ¹Department of Psychology, Florida Atlantic University, ²Stiles-Nicolson Brain Institute, Florida Atlantic University

When asked to visualize a large red apple, most people can effortlessly generate a vivid mental image. In contrast, individuals with aphantasia report being unable to voluntarily produce such visual imagery (Zeman et al., 2015). A central question is whether aphantasia reflects a failure to form neural representations of imagined content, or instead a disruption in conscious access to otherwise intact representations. The present study addressed this question using fMRI combined with multivariate pattern analysis (MVPA). Four participants (three with aphantasia) completed four scanning sessions over a two-week period: (1) retinotopy and localizer, (2) orientation decoding of visually presented gratings, (3) decoding of orientation held in working memory, and (4) decoding of visually imagined orientation. The working-memory task followed Harrison & Tong (2008): two oriented gratings (200 ms each) were presented sequentially, followed by a cue indicating which orientation to retain during an 11-s delay. In the imagery condition, gratings were replaced

with double-sided arrows, and participants were instructed to imagine a full-field grating of the indicated orientation. Across all participants—including those with aphantasia—MVPA reliably discriminated orientation-specific activity patterns in early visual cortex for (1) viewed gratings, (2) remembered gratings, and (3) imagined gratings. These findings indicate that individuals with aphantasia do form decodable sensory-like representations during imagery, suggesting that the condition may reflect a disruption of phenomenological experience rather than a lack of underlying neural representation.

Stiles-Nicolson Brain Institute, Florida Atlantic University

56.309 AI-GENERATED IMAGERY SUCCESSFULLY CAPTURES MENTAL VISUALIZATIONS

Zaynab Rehman^{1,2} (zaynab.rehman.de@gmail.com), Julia Thomas³, Wilma A. Bainbridge¹; ¹University of Chicago, ²University of Illinois Urbana-Champaign, ³Cardiff University

When reading a story, one often conjures up mental images of the scenes and events unfolding in the text. Mental imagery generated by individuals after reading a given text may differ depending on which experiences and biases that individual is drawing from. How similar are people in what they visualize, and is there a way to access these mental images by using generative artificial intelligence methods? To test this question, we presented participants (N = 124) with brief synopses of news stories and asked them to mentally visualize what they had read. Participants were then presented with four images (the “ground-truth” image used in the original news article, as well as three randomly intermixed AI-generated images) and rated how well these images matched their own personal mental imagery by providing a concordance score. The participants also described what specific details about the images led to higher or lower concordance with their own mental imagery. Of the 20 articles selected, participants gave the AI-generated images higher concordance scores than the ground-truth image used by the news article 80% of the time. This suggests AI better matches mental visualizations than the images chosen by news editorial teams and that AI platforms may serve as a powerful tool for uncovering these mental images. Consistent patterns in concordance were also observed across participants for the news articles, indicating frequent agreement in which images matched participant mental visualization the most and which matched the least. We are now assessing whether it is possible to improve concordance scores across participants by re-generating the images according to participants' qualitative comments about the AI-generated photographs. Leveraging AI-generated images and ratings of concordance could present a novel way of extracting information regarding mental imagery, memories, or individual biases.

56.310 EMERGENT VISUAL MENTAL IMAGERY IN LARGE LANGUAGE MODELS

Morgan McCarty¹ (mccarty.mo@northeastern.edu), Jorge Morales¹; ¹Northeastern University

Imagine a "D", rotate it 90 degrees to the left, now add a "J" right under it. What does that look like? If your response was "umbrella", you probably arrived there by manipulating the letters and visualizing the resulting composite object in your mind's eye. However, could tasks like this be solved by using language and reasoning alone? Large Language Models (LLMs) offer a perfect case for testing this hypothesis. We adapted this classic task and extended it with dozens of novel instruction-sets, thus avoiding data contamination in the models' training sets. Multiple state-of-the-art LLMs and human participants carried out step-by-step transformations of letters and shapes with the goal of naming the resulting object. We found that the best LLMs performed at, or even above, average human level. Despite this similarity, humans and machines did not succeed and fail in the same way. There were radically different patterns of human and LLM mistakes in a large proportion of the instruction-sets. LLMs provided more consistent answers (across models) than humans (across subjects), but human individual answers were often ranked higher. Finally, both humans and machines sometimes misconstrued (albeit somewhat differently) the intermediate steps of the instruction-sets. There were notable cases in which the most common answers were bimodal, indicating that some items were prone to occasional catastrophic failures or to systematic misunderstanding of the resulting object (e.g., a balloon with a string and bow may be misinterpreted as a stick person if the component proportions are wrong). Our results offer the enticing possibility that tasks long thought to require visual mental imagery can be solved exclusively via a (quasi-)propositional format. Our findings also offer a cautionary tale: Human and artificial intelligence systems' comparable performance in visual and cognitive tasks may hide important underlying computational differences.

This work was funded, in part, by a Northeastern University Undergraduate Research and Fellowships PEAK Summit Award.

Visual Memory: Mechanisms, models, individual differences

TUESDAY, MAY 19, 2:45 – 6:45 PM, BANYAN BREEZEWAY

56.311 THE CODING OF MULTIPLE VISUAL FEATURES IN VISUAL WORKING MEMORY

Yaoda Xu¹, Marvin Chun¹; ¹Yale University

Compared to visual perception, visual working memory has a limited storage capacity and, in general, evokes much lower neural responses. How does this impact the coding of multiple visual features in VWM? Here, we examined the coding of shape features from different objects and the coding of color and shape from the same object in VWM. In visual perception, neural responses for a pair of objects shown together can be approximated by the linear average of those of each object shown alone. Such response averaging is

considered an effective mechanism to combat representational distortion due to neuronal response saturation. Given that response amplitudes are much lower in VWM than those in perception, does averaging exist in VWM? Across two experiments, we found that, after accounting for task factors such as load, VWM representations for two objects can also be approximated by the linear average of those of each component object in both human occipitotemporal cortex (OTC) and posterior parietal cortex (PPC). In contrast, although an object's color and shape are coded in a largely independent manner during visual perception in OTC, we found in a third experiment that these features form a partially integrated representation in VWM. Specifically, a linear classifier trained to decode a pair of objects in one color showed a significant decoding drop when asked to decode the same pair of objects in a different color during VWM delay, but not during VWM encoding. This indicates interactive feature coding in VWM, unlike that in perception. Together, our results show that features from different objects maintain their representational separation in VWM as they do in perception, despite a drop in response amplitude; meanwhile, features from the same object appear to lose their representational independence from perception to VWM and form partially integrated representations in VWM.

Supported by NIH Grant R01EY030854 to YX.

56.312 RE-EXAMINING CATEGORY-SELECTIVE SENSORY REINSTATEMENT IN VISUAL RECALL

Deepasri Prasad¹, Adam Steel^{1,2}, Caroline E. Robertson¹; ¹Dartmouth College, ²University of Illinois Urbana-Champaign

Our classic understanding of visual recall proposes that category-selective areas are activated during perception and similarly re-activated when recalling their preferred stimulus (O'Craven & Kanwisher, 2000). Recent work has complicated this "sensory reinstatement" hypothesis, demonstrating variations like topographic shifts between encoding and recall (scenes; Steel et al., 2021) and intermixed neuronal subpopulations selective for encoding and recognition (faces; Quiroga et al., 2023; Chen et al., 2024). Given this variety, are there common principles or characteristic differences across how category-selective areas implement sensory reinstatement? Here, we examine the activation patterns during perception and recall in a range of category-selective visual areas using fMRI. Participants (N=25) performed visual recall and perception tasks during the scan. For recall, participants vividly recalled self-generated personally familiar faces, scenes, objects, and body parts. For perception, participants viewed dynamic videos from the same categories. Using functional areas defined from independent atlases (parahippocampal (PPA), occipital (OPA) place areas; occipital (OFA), fusiform face areas (FFA1, FFA2); extrastriate body subregions (ITG, LOS, MTG)), we identified category-specific regions of interest in both tasks (contrast: category of interest > all other categories). All areas showed some degree of sensory reinstatement, with stronger overall univariate activity for their preferred category in both perception and recall. For almost all areas, the most selective voxels for perception

were similarly the most selective for recall. Interestingly, in PPA, OPA, and OFA, the top-most selective voxels differed between the tasks. A significant anterior shift in recall versus perceptual activity drove this effect in PPA and OPA only (all p -values < 0.001). OFA activity showed no systematic shift between recall and perception. Instead, OFA contained intermixed subpopulations of mnemonic- and perceptually-selective voxels. Together, this indicates that no single implementation of sensory reinstatement is common across all visual categories; instead, the reinstatement method might reflect unique mnemonic demands of each visual category.

NSF Career Award 2144700

56.313 REPRESENTATIONAL DRIFT HELPS NEURAL NETWORKS FIND STABLE AND SPARSE SOLUTIONS FOR VISUAL WORKING MEMORY TASK

Chattarin Pongtubtim¹, Timothy F. Brady¹, John T. Serences¹; ¹Department of Psychology, UCSD

Representational drift is a phenomenon in which neural activity associated with the same stimulus gradually deviates from its initial pattern over a period of days and weeks, even after behavioral performance has stabilized. While the cause and functional role of this drift is still debated, recent studies have shown that representation drift is correlated with sparser and more separable patterns of neural activity (Ratzon et al., 2024; Kumar et al., 2025). Thus, drift might result from neural networks moving toward more efficient, sparser solutions that still support adequate task performance. In turn, as the separability of neural representations increases, any noisy perturbations of the representations will be less likely to cause confusion. Therefore, representational drift might also drive networks towards more stable and robust solutions. To test this hypothesis, we trained continuous time recurrent neural networks (ctRNNs) on a visual working memory task that required encoding and storing stimuli over a brief delay period. To ensure stimulus encoding independent of the required response, we presented a cue after the delay period indicating the stimulus-response mapping, and the model produced a recall response indicating the identity of the remembered stimulus. We observe that models drift further from their initial states even as performance remains stable: representational drift. As they drift, their activity becomes more sparse, and has greater separation between different memory representations. Crucially, we show that post-drift network configurations are more robust to increased stimulus noise, internal noise and increases in the duration of the delay period. Overall, we provide theoretical evidence that representational drift can play a functional role in making task representations become more stable, robust, and energy efficient over time.

56.314 DISSOCIATING STRENGTH AND PRECISION IN WORKING MEMORY REPRESENTATIONS: AN INTEGRATED MODELING APPROACH

Inik Kim¹, Tianye Ma¹, Lila Liu¹, Weiwei Zhang¹; ¹University of California, Riverside

Working memory (WM) is a multidimensional cognitive process that involves dynamic activation of information. Although recent research has focused on the noisy WM representations, many studies continue to rely on a unitary measure such as either precision or strength. In the present study we have developed an integrated model with two orthogonal components of WM representations, strength and precision, by modifying the Target Confusability Competition (TCC) model. To test our hypothesis, two experiments were conducted using a Random Dot Motion (RDM) task in which participants remembered and then estimated the direction of moving dots. Experiment 1 manipulated Gaussian bandwidth noise of motion stimuli and showed that increasing noise largely reduced WM precision, but not strength. Experiment 2 manipulated the set size and coherence of RDMs and the results showed that decreasing coherence of RDMs largely reduced WM strength, while precision remained relatively comparable. In summary, these findings demonstrate that WM representations can be computationally and experimentally dissociated into separate dimensions of strength and precision. This work highlights the importance of accurately characterizing the structure and limitation of WM representations.

56.315 THE EFFECTS OF FRONTO-PARIETAL ALPHA TACS ON VISUAL SHORT-TERM MEMORY CAPACITY

Sahereh Varastegan¹, Weiwei Zhang²; ¹Ph.D. Student at University of California, Riverside, ²Professor of Psychology at University of California, Riverside

Converging evidence suggests that alpha-band activities within the fronto-parietal networks are linked to the limited capacity of Visual short-term memory (VSTM). To provide causal evidence supporting the roles of fronto-parietal alpha synchronization for VSTM, the present study tested whether the phase relationship between frontal and parietal alpha oscillations affect VSTM capacity. Healthy adults performed a lateralized delayed-estimation task with a continuous color-wheel report while receiving bifocal alpha-frequency transcranial alternating current stimulation (tACS) over fronto-parietal electrode pairs. On each trial, stimulation within the left or right hemisphere was delivered either in-phase or anti-phase across frontal and parietal sites. Recall errors were analyzed with the mixture model to dissociate capacity and mnemonic precision. The findings from 26 participants showed that in-phase fronto-parietal tACS reliably increased VSTM capacity relative to anti-phase stimulation, and this enhancement was selective for items presented contralateral to the stimulated hemisphere. In contrast, mnemonic precision was comparable across conditions. These findings provide causal evidence that phase-aligned fronto-parietal alpha activity selectively increases the number of items that can be stored in VSTM, without substantially altering mnemonic precision of those representations.

56.316 STABLE INDIVIDUAL DIFFERENCES IN STRATEGIC VISUAL MEMORY USE

Trevor Bissert¹ (trevor.bissert001@umb.edu), Candice Koolhaas², Zsuzsa Kaldy³, Erik Blaser⁴; ¹University of Massachusetts Boston

Naturalistic visuomotor tasks, like assembling a new bookshelf from a diagram, allow individuals to modulate their use of external resources (e.g., looking at the diagram) versus internal resources (memorizing steps); a sampling-remembering trade-off (Ballard et al., 1995; van der Stigchel, 2020; Liang et al., 2025). There are large individual differences in the strategies participants use in the sampling-remembering trade-off (Koolhaas et al., 2025) - here we tested whether these individual differences are stable across tasks. We used two games that were superficially different but identical in underlying task structure. In our tablet-based 'Shopping Game' and 'Zoo Game', participants (N=30, ages 18-45, 83% female) were given a list of 10 food items (or animals, in Zoo Game) and asked to locate them in a store (or field) of 20 items (10 target items plus 10 distractors). The list and the store (or field) were not visible simultaneously, but participants could toggle freely between them. We measured participants' study time of the list per visit as a measure of their cognitive strategy: the more they choose to encode in their working memory (internal resource), the longer they will spend studying the list. Study times per block ranged from 1.05s to 11.1s (M=4.1s, SD=1.8s) and, despite these large individual differences, our results show that participants were highly stable in their study time both across tasks (Kendall's tau = 0.57, $p < .001$) and over time in the same task (tau = 0.52, $p < .001$): people who prefer to remember a lot and sample less tend to stay 'heavy-loaders' while those that prefer a lighter load stay light-loaders. We are also using task-evoked pupil responses (using a Pupil Labs Neon eyetracker) and a set of metacognitive debriefing questions to determine the implicit and explicit factors affecting these strategies.

The project was supported by NIH R15HD115244.

56.317 NOVELTY SEEKERS VS. FAMILIARITY SEEKERS: INDIVIDUAL DIFFERENCES IN MEMORY-DRIVEN VISUAL PREFERENCE AND THEIR PSYCHOLOGICAL CORRELATES.

Eiko Shimojo¹, Shengjie Zheng¹, Shinsuke Shimojo¹; ¹Caltech, Division of Biology & Biological Engineering / Computation and Neural System

People differ in whether they seek familiar or novel experiences in daily life. Visual preference is also shaped by memory: our prior work shows familiarity preference for faces and novelty preference for natural scenes (Park et al., 2010). Here, we asked how individual differences in familiarity/novelty (F/N) preference in a visual task relate to daily-life F/N tendencies and psychological traits. 46 participants completed a 2AFC preference task using image pairs from three categories (Faces, Natural Scenes, Geometric Figures). Familiar images were the median-rated stimuli from a prior evaluation, selected as the "old" stimulus for each participant, and paired with a randomly selected "new" stimulus on each trial. Participants also completed an eight-domain daily-life F/N questionnaire (food, activities, games, etc.), along with the Autism-Spectrum Quotient (AQ), Flow propensity, Maemuki (positive attitude), and Negative Capability scales. We replicated group-level patterns (familiarity preference for faces, novelty

preference for natural scenes) and observed substantial individual variability. Participants could be classified into "F/N types" based on their patterns across Faces–Natural Scenes–Geometric Figures (e.g., F–N–N), and these types showed distinct profiles on AQ, Flow, Maemuki, and daily-life F/N seeking, suggesting a potential new personality-profiling axis grounded in memory-driven preferences. Key correlations included: AQ positively related to familiarity preference for geometric figures ($r = .24$) and for food/cuisine and restaurant/snack choices (both $r = .35^*$); Flow positively related to novelty preference for natural scenes ($r = .27$); and Maemuki positively related to familiarity preference for faces ($r = .22$). Higher AQ was associated with lower Maemuki ($r = -.31^*$), stronger daily-life familiarity seeking ($r = -.41^{**}$), and Flow was strongly associated with Maemuki ($r = .62^{**}$). However, overall laboratory F/N indices did not correlate with daily-life F/N tendencies ($r = -.08$), suggesting distinct mechanisms underlying memory-based visual preferences vs. real-world exploration–exploitation.

This work was supported by JST [MOONSHOT R&D Program: JPMJMS2295-03] .

56.318 HIGH-LEVEL VISUAL DISCRIMINATION ABILITIES PREDICT MEMORY FOR IMAGE LOCATION AS A FUNCTION OF MEANINGFULNESS.

Conor Smithson¹, Nurit Gronau², Isabel Gauthier¹; ¹Vanderbilt University, ²The Open University of Israel

Individuals differ in the level of detail they can visually encode. Prior work (Gronau et al., submitted) suggested that a domain-general perceptual ability (Op) for discriminating among object identities predicts memory for contextual detail, specifically, where an object was seen. Additionally, Op predicted the 'meaningfulness effect', wherein the locations of objects that were independently rated as highly familiar were better remembered than the locations of objects rated as rare or difficult to recognize. Here, we extend this work by experimentally manipulating meaningfulness by changing the level of processing of the images. Storybook images were encoded during a repetition detection task in which participants attended to either the artist style or the deeper semantic content. We tested whether Op predicted the meaningfulness effect on memory for location as well as for temporal position within a sequence. We controlled for low-level visual ability, working memory, and intelligence. Although preregistered structural equation models were inconclusive due to imprecise parameter estimates, multiple regression analyses replicated the predictive effect of Op on memory for location, as well as the meaningfulness effect on location memory. Op did not, however, predict memory for sequence position, nor the meaningfulness effect on this type of memory. Domain-general scene perception ability was an even stronger predictor, capturing all that was accounted for by Op, as well as additional variance, likely because it also captures variance in the ability to use color, texture, and spatial relations between objects. Scene perception also predicted memory for sequence position, but like Op, it did not predict the meaningfulness

effect on memory for sequence position. High-level visual discrimination abilities appear to predict the meaningfulness effect on where stimuli were seen, but not on when they were seen. This is consistent with previous research showing that object location and object identity are not perceived entirely independently.

This work was supported by the David K. Wilson Chair Research Fund and NSF BCS Award 2316474

56.319 UNIVERSAL PATTERNS AND INDIVIDUAL DIFFERENCES IN VISUAL MEMORABILITY

Harry Green¹ (harryg@umd.edu), Chaoxiong Ye, Lijing Guo, Srihitha Akula, Alena Zheng, Heidi Lin, Nandani Sharma, Weizhen Xie; ¹Department of Psychology, University of Maryland, College Park, MD, USA, ²Department of Psychology, University of Jyväskylä, Finland

Visual recognition memory shows striking consistency across observers, a pattern often attributed to stimulus memorability – an intrinsic property of images predicting how likely they can be recognized. Yet, whether this consistency reflects only stimulus-driven factors or also stable individual differences remains unresolved. We examined this question using a continuous recognition paradigm with faces from the 10k U.S. Adult Faces Database, while assessing participants' individual differences in response criteria during memory-based decisions. Across two samples ($n = 146$, U.S.; $n = 240$, China), participants viewed the same set of 160 critical faces that repeated only after longer delays ranging from 30s up to 40min, intermixed with filler images and vigilance repeats within 2s to 30s. In both samples, participants correctly responded to the same subset of repeated faces even under randomized test orders. This consistency was strongly predicted by established memorability scores, confirming a robust stimulus-driven component. However, participants' correct detection of repetition systematically varied based on their response criteria. Individuals with more liberal response criteria, such as those with lower self-confidence and high psychosis traits, showed inflated rates of correct repetition detection. Their apparent "better memory" may reflect decisional bias rather than superior mnemonic encoding or higher stimulus memorability. These findings demonstrate that cross-observer consistency in visual recognition arises from the joint influence of stimulus-driven memorability and individual differences in response criteria. Conventional memorability estimates based on correct repetition detection rate (i.e., hit rate) is not solely a property of the images, but also a function of who is making the recognition decision. Together, these results underscore the need to account for individual differences in observer traits when interpreting "universal" patterns of visual recognition, and highlight that computer vision models trained exclusively on population-level hit rate to predict visual memorability may inherit systematic biases if individual differences are ignored.

56.320 CONTINUOUS AND CATEGORICAL REPRESENTATIONS IN WORKING MEMORY FOR FACIAL EXPRESSION

Jessica Golding¹, Tianye Ma¹, Weiwei Zhang¹; ¹University of California Riverside

Facial expressions are represented with both continuous sensory fidelity and categorical semantic labeling. While classical models debate the primacy of these representations, recent evidence suggests that the balance between continuous and categorical encoding is malleable, influenced by top-down biases and individual differences. The present study investigated individual differences in working memory (WM) for facial expressions and their potential relationship with psychopathology. We collected mouse-tracking data from 32 participants during a continuous estimation task for facial expression. Stimuli consisted of morphed faces from the NimStim set, forming a continuous valence half-circle response scale from 150% Angry to 150% Happy. Participants recalled expressions on the continuous scale. Analysis of mouse trajectories revealed significant early mu-shifts toward negative valence, with final responses showing a robust negative bias. Next, we integrated the Category Adjustment Model (CAM, prototype-biased) into the Target Confusability Competition (TCC, linear similarity-based) and developed 10 candidate models to capture categorical and continuous WM representations. While half of the participants were best fit by the continuous TCC-Linear model, the other half exhibited performance favoring categorical encoding models. Crucially, these strategy differences tracked with variations in depression, anxiety, and schizotypy. We propose that while normative facial processing preserves graded fidelity (TCC), psychopathological traits may potentiate top-down categorical warping (CAM), specifically strengthening negative attractors. These findings unify psychophysical models with emotional bias, demonstrating that the "geometry" of memory is reshaped by the observer's emotional state.

56.321 INDIVIDUALS USE ENCODING STRATEGIES TAILORED TO THEIR ABILITIES

Yin-ting Lin¹ (lin.3913@osu.edu), Andrew B. Leber¹; ¹The Ohio State University

Visual working memory only maintains limited information. One way we can overcome this limitation is to selectively encode information relevant to the ongoing task. Alternatively, we can use strategies to maximize performance, yet relatively fewer studies have investigated this. Here we examined what individuals choose to encode when the available information exceeds their capacity limits. In the choice task, participants viewed six colours and had to report the colour of three items on a continuous wheel. Importantly, we allowed participants to choose which three items they report. Across three experiments, results show that strategy use varied across individuals. Some participants chose to report items that share high colour similarity (colour clustering). Others chose to report items in certain spatial locations (spatial clustering), and many of these participants preferred items from the upper part of the display. Interestingly, there was a strong trade-off between colour and spatial clustering, with many individuals using a dominant strategy. We further examined whether

these individual differences reflect idiosyncratic preferences or differences in ability. Specifically, we also asked participants to perform a cued task, where they were pre-cued to report three items in the display. We found that many participants show better performance for some locations, and that participants' preferred locations in the choice task match the locations with best memory performance in the cued task. Moreover, participants who showed larger benefits for similar versus dissimilar colours were more likely to use colour clustering strategies. Furthermore, we trained a support vector regression model using cued task data and found that participants tend to choose items that yield the best predicted memory performance, suggesting that participants use strategies that maximize their own performance. Overall, our findings suggest that there are large individual differences in encoding strategy, and individuals appear to be using strategies tailored to their abilities.

This work was supported by NSF BCS-2021038 to ABL.

Face and Body Perception: Neural mechanisms

TUESDAY, MAY 19, 2:45 – 6:45 PM, BANYAN BREEZEWAY

56.322 PERCEPTOGRAPHY OF FACE NEURONS.

Elia Shahbazi¹, Drew Nguyen², Sophie Ledet¹, Rasel Ahmed Bhuiyan³, Adam Czajka³, Reza Azadi¹, Arash Afraz¹; ¹National Institutes of Health, ²University of Chicago, ³University of Notre Dame

We recently developed a paradigm called perceptography to visualize complex perceptual distortions induced by local brain stimulation. Perceptography uses machine-learning to create and optimize image distortions that are hard for the animal to distinguish from the state of being cortically stimulated. In our first project, we used optogenetics for brain-stimulation, which emphasized the concept but lacked simultaneous recording from stimulated neurons. Here and in our next move, using neural recording, we targeted specific groups of neurons within the inferior temporal (IT) cortex, including face selective neurons. We used two Utah-Array implants, one in the middle face-patch and one in the central IT-cortex. Macaque monkeys were trained to perform a cortical stimulation detection task, requiring fixation on a computer-generated image for 1s. At the midpoint of image presentation, we altered the image for 200ms and delivered an electrical stimulation impulse in 50% of the trials at random for 200ms. After the image, two targets appeared, and the animals indicated whether the stimulation occurred, earning liquid reward for correct reports. Our AI-generated framework, Perceptographer, enabled fine-grained perceptography for both face and non-face objects. Cortical stimulation in the face-patch produced perceptograms in which faces emerged for both face and object seed images, whereas stimulation outside the face-patch did not produce faces in object perceptograms.

Furthermore, faces that emerged during face-patch stimulation appeared only contralateral to the stimulated hemisphere, while stimulation outside the face-patch produced whole-image distortions for both face and object seeds. These results indicate that representation of faces in IT-cortex is strongly contralateral, in contrast to broader bilateral representation of nonface objects. Neural recording corroborated these effects. Contralateral distortion/presentation elicited more vigorous neuronal activity in face neurons with face stimuli, whereas outside the face-patch, central bilateral image presentation/distortion elicited more robust responses for both face and object stimuli.

56.323 INTRACRANIAL EEG AND EYE-TRACKING REVEAL PARAFOVEAL PREVIEW CONTRIBUTIONS TO FACE PROCESSING DURING NATURALISTIC VISION

Casey Becker¹, Mary Kate Richey¹, Arish Alreja¹, Jorge Gonzalez-Martinez¹, J Patrick Mayo¹, Avniel Ghuman¹; ¹University of Pittsburgh

Faces rarely appear suddenly in our central vision (fovea); instead, specialised neural systems detect faces in our periphery and direct our attention towards them. The blurry parafoveal preview of a face recruits prediction mechanisms that update when the fixation lands on the target face. And yet, much of what we know about the neuroscience of face perception relies on paradigms where faces are presented abruptly at central fixation, overlooking the dynamic role of eye movements in natural vision. Here, we examine the role of the parafoveal preview on faces and houses, using high-speed eye-tracking and intracranial EEG (depth electrodes). In a gaze-contingent paradigm, we compared traditional centrally presented faces and houses with parafoveally presented ones. Participants freely looked towards faces, houses, or blank squares that appeared in their parafovea. Parafoveal faces and houses could remain the same (valid preview), turn into blank squares (face-to-blank), or vice versa (blank-to-face). As expected, neural activity from faces that suddenly appear in the fovea differs from activity induced by previewing. This did not greatly differ between centrally presented faces and those that appeared during a saccade towards a blank square, suggesting similarity between central onset and an invalid preview. Interestingly, when an object disappears during a saccade, responses are similar to valid previews, suggesting much processing occurs before fixation. Responses to faces, but not houses, somewhat differ when the stimulus disappears during saccade, suggesting differences in predictive processing.

56.324 ONE FACE, MANY PATHS: TASK-DEPENDENT NEURAL ROUTING OF IDENTICAL FACIAL MOVEMENTS INTO DIFFERENT MEANINGS

Yu Chen¹ (y.chen.8@research.gla.ac.uk), Yuening Yan¹, Oliver G. B. Garrod¹, Robin A.A. Ince¹, Rachael E. Jack¹, Philippe G. Schyns¹; ¹University of Glasgow

Facial expressions are multi-component signals composed of Action Units (AUs), from which humans infer both discrete emotion categories

(e.g., anger) and affective dimensions (e.g., negative valence, high arousal). Yet the neural mechanisms through which the brain extracts these distinct signals remain poorly understood. Using a generative face model, we independently sampled eight emotion-relevant AUs to create 1,600 animations rendered on random 3D identities. Each AU was either absent (0 intensity, 50% probability) or present with a random non-zero intensity between 0 and 1 (maximum perceptual visibility), enabling us to trace AU-specific neural representations in 8,196 cortical MEG sources. Participants viewed all animations and completed three 7-AFC tasks: emotional category (“happy”, “fear”, “surprise”, “anger”, “disgust”, “sad” and “other”), valence (negative to positive), and arousal (low to high). Each participant saw the same 1,600 animations in each task (dense-sampled, 4,800-trial design), allowing us to isolate task-dependent effects on neural representations. We recorded whole-brain MEG and reconstructed source activity at 8,196 cortical locations. AU representations were quantified using mutual information between single-trial AU intensities and source-level MEG amplitudes every 4ms. After animation onset, AUs were first represented in occipital cortex (~210ms) and then propagated along the third pathway to posterior STS and STG (~310ms), replicated across all participants. Whereas occipital representations were largely task-invariant, pSTS/STG showed systematic task-dependent reweighting. For example, 3×2 ANOVAs (Task×Hemisphere) of one participant revealed stronger STG representations (e.g., Brow Lowerer and Jaw Drop) during category task (e.g., responding “surprise”), with distinct lateralization in pSTS—rightward for Brow Lowerer and leftward for Jaw Drop (relative to other tasks and hemispheres, $p < 0.05$). Thus, while early visual representations of facial movements are task-invariant, higher-level pSTS/STG flexibly and laterally reweights identical facial movement inputs depending on whether observers judge categories or dimensions. This dynamic routing supports divergent emotional interpretations from the same facial inputs.

This work was funded by the Wellcome Trust (Senior Investigator Award, UK; 107802) and the Multidisciplinary University Research Initiative/Engineering and Physical Sciences Research Council (USA, UK; 172046-01) (P.G.S.); ERC [FACESYNTAX; 759796] (R.E.J.); the Wellcome Trust [214120/Z/18/Z] (R.A.AI.).

56.325 PUSHING THE EYES TO THE LIMIT: STABLE N170 SENSITIVITY TO EYES DESPITE INCREASED DIAGNOSTICITY IN EXPRESSIVE FACE JUDGMENTS

Chanelle Demeule¹, Anthony Proulx¹, Isabelle Charbonneau¹, Justin Duncan², Allyson Bastien¹, Caroline Blais¹, Daniel Fiset¹; ¹Université du Québec en Outaouais, ²Université d'Ottawa

Previous work suggests that early perceptual brain activity in face recognition, indexed by the N170, is jointly influenced by sensitivity to the eye region and the processing of diagnostic, task-relevant information (Schyns et al., 2007; Smith et al., 2004). However, these two factors are often confounded because the eyes typically carry diagnostic information. In earlier work, we dissociated them by

equating the eye region across neutral/smiling versions of the same identity and showed that N170 amplitude tracked the eyes and the mouth. Here, we asked whether making the eyes more diagnostic would further boost this effect. We recorded EEG from 29 adults (19 women) performing an expressive/neutral (EXNEX) categorization task while viewing faces revealed through randomly located Gaussian apertures (“Bubbles”; Gosselin & Schyns, 2001). The design included two stimulus sets: (1) faces with neutral eyes identical in the smiling and neutral versions, and (2) faces with a Duchenne smile in which the eyes were markedly more expressive in the smiling than in the neutral version. Time-resolved classification images were computed at PO8 from -200 to 800 ms post-stimulus to identify pixels associated with N170 amplitude on a participant basis. In both sets, pixel tests ($p \leq .025$; Stat4Ci Toolbox; Chauvin et al., 2005) revealed significant associations between N170 amplitude and both eyes and the mouth. Critically, these associations did not differ between sets, even when analyses were restricted to the 12 participants who behaviorally relied most on the eyes. This pattern suggests that N170 sensitivity to the eyes may already operate near ceiling, such that increasing their diagnostic value fails to further enhance the response, whereas the strong involvement of the mouth—highly diagnostic for the present expressivity judgment but not for tasks where it is unnecessary—illustrates how goal-driven diagnostic information can shape processing without further boosting this early eye-driven activity.

56.326 NEURAL CORRELATES OF FACE LEARNING: THE N250 DISTINGUISHES NEWLY-LEARNED FROM SIMILAR-LOOKING NOVEL FACES

Alice Nevard¹, Soazig Casteau¹, Ulrik Beierholm¹, Holger Wiese¹; ¹Durham University

Learning a new face can be described as gradual formation of a robust representation that supports recognition of an individual in novel contexts. Previous event-related potential (ERP) research has demonstrated that more negative amplitudes from approximately 200 ms onwards are associated with newly learnt faces compared to novel ones, suggesting a robust neural learning effect. The current study examines how representations of newly learnt identities are formed following a brief real-life interaction, compared to viewing novel matched (i.e. similar-looking) and unmatched (i.e. dissimilar) identities. Participants had a 10-minute conversation with one of four individuals they had not met previously. Participants then saw multiple ambient images with high within-person variability, of the “learnt”, an “unmatched” and a “matched” face while their EEG was recorded. N250 amplitudes (200-300 ms, occipito-temporal channels) were more negative for the learnt face compared to both the unmatched and matched identities, indicating the emergence of a distinct neural representation for the newly learnt face. Additionally, no differences were found between the matched and unmatched novel identities, indicating that both novel faces were similarly differentiated from the learnt identity. Furthermore, this ERP learning effect was carried over into later time windows showing sustained familiarity effects, and

learning effects were generally stronger over the right hemisphere after 300 ms. Together these findings indicate that brief real-world interactions allow to establish robust neural representations that support the recognition of newly learnt faces – even in comparison to similar-looking novel identities.

56.327 RELATIVE SALIENCE DETERMINES WHETHER WE REPRESENT TWO FACES SIMULTANEOUSLY: EVIDENCE FROM EVENT-RELATED BRAIN POTENTIALS

Holger Wiese¹ (holger.wiese@durham.ac.uk), Linda H. Lidborg¹, A. Mike Burton^{2,3}; ¹Durham University, UK, ²University of York, UK, ³Bond University, Australia

Research suggests that our resources for representing multiple faces simultaneously are severely restricted. Yet, in our daily life we are often confronted with more than one person simultaneously, a situation which usually does not seem to pose substantial problems to the face processing system. Here, we used event-related potentials (ERP) in combination with repetition priming to test under which conditions two faces can be represented simultaneously. All experiments presented two faces as primes and a single face as the target. When two different familiar face primes were presented, the repetition of either elicited ERP priming effects, consisting of more negative amplitudes at occipito-temporal channels with a maximum at approximately 400 ms after target onset. This effect was observed when the two primes were presented at equal distances relative to fixation (E1), and it neither depended on the predictability of prime location (E2), nor on the physiognomic similarity of the prime identities (E3). These findings suggest that two faces are represented simultaneously when processing resources can be divided between two equally salient faces. However, as shown in further experiments, if one prime is presented at a central location while the other prime is presented either left or right of fixation, this flanker stimulus does not elicit priming (E4). This absence of flanker priming is independent of whether the central stimulus is familiar or unfamiliar (E5). Accordingly, a face, appearing in a peripheral and unpredictable location, can be ignored. Finally, subsequent experiments did detect priming from a flanker face when the central stimulus was a written name (E6) or when the flanker was task-relevant (E7), ruling out reduced accuracy in peripheral vision as an explanation for absent flanker priming in E4/E5. Together, these findings demonstrate the flexibility of the face recognition system to prioritize the most salient information.

Funded by the Economic and Social Research Council, UK (ES/X002063/1)

56.328 LATERALIZED REPRESENTATION OF CATEGORICAL FACE INFORMATION IN FUSIFORM CORTEX

Samantha C Lee¹ (samanthalee@unr.edu), Lars Strother¹; ¹University of Nevada, Reno

Our visual system allows us to rapidly and effortlessly perceive identity-related face information (e.g., sex and age) as well as non-identity information (e.g., emotional expression). We used fMRI multivariate pattern analysis (MVPA) to study how both types of information are represented neurally. Of particular interest in our study was whether the neural representation of identity-related and non-identity information was similarly lateralized in fusiform cortex (FC). Many studies have reported right-lateralized representation of face identity information, but some recent studies have refuted this. Other fMRI studies have reported left-lateralized representation of categorical face information (e.g., sex; Zhou & Strother, VSS 2018). We wondered if other types of face information would also exhibit left-lateralized representation in FC, and whether this would be limited to identity-related face information (i.e., in addition to sex) or not. Observers in our experiment performed a one-back task while viewing 16 different faces comprising four binary categories: (1) two distinct (unfamiliar) identities; (2) sex (female/male); (3) age (old/young); and (4) emotional expression (happy/sad). Our method enabled us to determine whether all four categories could be decoded in FC and the degree to which underlying representations might be lateralized. We found that all four categories were decodable in FC but exhibited different patterns of laterality. Whereas left-lateralization was abundant (i.e., occurred for various categories and category combinations), right-lateralized decoding only occurred for age (or age combined with other categories). Emotional expression decoding was not associated with consistently lateralized representations. Our results suggest that left-lateralized face representation can be revealed by creating binary categorical face information. Nevertheless, right-lateralized representation of identity-related information can occur even when such information is treated categorically (despite being continuous) within our set of face stimuli. Our findings highlight the need for further research into the relationship between laterality in FC and categorical face perception.

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under grant number P30 GM145646.

56.329 NEURAL REPRESENTATIONS OF DYNAMIC FACIAL EXPRESSIONS IN RSTS REFLECT PERCEIVED EMOTIONAL AND SOCIAL SIMILARITY

Hilal Nizamoğlu¹, Katharina Dobs¹; ¹Justus Liebig University in Giessen

The right superior temporal sulcus (rSTS) is critically involved in processing dynamic social cues, yet how it encodes complex facial expressions remains unclear. Here, we investigated how well both stimulus-derived features and perceptual similarity judgments predict neural representations of dynamic facial expressions in this region. We recorded fMRI data while participants viewed 48 video clips (4 actors × 12 expressions) depicting emotional and conversational facial expressions. As model predictors, we used perceptual similarity judgments from a behavioral experiment and independent ratings of

emotional (valence, arousal, affectiveness), social (communicativeness, friendliness), and motion-based properties for the same stimuli. Using representational similarity analysis (RSA), we tested whether these models could explain multivoxel response patterns in rSTS. Preliminary results (N=10) showed that rSTS activity patterns were best predicted by arousal, affectiveness, communicativeness—and critically, by behavioral similarity judgments. Among all predictors, only the behavioral model uniquely explained variance in rSTS, suggesting that this region encodes a perceptually grounded, multidimensional representation organized by social-emotional meaning. Our findings reveal an alignment between neural and perceptual representations of dynamic facial expressions, bridging behavior and brain.

56.330 THE TEMPORAL POLE AS THE CORTICAL ENDPOINT OF THE LATERAL FACE PROCESSING PATHWAY

Abigail Gantz¹, Ben Deen¹; ¹Tulane University

Face processing has been argued to rely on two distinct neural pathways: a ventral pathway within ventral temporal cortex, and a lateral pathway within the superior temporal sulcus (STS) specialized for processing dynamic information for social perception. Within both pathways, visual information proceeds along a posterior-to-anterior path from early visual areas in occipital cortex to the anterior temporal lobe. However, the anterior-most part of this system has been difficult to study using neuroimaging, due to magnetic susceptibility artifacts reducing signal quality in the anterior temporal lobe. Recent work using advanced fMRI methods has identified a face-selective area in the temporal pole (TPFA). Here we ask how the TPFA is situated with respect the broader functional architecture of the face processing system, and whether it belongs to the ventral or lateral pathway. In N=10 participants, we tested whether the TPFA has signature functional properties of the lateral pathway face areas: a stronger response to dynamic over static faces, and an auditory response to voices. To test sensitivity to motion content, participants viewed dynamic videos of faces or scenes as well as static images taken from video frames. To test for a voice response, participants listened to audio clips of speech (sentences and nonword lists) in addition to a synthetic non-vocal control with matched acoustic features. Results (region-of-interest analysis) showed that TPFA showed a preference for dynamic over static faces, as well as a preference for speech over synthetic sounds. The response of TPFA across conditions was closer to face areas in the STS than to ventral stream areas such as the occipital and fusiform face areas. Given the functional profile demonstrated here, and the anatomical positioning of the TPFA at the anterior tip of the temporal lobe, we propose that TPFA constitutes the cortical endpoint of the lateral face processing pathway.

56.331 SHARED AND DIVERGENT NEURAL CODES FOR FACE IDENTITY IN PERCEPTION AND IMAGERY

Shaofeng Liu¹, Eric Nemrodov², Kinkini Monaragala¹, Ilya Nudnou³, Dan Nemrodov¹, Matthias Niemeier¹, Adrian Nestor¹; ¹University of Toronto, ²University of Waterloo, ³North Dakota State University

Visual imagery is thought to recruit some of the same neural machinery as perception, but how identity-level representations unfold over time during imagery, and how they relate to perceptual processing, remains unclear. Here we used EEG and multivariate decoding to compare the spatiotemporal profiles of face perception and imagery. Seventeen participants viewed 25 famous and 25 unfamiliar female faces shown upright or inverted, and later generated mental images of a subset of five highly familiar identities cued by their names. We used temporally cumulative and time-resolved analyses to decode face identity, along with multivariate channel selection, representational similarity analysis, and cross-temporal generalization to characterize the underlying code. Identity could be reliably decoded from neural signals elicited by both perceived and imagined faces. Perceptual decoding emerged early and was modulated by both face inversion and familiarity, which contributed largely independently to performance. Imagery-based decoding showed a markedly later onset and more extended time course, but relied on scalp topographies that overlapped with those supporting familiar upright face perception. Across the five imagery targets, representational dissimilarity matrices for perception and imagery were significantly correlated, indicating a partially shared identity space despite differences in temporal dynamics. Together, these findings demonstrate that EEG carries fine-grained information about individual face representations not only during perception but also during voluntary imagery, and they highlight both common and distinct organizational principles for perceptual and mnemonic face codes.

56.332 HOW MUCH IDENTITY INFORMATION IS NECESSARY FOR FAMILIAR FACE RECOGNITION? EVIDENCE FROM EVENT-RELATED BRAIN POTENTIALS

Linda H. Lidborg¹, Anna Yue Gao¹, A. Mike Burton^{2,3}, Holger Wiese¹; ¹Durham University, ²University of York, ³Bond University

Most people can recognise thousands of familiar faces, with mix-ups occurring only rarely. This is remarkable, considering that the same face can look very different under varying conditions. For example, we recognise familiar faces despite changes in lighting, viewing angle, facial expression etc. But how much facial identity information is necessary for reliable recognition? This can be tested by using morphing techniques and varying the contribution of each face to the morph. Behavioural findings using such morph images suggest that recognition is possible with a minimal amount – possibly as little as 10% – of a face's identity information preserved. However, many studies have used either unfamiliar-familiar morphs or morphs based on multiple faces, which reduces identity-specific noise and may facilitate recognition with minimal identity information. Here, we used event-related potentials (ERP) and immediate repetition priming to investigate the amount of identity information required for recognition from morphs across two identities (ID1, ID2). Primes were morphs

created from pairs of familiar celebrity faces. For each of 24 celebrity pairs, six morph continua were generated, each with six identity steps in 20% increments (0% ID1/100% ID2 to 100% ID1/0% ID2). Target images immediately followed the primes and displayed full identity information. We observed clear ERP priming effects from 300 ms onwards at occipito-temporal electrodes, with more negative amplitudes for 100%, 80%, and 60% morphs, but no priming effects for 40% and 20% relative to 0% morphs. This suggests that the critical threshold for recognition lies between 60-40% identity information. Moreover, the 'majority' identity inhibited the processing of the 'minority' identity, consistent with an inhibitory mechanism allowing only the processing of the most salient facial identity.

This work was supported by the Economic and Social Research Council (ESRC; grant number ES/X002063/1) to HW and AMB.

56.333 EYES ON THE FFA: ABILITY-DEPENDENT EFFECTS OF HD-TDCS ON FACE RECOGNITION AND LEFT-EYE INFORMATION USE

Daniel Fiset¹ (daniel.fiset@uqo.ca), Kim Calvé¹, Jérémy Lamontagne¹, Marie-Claude Desjardins¹, Mélodie Potvin-Poirier¹, Caroline Blais^{1,2}, Sara Tremblay; ¹Université du Québec en Outaouais, ²The Royal's Institute of Mental Health Research

Previous attempts to modulate face-recognition performance with non-invasive brain stimulation over the right FFA have yielded mixed results (e.g., Brunyé et al., 2017; Wu et al., 2021; Kho et al., 2023). We asked whether high-definition tDCS (HD-tDCS) over right FFA changes face-recognition performance and the use of diagnostic information as a function of individual ability. Thirty adults were randomly assigned to anodal, cathodal, or sham HD-tDCS (N = 10 per group) targeting right FFA with a 5-electrode montage. In Session 1, participants completed the Cambridge Face Memory Test – Australian version (CFMT-Aus). In Session 2, they received offline HD-tDCS or sham and then performed the CFMT+ and 700 identity-recognition trials using the Bubbles technique (Gosselin & Schyns, 2001), in which faces are revealed through randomly positioned Gaussian apertures, allowing us to derive classification images indexing the use of visual information (in terms of facial features). Group-level ANOVAs on CFMT change scores (post-pre) showed no reliable difference between stimulation and sham conditions [$F(2,27) = 0.62$, $p = 0.55$]. However, regression analyses revealed a robust interaction between baseline CFMT and stimulation: in both anodal and cathodal groups, poorer recognizers improved after HD-tDCS, whereas better recognizers tended to worsen; no such pattern emerged under sham [$F(1,18) = 13.00$, $p = 0.002$, $\beta = -0.65$, $r^2 = 0.42$]. We then compared classification images across groups by matching participants on ability. Among low-ability observers, HD-tDCS over right FFA (cathodal or anodal) increased left-eye use relative to low-ability observers in the sham group, whereas among high-ability observers, HD-tDCS decreased left-eye use relative to high-ability sham participants, with no systematic change for the right eye. These findings provide evidence that FFA modulation reconfigures face-

recognition performance and visual information use in a state-dependent manner, consistent with models where tDCS perturbs excitation-inhibition balance rather than exerting polarity-specific effects.

Natural Sciences and Engineering Research Council of Canada

Face and Body Perception: Emotion

TUESDAY, MAY 19, 2:45 – 6:45 PM, BANYAN BREEZEWAY

56.334 PERCEPTUAL STRATEGIES IN EMOTION RECOGNITION ACROSS LEVELS OF AUTISTIC TRAITS: AN EYE TRACKING STUDY

Golnoosh Soroor¹, Alexandra McGowan¹, Amaya Aquino Aquino¹, Nicolas Davidenko¹; ¹University of California, Santa Cruz

Emotion recognition is crucial for functional social interaction. In some conditions, such as autism spectrum disorder (ASD), difficulties in emotion recognition have been reported and may be related to different gaze strategies (e.g., eye avoidance) and atypical perceptual processing, including a local processing bias. However, it is not clear whether variation in autistic traits across the general population predicts gaze and emotion recognition differences. In this study, ninety-five neurotypical participants completed an emotion recognition task evaluating carefully controlled face stimuli across the six basic emotion categories (anger, fear, disgust, happiness, sadness, and surprise). Before each block of trials, participants completed a Navon task designed to prime global processing, local processing, or intermediate processing (control). Eye tracking was used to measure gaze patterns across six basic emotions, including fixation distribution and dwell time in different facial regions. Participants then completed the Autism Quotient (AQ; Baron-Cohen et al., 2001) to assess autistic traits. Behavioral results showed that priming did not lead to differences in accuracy or reaction time. However, both global and local priming resulted in more broadly distributed gaze patterns, reflected by higher standard deviations of fixation locations, compared to the control condition. Across all participants, fixations were consistently more focused on the eyes than the mouth, regardless of AQ scores. However, participants with higher AQ scores (particularly in the attention-switching subscale) spent less time fixating on non-face areas. Overall, our findings suggest that the variability of AQ scores across the general population may not be sufficient to detect behavioral or gaze differences in emotion recognition previously shown in autism. Future research on emotional processing differences across the general population may need to examine emotion processing in more naturalistic contexts using dynamic faces.

56.335 DISSOCIATING MENTAL REPRESENTATION FROM PERCEPTION IN GENDERED PAIN AND

EMOTIONAL FACIAL EXPRESSIONS: A REVERSE CORRELATION STUDY USING GENETIC ALGORITHM AND MACHINE LEARNING

Arianne Richer¹, Jeanine Ohene-Agyei², Pénélope Pelland-Goulet¹, Alexis Bellerose¹, Zohair Mharchat¹, Michèle D. Berthaud¹, Francis Gingras^{1,3}, Émilie St-Pierre¹, Camille Saumure⁴, Roberto Caldara⁴, Daniel Fiset¹, Caroline Blais¹; ¹University of Quebec in Outaouais, Canada, ²University of California Davis, USA, ³University of Quebec in Montreal, Canada, ⁴University of Fribourg, Switzerland

Facial expressions are used as cues to infer others' affective states. However, observers often make mistakes when identifying facial expressions, especially for pain facial expressions (PFE). Moreover, these difficulties are exacerbated when PFE are expressed by women. Drawing on evidence that imprecise mental representations (MR) of emotion can induce perceptual errors, we hypothesized that difficulties in adequately identifying PFE, especially on female faces, may emerge during MR formation. To investigate potential gender differences in the MR of PFE, 49 participants (27 women) completed a genetic-algorithm-based reverse correlation task. This procedure generated individual MR of pain and six basic emotions (anger, disgust, joy, sadness, surprise and fear) on both male and female faces (using MakeHuman software). We analyzed differences in the facial movements captured in these MR proxies and collected independent observer (20 observers, 10 women) ratings of perceived intensity of pain and six basic emotions. Four machine learning models (Decision tree, Support Vector Machine, K-Nearest Neighbors & Naive Bayes) trained on facial movements of the MR collected revealed no significant distinctions between male and female PFE (performing at chance level), although they clearly distinguished between affective states. Furthermore, a MANOVA (7 affective states x 2 genders) revealed that independent observers perceived significantly more sadness and pain in MR of PFE of female compared to male faces. These findings suggest that people hold largely similar mental representations of what a pain expression looks like for male and female targets. The perception task, however, revealed differences in the perceived intensity of affective states for these generated proxies between the two genders.

The present study is supported by the Canada Research Chair in cognitive and social vision to Caroline Blais (#CRC-2023-00019) and by the Canada Graduate Scholarship - Doctoral program of the Natural Sciences and Engineering Research Council of Canada to Arianne Richer (#CGS D - 589787).

56.336 MAPPING HOW DYNAMIC AND STATIC FEAR AND SURPRISE EXPRESSIONS ARE RECOGNIZED ACROSS AGES

Fanny Poncet¹ (fanny.poncet@unifr.ch), Marie Smith², Roberto Caldara¹, Anne-Raphaëlle Richoz¹; ¹Eye and Brain Mapping Laboratory (iBMLab), Department of Psychology, University of

Fribourg, Fribourg, Switzerland, ²School of Psychological Sciences, Birkbeck, University of London, London, United Kingdom

Distinguishing between fear and surprise expressions is difficult, as these emotions share several overlapping facial features. Prior research shows a dynamic advantage for recognizing surprise, but not fear, and age-related differences in eye-movement strategies. However, it remains unclear how the diagnostic information used to recognize these expressions varies across ages and stimulus modalities. Here, we used the "Bubbles" reverse-correlation technique to isolate the diagnostic facial information underlying static and dynamic facial expression recognition (FER) across ages. Participants were 120 young children (6–9-years), 110 older children (10–13-years) tested in either static or dynamic condition, and 45 young adults (18–30-years) tested in both modalities. The number of bubbles was set a priori and varied by age group to balance task difficulty. All groups performed better with dynamic presentations overall. At the expression level, recognition of both fear and surprise improved with dynamic cues in older groups, and confusion between the two expressions was significantly reduced in the dynamic condition, most strongly for surprise. Diagnostic information also varied as a function of age and modality. Six- to nine-year-olds primarily relied on the mouth but started to include the eyebrow information in the dynamic modality. Ten- to thirteen-year-olds and adults showed similar patterns: static recognition involved mouth, eyes and nose (children only), while dynamic recognition incorporated the eyebrows (with specific eyebrow regions depending on the expression) alongside the mouth. Interestingly, specific diagnostic patterns were observed when examining each expression separately, interacting with both modality and age. In conclusion, dynamic facial expressions modulate facial information use and improve FER performance. Our data reveal the facial features involved in recognizing fear and surprise, highlighting the strong influence of age and presentation modality. These findings provide novel insights into the developmental trajectory underlying the well-established confusion between these expressions.

This study, was supported by the grant no 10001C_201145 from the Swiss National Science Foundation awarded to RC and ARR.

56.337 PROCESSING OF FACIAL EXPRESSIONS IN PRIMATE SUPERIOR COLLICULUS NEURONS

Gongchen Yu¹, Leor Katz¹, Richard Krauzlis¹; ¹Laboratory of Sensorimotor Research, National Eye Institute, NIH

The fast processing of faces and emotionally salient facial expressions is important for primate behavior and has long been proposed to rely on a subcortical visual pathway connecting the superior colliculus (SC), pulvinar, and amygdala. Despite the prominence of this hypothesis, there is no direct neuronal evidence (to our knowledge) for facial-expression processing in the SC. We recently reported that SC neurons show a robust preference for face images at extremely short latencies, raising the possibility that expression-specific information may also be present at these early processing stages. To test this idea,

we recorded neuronal activity from the macaque SC while presenting images with a range of facial expressions, including images shown to evoke expression-selective activity in the amygdala. We found that a subset of SC neurons (10–15%) exhibited a short-latency (~50ms) modulation by facial expressions. Most of the expression-selective neurons responded most strongly to images that contained expressions of either threat or fear. By comparison, the short-latency face preference (replicated here) was found in ~50% of our visually responsive neurons. Our findings show that the primate SC not only encodes the presence of a face, but also key aspects of its emotional expression, and that this encoding takes place at remarkably short latencies. By demonstrating early expression-selective signals within the SC, our results fill a critical missing link in the proposed subcortical pathway for rapid evaluation of emotionally salient faces and inform models of how subcortical circuits support fast social perception.

This work is supported by the National Eye Institute Intramural Research Program at the National Institutes of Health ZIA EY000511.

56.338 UPDATING THE CROWD: HOW ATTENTIONAL LOAD SHAPES ENSEMBLE EMOTION REPRESENTATIONS IN WORKING MEMORY UPDATING

Bryce C. Romero¹, Bradley E. Buchanan¹, Emily A. Rancorn¹, Laramie L.J. Starling¹, Jacob Zepp², Rachel Gaynor¹, Liana Y. Thielecke¹, Emma Sonenblum¹, Connor Murray³, Chad Dubé¹; ¹University of South Florida, ²University of Zürich, ³American University

When presenting a conference poster, the aim is to keep our audience engaged and deliver information in a clear way, to do so, we must make quick judgments of how our audience is feeling to evaluate and adapt our own behaviors. Ensemble emotional processing (EEP), or the ability to perceive a group's overall emotion, assists us in making these quick assessments of our audiences' average emotionality. However, current models of EEP offer conflicting accounts of the dynamic roles of attention and memory during ensemble judgment formation. Specifically, Liu and Ji (2024) reported no effect of attentional load, or the amount of attentional resources required for selective attention, on behavioral markers of EEP. To address this discrepancy, we conducted a replication of their dual-task paradigm while also measuring neural markers in a sample of 33 college students. In particular, we examined the P3b, an event-related brain potential associated with updating new or unexpected information in working memory through selective attentional processes, measured 300–500 ms after stimulus onset. Bayesian linear mixed-effects models revealed that high loads reduced the neural markers of working memory updating for ensemble judgments to a similar degree as for item judgments. Considering that working memory updating did not differ between item and ensemble conditions, this suggests that both types of representations operate similarly within working memory and are subject to comparable attentional constraints when updating is required.

Undergraduate Just-In-Time 3

TUESDAY, MAY 19, 2:45 – 6:45 PM, BANYAN BREEZEWAY

56.339 GLOBAL SPATIAL STRUCTURE DIFFERENTIALLY RESHAPES LOCAL CHROMATIC REPRESENTATIONS ACROSS EARLY AND LATE VISUAL WORKING MEMORY MAINTENANCE

Craig Tomlin¹ (ctomlin@rollins.edu); ¹Rollins college, Winter Park, FL

Visual working memory (VWM) is a capacity-limited system for retaining visual information over brief intervals, yet its representational format and organizational principles remain debated. To the investigator's knowledge, no known work has directly examined whether geometric boundary structure systematically modulates memory for features defined in independent dimensions across hierarchical levels, or whether such effects unfold dynamically as maintenance progresses. To test this, participants performed a whole-display color change detection task in which local chromatic elements were enclosed within parametrically varied boundary shapes (ellipse, octagon, hexagon, pentagon, rectangle) and probed at short (500 ms) and longer (1500 ms) retention intervals. Results reveal a systematic temporal dissociation in boundary geometry effects. At 500 ms, cardinal-rich geometries produced the highest discriminability ($d' \approx 1.93$ – 1.98), whereas smooth contours yielded lower sensitivity (ellipse $d' \approx 0.90$). At 1500 ms, this pattern reversed: ellipses produced the highest sensitivity ($d' \approx 2.06$), while cardinal-rich shapes declined by 32–42%. Performance change scaled monotonically with cardinal density (Spearman $\rho = -.90$) and interior angle magnitude ($r = -.98$), indicating a systematic geometric gradient. Crossover effects emerged selectively under higher memory load. Boundary detection accuracy was largely independent of chromatic memory ($r = -.16$), ruling out resource competition and spatial indexing accounts. One possibility is that boundary shapes provide a relational spatial scaffold within which chromatic elements are encoded referentially across multiple anchor points. Cardinal-rich structures may facilitate early encoding via preferential neural tuning before degrading from compensatory repulsion drift over time, which introduces higher variance into chromatic-spatial representational space which reduces discriminability. Smooth, drift-resistant contours may preserve chromatic encoding integrity as encoding transitions toward relational formats at longer delays. extend hierarchical feature-bundle and conjunctive coding models, suggesting that cross-dimensional structural influences in VWM reflect geometry-dependent organization of mnemonic representations rather than fixed capacity constraints.

56.340 NEURAL SIGNATURES OF REAL AND AI-GENERATED FACE PERCEPTION

Heaven Dixon¹, Sofia DiCostanzo¹, Keith A. Schneider¹, Adrian W. Gilmore¹; ¹Department of Psychological & Brain Sciences, University of Delaware

The increasing realism of AI-generated faces presents a challenge for the human visual system, raising questions about our ability to distinguish artificial from real facial stimuli. Research examining individuals' ability to differentiate real faces from hyper-realistic AI-generated faces suggests that behavioral performance is often at or near chance levels. Despite this behavioral limitation, EEG evidence indicates that real and artificial faces can be decoded at a neural level, with significant classification emerging as early as ~150 ms following stimulus onset (Moshel et al. 2022). However, the identity of brain regions that allow for accurate classification is not presently known. To address this gap, we used functional magnetic resonance imaging (fMRI) to examine neural responses to face stimuli in healthy adults aged 18–35. Participants viewed images of real and AI-generated faces and made a binary judgment about whether each face was real or artificially generated. AI-generated faces were produced using StyleGAN3, and the real face stimuli were drawn from the real-world face images used to train the model. Consistent with recent behavioral findings, participants performed at chance when distinguishing between real and AI-generated faces. Critically, early ventral visual regions exhibited stronger responses to real faces compared to AI-generated faces ($q < .05$). These findings suggest that early visual processing stages retain sensitivity to facial authenticity, even when subsequent behavioral responses do not achieve above-chance accuracy.

NSF Grant No. 2225805

56.341 SCHIZOPHRENIA AFFECTS DIRECTNESS AND SMOOTHNESS OF REACHING TRAJECTORIES GUIDED BY VISION AND PROPRIOCEPTION

Maya Glasman^{1,3}, Jose Reynoso^{1,2,3}, Brian Keane^{1,2,3}, Duje Tadin^{1,2,3}, Diana Freed^{1,2,3}; ¹University of Rochester, ²University of Rochester Medical Center, ³Center for Visual Science

Proprioceptively-guided reaching relies on internal sensory feedback, which is often impaired in schizophrenia. While prior research illustrates sensorimotor integration abnormalities in schizophrenia, less is known about how it affects reaching trajectories in virtual reality. Here, we assess the trajectories of reaches guided mainly by vision or proprioception. We hypothesized that individuals with schizophrenia will show less direct and less smooth reaching movements compared to our healthy control group. We used a virtual reality test to study reaching accuracy guided by proprioception or vision and recorded subjects' reaching trajectories throughout. Additionally, directness and smoothness of reaching trajectories in 19 participants with schizophrenia and 18 age-matched controls was assessed. Participants reached to tap a virtual target using their index finger with either a visible (visually-guided reach) or invisible hand (proprioceptively-guided reach). Directness was measured as the ratio

of total path length to straight-line distance between movement start and end points, with values closer to 1 indicating a more direct path. Smoothness was quantified as the R^2 of a 3D cubic fit to the movement trajectory (made to begin and end at the true endpoints), with values closer to 1 indicating smoother, less complex movements. We found that when the hand was visible during the reach movement, reach trajectories were more direct ($p=1e-05$) and smoother ($p=1e-07$) relative to proprioceptively-guided reaches. This was the case for both healthy controls (all $p<0.035$) and patients (all $p<0.0001$). We found no differences between patients and controls in reach smoothness (all $p>0.22$). However, we found a trend for less direct reach movements (all trials $p=0.08$; hand visible $p=0.12$; hand invisible $p=0.063$) in patients. Our findings suggest that visual feedback improves reaching abilities in both groups. However, those with schizophrenia show reduced movement efficiency, illustrated through less direct trajectories, even when smoothness is preserved.

56.342 ARE WE ATTENTIONALLY BIASED TOWARD REGULARITIES? NOT IF THEY FAIL TO BENEFIT BEHAVIOR

Caiden Ross-Moody¹ (caidenmoody@gmail.com), Andrew Leber¹; ¹The Ohio State University, Columbus, United States

Our world is filled with visual regularities, as we often navigate the same environments each day. Much research has investigated how we behaviorally exploit these regularities, but questions remain about how we learn to prioritize such regularities. Zhao, Al-Aidroos & Turk-Browne (2013) proposed we are spontaneously biased toward regularities, enabling us to learn whether such regularities can benefit behavior. However, as in our familiar everyday environments, many regularities prove to be uninformative (cf. learned irrelevance). Under these circumstances, we should prioritize novel information. Infants demonstrate preferences for either familiar or novel stimuli depending on their habituation to the former. Here, we investigate if adults flexibly prioritize repeated vs. novel stimuli, based on whether each type of information has the potential to afford a behavioral advantage. In Experiment 1, 50 participants completed a visual search task for a T target among L distractors, with the display split into left and right sides. 8-item configurations in each side repeated throughout, except for occasional (17%) 'test' trials for which either the left- or right-side configuration was novel. Critically, unlike standard contextual cueing experiments, no configuration predicted the target location, although participants would presumably attend these configurations to determine their predictiveness. Results showed a regularity bias: critical trial RTs were faster to targets in the repeated than novel display half. In Experiment 2, we attempted to more strongly habituate participants to the regularities by using only one repeated configuration on each display side. Here, we expected participants to learn that repeated configurations clearly conferred no behavioral advantage. Now, results revealed a novelty bias: participants were faster to identify targets in the novel half of the display. Altogether, these results confirm that human adults deploy attention flexibly to

either statistical regularities or novel information, depending on the potential behavioral utility of this information.

56.343 EVALUATION OF MODEL-GENERATED SALIENCY MAPS FOR AFFECTIVE SCENES

Simona Sartan¹ (simona.sartan@ufl.edu), Yujun Chen¹, Peng Liu¹, Ruogu Fang¹, Mingzhou Ding¹; ¹UNIVERSITY OF FLORIDA

Saliency maps are bottom-up, image-computable priority maps that highlight locations in natural scenes likely to attract visual attention. Deep learning-based saliency models, trained predominantly on emotionally-neutral images, have demonstrated strong performance in predicting human gaze patterns. Whether these models capture visual features that convey emotional significance in affective scenes is unclear. We addressed this by combining saliency-guided occlusion with large language model (LLM)-based emotion evaluation to test whether saliency maps from these models encode emotion-relevant information. Saliency maps for affective scenes from the International Affective Picture System (IAPS) were generated using three deep saliency models: DeepGaze IIE (Linardos et al., 2021), SalFBNet (Ding et al., 2022), and TranSalNet (Lou et al., 2022). A human eye-tracking saliency baseline from the EMOD dataset (Fan et al., 2018) was included for comparison. Three conditions were tested: occlusion guided by the original saliency map, occlusion of randomly relocated saliency patches preserving patch size and shape, and occlusion at random locations. Valence and arousal ratings for each occluded image were obtained from ChatGPT-3.5 using structured prompts. Correlation between LLM ratings and normative IAPS ratings were computed across occlusion levels. The results showed that saliency-guided occlusion produced steeper declines in both valence and arousal correlations compared to the two control conditions, confirming that model-identified salient regions carry emotionally-significant information. This effect was more pronounced for arousal than valence, suggesting that valence-relevant information is more spatially distributed across a scene. Performance under saliency-guided occlusion converged with the human gaze baseline, indicating that neutral scene-trained models identify regions that overlap with those prioritized by human visual attention. These findings demonstrate that (1) saliency models trained on neutral content capture emotionally-critical visual locations in affective scenes and (2) LLM-based ratings offer a scalable and efficient alternative to human ratings for affective image analysis.

56.344 CAN A BATHTUB BE IN THE BEDROOM SCENE?: 2 DIAGNOSTIC OBJECT COMBINING STRATEGY

Phutanik Setasartit¹ (phutanik@gmail.com), Payachana Victoria Chareunsouk^{1,2}, Waragon Phusuwan^{1,2}, Chaipat Chunharas^{1,3,4}; ¹Cognitive Clinical and Computational Neuroscience Center of Excellence, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand, ²Medical Sciences, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand, ³Division of Neurology, Department of Medicine, Faculty of Medicine, King Chulalongkorn Memorial Hospital, Thai Red Cross Society, Bangkok, Thailand, ⁴Chula Neuroscience

Center, King Chulalongkorn Memorial Hospital, Thai Red Cross Society, Bangkok, Thailand

Everyday scenes contain multiple objects, yet we effortlessly classify them. Prior research suggests that the properties of objects in a scene, known as diagnosticity, play a crucial role in scene categorization. A single diagnostic object can be sufficient to support scene classification. However, when multiple diagnostic objects provide converging or conflicting diagnosticities, the way they are jointly interpreted remains poorly understood. We hypothesize that room classification is primarily determined by how individual object diagnosticities are co-evaluated. Specifically, we focused on four integration models: (1) Winner-take-all, relying on the most diagnostic object; (2) Interactive, in which the most unlikely object constrains the overall diagnosticities; (3) Additive, summing object diagnosticities; and (4) Average, which proportionates to the total diagnosticity. We evaluated these models against responses from nine participants who completed a two-alternative forced-choice task across 4,860 trials. In each trial, they viewed a brief visual scene containing either one or two objects, selected from nine objects across three room categories: bedroom, bathroom, and kitchen. After 300 ms, the text probe of the room type was displayed, and participants judged whether the composition matched the probe. For two-object trials, the average model provided the best fit to participant responses ($r = .85$), outperforming the other models ($r \approx .70-.80$). To further examine the remaining model-data mismatches, we developed an extended model in which object diagnosticities were learned from the data. This substantially improved model performance ($r = .93$), with the largest increase observed in the bedroom condition ($r = .98$). Adding an interaction term resulted in a slight to no benefit ($r = .94$) with a lower Bayesian Information Criterion (BIC) than earlier models. These findings suggest that the weighted-sum model is most compatible with all hypotheses, with some rooms (e.g., bedroom) depending on highly diagnostic individual objects (e.g., bed).

56.345 INDEPENDENT ENSEMBLE FEATURE REPRESENTATION AT THE OBJECT LEVEL.

Mikayla Shirley¹, Jason Haberman; ¹Rhodes College

A prevailing question in the study of ensemble perception is whether it is supported by domain-general or domain-specific mechanisms. Evidence for domain-specificity comes from Haberman et al., 2015, which showed that error in low-level ensemble representation does not predict error in high-level ensemble representation. This raises another question: what happens to that relationship when high- and low-level ensemble features are embedded within a single object? We created stimuli in which two features, orientation and emotion, appeared within a single face. Dimensions could be manipulated independently, depending on the task. In separate blocks, observers ($N = 30$) adjusted the test stimulus to match either the average orientation or average expression of a set of faces. In expression blocks, the orientation of the faces varied, but the test stimulus observers adjusted was always upright. In orientation blocks, the

emotion of the test stimulus was orthogonal to the average emotion of the set. Orientation parameters prevented individual faces from tilting below the horizontal meridian, ensuring judgements of emotion would not be subjected to face inversion effects. In line with previous work, preliminary results revealed a non-significant relationship ($r = .12$, $p > .05$) between average orientation precision and average emotion precision, suggesting that ensemble representations do not become entangled when combined within a single stimulus. Ensembles remain hierarchically organized and independently represented.

56.346 STIMULUS-SPECIFIC MEMORABILITY EFFECTS GO BEYOND MEMORY

Darius Suplica¹, Igor Utochkin¹, Edward Awh¹; ¹University of Chicago

Stimulus-specific memorability refers to the finding that recognition memory performance is enhanced for specific items, independent of study and retrieval context. Nevertheless, recent work has raised the possibility that memorability effects are not specific to memory, *per se*. For instance, the efficiency of low-level perceptual discriminations – in the absence of memory demands – is higher for memorable stimuli (Deng et al., 2024). Following this thread, we investigated whether memorability affects visual search performance when recognition memory is not a limiting factor. Critically, memorability appears to have two distinct facets: “Hittability” and “false-alarmability” refer to the probability of correctly recognizing previously studied items, or falsely recognizing unstudied items, respectively. While both facets can strongly impact memory performance, they are virtually uncorrelated at the item level. Thus, we independently manipulated the hittability and fa-ability of targets and distractors in four separate samples. Visual search performance was strongly affected by both factors. Subjects were faster to find more hittable (rm-ANOVA, $p = 1 \times 10^{-11}$) and less fa-able targets ($p = 6.9 \times 10^{-10}$), with no effect of distractor memorability (hittability $p = 0.18$, fa-ability $p = 0.08$). Experiment 2 replicated these findings while employing a continuous manipulation of hittability and fa-ability within a single set of stimuli. Once again, subjects were faster to recognize more hittable and less fa-able targets (linear mixed-effects model, hittability $p = 7 \times 10^{-17}$, fa-ability $p = 1.2 \times 10^{-6}$), with a significant ($p = 2 \times 10^{-19}$) negative interaction between the two. In line with the argument that hittability and fa-ability are distinct aspects of memorability, search performance was better explained by keeping hittability and fa-ability as independent factors, rather than combining them into single d-prime measure (WAIC = -2954 vs -3048). Therefore, stimulus-specific memorability powerfully affects performance in tasks that do not require recognition memory, motivating further work on precisely which computations are impacted by memorability.

The authors were supported by NSF grant #2402391 to E.A.

56.347 STEREOSCOPIC PRESENTATION OF MASKING OCCLUDERS REVEALS THE TIME COURSE REQUIRED FOR PERCEPTUAL COMPLETION OF OBJECTS

Hugo Turro-Serrano¹ (hugo.c.turro-serrano@vanderbilt.edu), Lasyapriya Pidaparthi¹, Frank Tong¹; ¹Vanderbilt University

In everyday vision, many objects will appear partially occluded due to the presence of foreground objects. Amodal completion allows us to perceive partially visible objects by “filling in” the missing information caused by occlusion (Nakayama et al., 1990). Previous studies have investigated perceptual completion of partially occluded objects using 2D displays (Johnson & Olshausen, 2005; Tang et al., 2018), but due to the lack of stereo-depth cues, it remains unclear how these findings would generalize to realistic 3D conditions of visual occlusion. In our study, participants ($n = 17$) viewed stereoscopic grayscale images drawn from eight possible animal categories, which were fragmented by textured horizontal bars. These bar occluders appeared either in front of, behind, or at the same disparity relative to the object. To investigate how perceptual completion unfolds over time, the object images and occluders were briefly presented for either 27, 53, 107 or 213 ms. Subjects pressed a key to identify the object category. We found that participants were more accurate when the stimulus was presented for longer durations ($p < .001$) and when the occluder appeared in front of the object ($p < .001$). Importantly, the effect of occluder depth on recognition accuracy increased with stimulus duration ($p = .015$), requiring 107ms for foreground occlusion to have a significant beneficial effect. Surprisingly, recognition was no more accurate for occluders at zero disparity than at far disparity even though the latter condition was highly unnatural. Our results suggest that object completion based on 3D depth perception requires time to take effect, presumably because of the recurrent or top-down interactions that are required to support amodal completion. Together, our findings add to our knowledge of how the visual system leverages stereoscopic depth cues to complete partially occluded objects and facilitate object recognition in our daily lives.

This research was supported by National Eye Institute grant R01EY035157 to FT, P30EY008126 to the Vanderbilt Vision Research Center, and Vanderbilt’s MARC undergraduate program supported by grant T34GM136451 from the National Institutes of Health.

56.348 COMPREHENSIVE CHARACTERIZATION OF COLOR DISCRIMINATION THRESHOLDS IN DEUTERANOPES

Kara Wang¹, Cameron Zhang¹, Jason Chow², Philip Guan², Alex Williams³, David H. Brainard¹, Fangfang Hong¹; ¹Department of Psychology, University of Pennsylvania, ²Reality Lab Research, Meta, ³New York University, Center of Neural Science

Color discrimination thresholds—the smallest perceptible differences between colors—have been extensively explored in trichromats, but remain less well-characterized in dichromats. Measuring these thresholds provides foundational data to constrain mechanistic models of color vision and can inform the design of color-rendering algorithms that make visual information more accessible to dichromatic populations. We recruited two deuteranope participants and measured

their color discrimination thresholds in an LS cone-isolating plane. Participants performed a three-alternative forced-choice oddity task in which three matte, blobby stimuli were presented for 1 s: two shared the same surface color, one differed, and participants reported the odd one out. Reference and comparison stimuli were selected using a non-parametric adaptive procedure targeting 66.7% correct performance. Each participant completed 7,500 trials, along with additional fallback trials inserted when the adaptive procedure timed out. We fit a semi-parametric Wishart Process Psychophysical Model (WPPM) to the data for each participant. Once fit, the model enables prediction of discrimination performance for any pair of reference and comparison stimuli, effectively characterizing the entire psychometric field in the LS cone-isolating plane. We then compared threshold contours with those obtained from a larger trichromat dataset collected over the full 3D RGB cube, from which thresholds in the LS plane were read out by slicing the fitted psychometric field. The comparison revealed three key regularities across participants: (1) thresholds along the S-cone contrast were comparable between deuteranopes and trichromats, (2) thresholds along the L-cone contrast were substantially larger in deuteranopes, and (3) the major and minor axes of threshold ellipses in deuteranopes align roughly with S+L and S-L, respectively. Together, these results suggest that color discrimination in deuteranopes is governed by an additive mechanism and an opponent mechanism constructed from the available L and S cone signals, with greater sensitivity along the S-L opponent direction than along the additive direction.

56.349 SIMULTANEOUS REPRESENTATION OF LOCATION WITH RESPECT TO LANDMARKS AND BOUNDARIES

Kyra Warmuth¹, Sami Yousif; ¹The Ohio State University, ²The Ohio State University

Suppose you return to a place you have not visited in many years and try to navigate to a favorite restaurant. You do not remember its exact location, but you are confident you can find it. Yet once you begin, you realize that much has changed: restaurants have closed or opened, new parks have been established, and street names differ. The landmarks you once relied on are no longer reliable. How do you find your way? Some theories propose that cognitive maps integrate multiple sources of spatial information into a kind of 'cognitive collage', so that, when one source of information fails, there are other available cues. Empirical evidence for such ideas remains limited, however. Here, we provide some. Participants completed a spatial memory task in which they remembered the location of a dot relative to both a landmark and a boundary. In Experiment 1, both cues were present during encoding, but at retrieval participants saw either the landmark, the boundary, or both. Error patterns differed across conditions, with the lowest overall error when both cues were available, perhaps suggesting that locations were encoded relative to both. We also found different characteristic patterns of errors in each condition (e.g., participants were biased toward the landmark when it was present at

retrieval, but not otherwise). In Experiments 2 and 3, the task was identical except that retrieval trials were weighted toward the landmark condition (Experiment 2) or the boundary condition (Experiment 3) on 80% of trials (with 10% for each alternative). Memory improved for the high-probability cue and worsened for others, indicating that encoding flexibly prioritizes the most reliable environmental cues. These results suggest that spatial information may be encoded with respect to multiple reference frames simultaneously, supporting the idea that redundant representations can benefit memory when reliable cues are disrupted.

56.350 WHERE DOES THE EYE GO? SHAPE-BASED AND GLOBAL SPATIAL BIASES IN GENERATIVE FACE PAREIDOLIA.

Emily Westrick¹, Benjamin Balas¹; ¹North Dakota State University

Face pareidolia research has contributed to our understanding of how face-like features are organized and identified throughout childhood and into adulthood. Most pareidolia research has either examined participants' responses to images with features similar to the 1st-order configuration of faces, or asked participants to identify such features in random or textured patterns. McCloud (1992) identified another aspect of face pareidolia that specifically emphasizes top-down processes and perceptual organization: Placing a single eye within arbitrary silhouettes to impose face-like appearance on random shapes. His claim was that eye placement is sufficient to impose face-ness on such closed forms, with participants re-organizing and re-interpreting silhouettes as faces to accommodate the eye micro-pattern. Presently, we used this unique type of face pareidolia to ask whether children and adults exhibit systematic preferences regarding where an eye should be placed in an arbitrary shape and whether these preferences are differentially affected by broad spatial biases and shape-based biases as a function of age. We presented children (#N=137) and adults (#N=47) with four ambiguous silhouettes and prompted them to place one eye sticker on each shape wherever they felt was best. Across participant groups, these silhouettes were presented either in their original orientation, or flipped upside-down. Based on prior research describing young children's tendency to disregard orientation relative to paper borders when drawing human figures (Goodnow and Friedman 1972), we hypothesized that younger children would place eyes in a more shape-dependent manner (being less affected by silhouette orientation), and older children would share a tendency with adults for more space-dependent placement. Our analyses revealed just such an interaction between age group and silhouette orientation with regard to vertical eye position. We discuss these results in the context of a simple image-based model to predict eye placement in arbitrary closed forms.

Funding from NSF BCS-2338600

56.351 EXPLORING INTEROCULAR INTERACTIONS IN HEALTHY CONTROLS DURING BINOCULAR RIVALRY WITH ASYMMETRIC STIMULUS CONTRAST USING SSVEP

Riley Wiland¹, Enoch Ansah Asiedu¹, Kimberly Meier¹; ¹University of Houston

Frequency-tagged SSVEP can isolate brain responses to signals sent to each eye, by flickering the two eye images at different frequencies. Previous research using this approach in a binocular rivalry paradigm found that greater imbalance in the amplitude of the eye-specific responses was associated with weaker intermodulation (IM) responses in amblyopia (Hu et al., *Heliyon* 2024;10:e39358). One proposed explanation was that weaker IM responses could partly reflect less time in a 'mixed' percept state if one eye dominates perception, rather than the amblyopic deficit itself. To test this hypothesis, we systematically reduced contrast in one eye of healthy controls to determine if unequal input reduces time spent seeing mixed percepts, and whether reduced mixed percepts are associated with weaker IM responses. Participants (n=7) viewed a red patterned stimulus in one eye and green in the other during 30-sec trials. Contrast was fixed in the left eye (80%) and varied in the right eye (0-100%) across trials. Observers continuously reported whether they saw a green, red, or mixed image. We extracted the amplitudes of the second harmonics for left-eye (6 Hz) and right-eye (7.5 Hz) frequencies, and the IM component (13.5 Hz), from the SSVEP responses. We found that mixed percept reports did not decrease monotonically as contrast was reduced in one eye: during rivalry, the main effect of right-eye contrast on mixed percepts was not significant ($p = 0.17$), and the mean within-subject slope (0.02) was not different from zero ($p = 0.74$). Time spent seeing mixed percepts did not significantly predict IM amplitude within observers ($p = 0.26$). Thus, reducing contrast in one eye did not produce the predicted decrease in mixed percepts or relationship with IM amplitude. Together, these results suggest unequal input alone does not fully account for the weaker intermodulation responses previously reported in amblyopia.

NIH Center Core Grant P30 EY007551

56.352 INFLUENCE OF SPATIAL RELEVANCE AND STATISTICAL LEARNING ON ATTENTIONAL CAPTURE

Jordyn R. Wright¹ (jordynwright@ucsb.edu), Sarah E. Purnell¹, Daniel D. Thayer¹, Thomas C. Sprague¹; ¹University of California, Santa Barbara

Statistical regularities in the environment can shape attentional priority in space (Theeuwes, 2022). For example, attentional selection of a target is enhanced when it appears at high probability locations (HPL), while interference from distractors is reduced when they are presented at HPLs (Ferrante et al., 2018). Although these statistical regularities can be learned even when task-irrelevant or without explicit awareness (Duncan & Theeuwes, 2020), some degree of top-down or bottom-up attention to stimuli containing these regularities may be necessary for learning to occur (Duncan et. al, 2025). Here, we tested whether learning statistical regularities depends on the top-down allocation of spatial attention to relevant locations (Golan et al., 2024). Participants covertly searched an 8-item array for a uniquely shaped target and

indicated the orientation of a line within the target (horizontal or vertical). On most trials, a central cue indicated the spatially relevant hemifield of the search array (left or right; 4 of 8 items) in which the target would always appear. Half of all trials contained a task-irrelevant color singleton distractor, which could appear within or outside the cued hemifield. Critically, the distractor appeared more often at two HPLs within the array (one in each hemifield; left and right). Statistical regularities of distractor location interacted with spatial cueing to reduce interference from salient distractors, however, results varied depending on the spatial relevance of the HPL. These results suggest that learning distractor-location regularities may depend on whether those regularities are spatially relevant, and support future research on the flexibility of learning environmental regularities to guide visual search.

Funding: National Eye Institute R01-EY035300

56.353 IDENTIFYING BRAIN REGIONS ENGAGED IN VISUAL EMOTIONAL PROCESS: COMPARISON OF UNIVARIATE VS MULTIVARIATE METHODS

Benjamin Cheng Yin¹, Ana Clara Villegas Botero¹, Max Lobel¹, Mingzhou Ding¹; ¹University of Florida

Presently, to identify brain regions engaged in visual emotional processing using fMRI data, one of three approaches is typically applied: univariate analysis based on general linear model (GLM), within-subject multivariate decoding analysis (wMVPA) where the classifiers are trained and tested within each subject, or across-subject multivariate decoding analysis (aMVPA) where the classifiers are trained on N-1 subjects and tested on the remaining subject. To understand the similarities and differences among these methods, we analyzed fMRI-BOLD responses from N=20 participants viewing pleasant, neutral, and unpleasant natural scene images drawn from the International Affective Picture System (IAPS). For univariate analysis, two contrast maps were generated: pleasant-neutral and unpleasant-neutral. For the two MVPA methods, a whole-brain searchlight analysis was carried out to generate two decoding accuracy maps: pleasant-neutral and unpleasant-neutral. At voxel level, the strongest correlations were between the two MVPA methods (pleasant-neutral $r = 0.63$; unpleasant-neutral $r = 0.66$), followed by aMVPA and univariate analysis ($r = 0.39$; $r = 0.47$), with the weakest correlations found between wMVPA and univariate analysis ($r = 0.34$; $r = 0.37$). At the brain region level, across all three methods, the most consistent effects were observed in posterior visual and lateral parietal cortices. wMVPA and aMVPA further converged on fusiform cortex, inferior parietal lobule, supramarginal gyrus, superior parietal lobule, precuneus, and cerebellar lobule VI. Conversely, univariate GLM analysis revealed significant effects in bilateral amygdala, hippocampus, posterior cingulate cortex, posterior orbitofrontal cortex, mid temporal cortex, bilateral Heschl's gyrus, superior temporal cortex, mid temporal cortex, thalamus, and bilateral precuneus. These results suggest that (1) convergence between univariate and multivariate methods is low, (2) the two MVPA methods, while showing more

convergence, can still differ significantly, and (3) applying all three methods together will yield the most comprehensive understanding of the brain regions engaged in visual emotion processing.

NIH grants MH125615 and MH112558 and NSF grant BCS2318984

56.354 REFLEXIVE INHIBITION OF EYE MOVEMENTS RELIES ON INTACT CORTICAL VISUAL PROCESSING

Klaury Youchom-Tagheu¹, Philipp Kreyenmeier¹, Krystel R. Huxlin², Elisha P. Merriam¹; ¹National Institute of Mental Health, ²University of Rochester

Motor circuits in the brain have privileged access to visual information, allowing rapid responses to external stimuli. Whether such rapid visuomotor behaviors are mediated via subcortical pathways or rely on cortical visual processing is debated. Here we study oculomotor freezing, the reflexive cessation of saccades in response to the onset of visual or auditory stimuli, to test whether this short-latency oculomotor behavior relies on intact cortical visual processing in humans. We recorded eye movements of two patients (p01:65-year-old male; p02:42-year-old female) with unilateral stroke-induced damage to early visual cortex. The patients continuously fixated at the center of a screen while we presented visual gratings (1cycle/degree, 3degrees diameter, 100% contrast), auditory white noise bursts, or both combined. The grating was either presented inside the patients' blind field (p01:[-18,-5]; p02:[7,5]) or at a corresponding location within the intact visual hemifield. The auditory stimulus was spatially unbiased. In both patients, stimulus-locked microsaccade rates exhibited a rapid decrease in response to the onset of the peripheral grating presented in the intact hemifield (indicating oculomotor freezing). The inhibition latency was ~100ms and lasted for ~300ms. In contrast, when iso-eccentric gratings were presented inside the blind field, the same saccade rate modulation was not observed. In trials where gratings were paired with the auditory stimulus, oculomotor freezing was enhanced for stimulus presentation in the intact hemifield, compared to auditory-only trials. No enhancement was observed when gratings were presented inside the blind field, indicating that multisensory enhancement only occurred when the visual stimulus was presented inside the intact field. Our results provide neuropsychological evidence that intact visual cortical processing of stimulus onsets is necessary to elicit oculomotor freezing. These findings are consistent with previously reported oculomotor abnormalities in cortically blind patients, and psychophysical studies suggesting that perception and oculomotor freezing rely on shared visual signals.

56.355 THE ASSOCIATION OF CLASSROOM SEATING LOCATION ON VISUAL GAZE BEHAVIOR AND MIND-WANDERING

Nanyu Zhang¹, Rebekah Che¹, Xiaorui Xue¹, Ido Davidesco¹; ¹Boston College

Prior research suggests that seating location influences academic engagement, but most evidence focuses on front-to-back distance and relies on subjective self-reports. This study examines whether central versus peripheral classroom positioning is associated with differences in mind-wandering (MW) frequency and gaze behavior. We hypothesized that students in peripheral seats would show reduced visual attention to instructional areas of interest (AOIs) and higher rates of self-reported MW, potentially because these seats expose students to more visual distractions, such as peers and devices. Wearable eye-tracking glasses (Pupil Labs Neon) were used to collect data in a college classroom from 15 undergraduates across six class sessions. Each participant wore eye-tracking glasses during two sessions. Students were prompted 6-8 times throughout class to report via an online survey whether they were mind-wandering immediately before the probe. Eye-tracking data from the 1-minute window preceding each probe were coded for specific AOIs, categorized as Instructional (slides, instructor, note-taking) or Non-instructional (other students, others' devices, etc.). Descriptively, students in center seats (n=7) spent a greater proportion of time looking at instructional AOIs (M=61.8%, SD=0.095) than students in peripheral seats (n=8, M=48.4%, SD=0.276). Self-reported MW showed a similar directional pattern, with students in the periphery reporting somewhat higher average MW (M=25.2%, SD=0.270) than center-seated students (M=20.4%, SD=0.184). The peripheral group also showed greater variability in both instructional gaze and MW frequency. These preliminary findings suggest an association between seating location and gaze-based measures of classroom attention, while also indicating substantial heterogeneity among students in peripheral seats. Ongoing data collection will expand the sample and allow us to test whether these patterns replicate and to examine possible mechanisms underlying seating-related differences in classroom attention.

This material is based upon work supported by the National Science Foundation under grant number 2526250.

56.356 EXPECTATION CONFIRMATION RESULTS IN MORE CONSERVATIVE PERCEPTUAL DECISIONS: A STATISTICAL LEARNING STUDY

Zihui Zhao¹, Joey Zhou¹; ¹Shenzhen University

Statistical learning (SL) enables individuals to extract regularities and form expectations. While extensive research has illustrated that SL is prevalent in both humans and animals, it remains unclear how the learned regularities shape perception. Some studies suggest that it shifts the criterion, whereas others indicate that it shapes sensitivity. Here, we tested whether regularities learned via SL modulate sensitivity and/or criterion in simple perceptual decisions. Participants (N = 34) completed a training and a testing session on the same day. In the training session, participants viewed sequences of four gratings with varying spatial frequencies (SF), arranged in either a predictable or random order. In the testing session, three types of grating sequences were displayed: expected, random, and unexpected (i.e.,

the final grating violated the learned predictable sequence). Participants were required to report if the last grating of the sequence was the same as or different ($\pm 3^\circ$, $\pm 6^\circ$ or $\pm 9^\circ$) from the previous ones. By modeling the participants' proportion of "same" responses as a function of the last grating's orientation offset, we parametrically tested whether the width and/or baseline of the psychometric curves differed across conditions, thereby assessing whether prediction sharpens discriminability and/or shifts the decision criterion. The results showed that accuracy in the expected condition was significantly higher than in the unexpected condition. Furthermore, inspection of the psychometric functions revealed that the expected condition had a significantly lower baseline compared to the other two conditions. In contrast, the unexpected condition was associated with a higher baseline and a narrower width relative to the random condition. Together, these findings demonstrated that participants are able to acquire regularities and form expectations via SL. When stimuli are expected, participants tend to adopt a more conservative criterion. In contrast, violation of expectation leads to a more liberal criterion, accompanied by reduced sensitivity.

56.357 A MNEMONIC TASK ATTENUATES OTHER-RACE DEFICITS IN RECOGNITION MEMORY AND IMPLICIT BIAS

Yoseph Ahmad¹ (yosephahmad2@yahoo.com), Caeley Tackitt¹, Madison Antkowiak¹, Kaylee Do¹, Sarah Newbolds¹, James Townsend², Michael Wenger¹; ¹The University of Oklahoma, ²Indiana University

This study examines how mnemonic training can improve recognition memory and reduce implicit bias for faces of other races, replicating Lebrecht et al. (2009). The study incorporates EEG measures and an extended training design, and the present report focuses on the behavioral data pertinent to recognition memory and implicit bias. Participants completed a 10-session protocol comprised of pre- and post-assessments and eight training sessions. The baseline and endline sessions included three rounds of a recognition memory task in which participants had to discriminate between studied and unstudied faces. This was followed by the affective lexical priming score (ALPS) task, in which participants observed a prime of a same or other race face, followed by a positive, negative, or neutral word, or a non-word, after which they made their response as to whether the word was a real or a non-word. During training, White participants were assigned to individuate either Black or White faces. Changes in recognition performance and ALPS scores from pre- to post-training were used to assess modulation of the ORE and implicit bias. Results from the recognition memory task showed statistically significant increases in hit rates, decreases in false alarm rates, and increases in recognition discriminability, with no significant change in response bias, replicating the major findings of Lebrecht et al. (2009). Changes in ALP scores showed significant reductions in implicit bias, highlighting the potential for mnemonic training to reduce both the ORE and implicit bias. Ongoing work is intended to determine the extent to which the same effects can be obtained with a perceptual

rather than a mnemonic task and the extent to which neural measures are concurrently affected.

The University of Oklahoma Vice President for Research and Partnerships

56.358 THE ROLE OF COLOR AND SHAPE IN BOTTOM-UP PROCESSING WITH REALISTIC STIMULI

Sandra Park¹ (sandrapark1250@gmail.com), Alejandro Lleras², Junming Yu³, Simona Buetti⁴; ¹University of Illinois Urbana-Champaign

While substantial work exists on characterizing top-down guidance, much less systematic work has been done on bottom-up processing. It is not currently well understood how signals from different visual dimensions combine to capture attention and the eyes, nor how these signals operate with realistic stimuli. Here, we used naturalistic stimuli (apples) presented in a 9 x 9 grid with the center removed on a uniform gray background. Participants ($n = 23$) searched for a target that differed from distractors by color (green or yellow among red apples), shape (one or two bites taken from an apple among whole apples), or both features combined, while eye movements were recorded using an EyeLink 1000. Overall, color contrasts produce faster saccadic initiation times (188.84 ms vs. 233.79 ms) and more accurate saccades (guessing rate of 0.097% vs. 0.5583%) than shape contrasts. In the feature-combined conditions, we observed a cross-over interaction between shape and color contrasts: shape contrasts that were highly dissimilar from the background apples reduced the contribution of color contrasts in the color-similar (but not in the color-dissimilar) condition. Likely reflecting that the shape manipulation (taking two bites out of the apple) disproportionately reduced the ability of color in the remaining part of the apple to guide attention. This suggests that the two features do not always combine additively in peripheral vision to guide attention. Overall, these results indicate that color and shape contribute to bottom-up attentional guidance in somewhat different ways. Color produces faster and more reliable guidance while shape produces smaller guidance effects. Importantly, across all conditions, saccadic accuracy strongly predicted saccadic initiation time ($R^2 = 0.918$), with faster initiation times being associated with smaller guessing rates. This final result suggests both feature dimensions contribute in analogous though unequal fashion to the overall saccade execution system.

56.359 CONTRAST MANIPULATIONS INFLUENCE ON CONGRUENCY EFFECT IN EXPERT MUSIC READING

Darshan Shah¹, Zoe Epstein¹, Ziqing Zhang¹, Brianna Charlton¹, Allan Collins¹, Cindy Bukach¹; ¹University of Richmond

Perceptual expertise is associated with holistic processing across many modalities, including faces, cars, and birds, but music reading remains understudied. As a complex perceptual system, music reading can be examined with systems factorial technology, which models holistic processing through contrast manipulations that affect

perceptual processing speed. We asked how this contrast manipulation influences holistic processing, indexed by congruency effects ($\Delta d'$ and Δc). We hypothesized that if experts coactively pool music notes, congruency effects should be similar in high- and low-contrast conditions, as global form should still be salient in low contrast conditions. We had no predictions for conditions in which contrast masked half of the stimuli. Four experts completed three-six sessions in which they responded to sequentially presented composite four-note music sequences with low- or high-contrast on the left and/or right two notes. Hierarchical generalized linear mixed-effects models (binomial probit, with observer-level random effects) estimated congruency effects across contrast levels. Bayesian information criterion favored a model in which contrast moderated congruency effects. On average, discriminability (d') was significantly higher and response bias (c) was significantly lower in high- than low-contrast conditions. $\Delta d'$ and Δc main effects were also observed. As predicted, contrast moderated congruency where congruency effects did not differ for completely high or completely low contrast condition. Interestingly, congruency effects decreased when only part of the stimulus was masked. Overall, contrast manipulations influenced congruency effects suggesting that systems factorial technology may further inform the nature of holistic processing.

University of Richmond

TUESDAY AFTERNOON POSTERS IN PAVILION

Visual Working Memory: Spatial

TUESDAY, MAY 19, 2:45 – 6:45 PM, PAVILION

56.401 DOES NEURAL ACTIVITY IN HUMAN V1 PERSIST DURING WORKING MEMORY, AND IF NOT, WHY?

Wing K. H. Chu^{1,2} (wkc267@nyu.edu), Hsin-Hung Li³, Thomas C. Sprague⁴, Kartik K. Sreenivasan², Clayton E. Curtis¹; ¹New York University, ²New York University Abu Dhabi, ³The Ohio State University, ⁴UC Santa Barbara

Despite ample evidence that working memory content can be decoded from human primary visual cortex (V1), single-neuron activity in macaque V1 rarely persists above pretrial baseline. Persistent activity is the most established neural mechanism supporting working memory; thus, its absence in V1 has led many to conclude that V1 is not involved in working memory storage. Here, we address this contradiction by carefully estimating delay period activity in human V1 from three fMRI studies that successfully decoded spatial working memory content from V1. Leveraging population receptive field (pRF) mapping, we estimated BOLD responses in V1 voxels whose pRFs contained the memory target location (in-RF) and compared these to voxels that did not (out-RF). Across all three datasets, the amplitude

of BOLD activity in the in-RF voxels was greater than in the out-RF voxels throughout the memory retention interval, demonstrating that working memory content modulated V1 activity. The BOLD time courses were revealing: following a transient increase in BOLD activity time-locked to the visual target onset, the in-RF response returned to baseline while the out-RF response dipped below baseline. Indeed, only V1 voxels whose pRF overlapped the target location returned to baseline, whereas voxels tuned elsewhere exhibited widespread suppression. Therefore, human V1 activity during working memory delays did not persist above pretrial baseline in the classical sense; instead, spatially-defined differences across subpopulations were sustained. Furthermore, the in- vs. out-RF difference in neural responses predicted decoding performance, suggesting that decoding succeeds precisely because in-RF responses remain distinguishable from suppressed responses elsewhere. Finally, we used a hierarchical neural circuit with recurrence, normalization and interareal feedback to test which computations are necessary to reproduce V1's empirical dynamics. Together these results reveal how selective V1 activity supports robust decoding of working memory features without persistent firing, explaining V1's seemingly paradoxical role in working memory.

NIH R01 EY-016407 and R01 EY-033925 (CEC), NYUAD Research Institute Grant CG012 (KKS), Global PhD Fellowship (WKHC)

56.402 EFFECT OF LANDMARK USEFULNESS ON SPATIAL WORKING MEMORY REPRESENTATIONS

Nicholas Schmitz¹ (nschmitz@vols.utk.edu), Sage Bendickson¹, Alexis McCraw¹, Ilke Kavusturan¹, David Sutterer¹; ¹University of Tennessee - Knoxville

Patterns of EEG alpha-band power track sustained maintenance of remembered locations (Foster et. al. 2016; Sutterer et. al. 2021). Studies using this method have generally employed a blank background while a growing body of research shows that background landmarks contribute to how we remember locations (Aagten-Murphy, Bays 2019). An open question is whether background landmarks enhance or reduce active maintenance of object locations. To answer this question, we performed two experiments where we recorded electroencephalography (EEG) activity while participants completed a delayed estimation task in which the background alternated between a grid of lines intersecting stimulus locations and concentric circles parallel to stimulus locations. In Experiment 1, observers were asked to remember the location of a target dot. We found that participants were significantly more accurate when the background consisted of intersecting lines compared to parallel circles. Similarly, applying an inverted encoding model (IEM) to patterns of alpha-band power revealed enhanced spatial selectivity for remembered locations in the presence of intersecting lines compared to parallel circles. In Experiment 2, observers were asked to remember the color of a target dot and item location was never tested. We found no difference in color memory performance between background conditions. However, spatial selectivity for target locations was enhanced when the

background consisted of intersecting lines versus parallel circles. In contrast to Experiment 1, above chance spatial selectivity did not persist through the end of the retention interval. Taken together our results reveal that observers continue to rely on active memory maintenance in the presence of useful background landmarks. Instead of using landmarks to free up spatial memory, the presence of landmarks strengthens online memory representations. However, this enhancement does not sustain when space is irrelevant, suggesting that the duration of active spatial memory maintenance is task dependent.

56.403 N2PC SET-SIZE EFFECTS EMERGE WHEN WORKING MEMORY REQUIRES FLEXIBLE SPATIAL REPRESENTATIONS

Gengshi Hu¹ (gengshihu@uchicago.edu), Chong Zhao¹, Edward K. Vogel¹; ¹University of Chicago

In visual working memory (VWM) tasks the contralateral delay activity (CDA) scales with the number of items stored, while the N2pc component—reflecting attentional selection—does not vary with set size. This dissociation has been replicated many times and suggests that VWM load effects emerge mainly during maintenance, not initial selection. A notable exception is found in multiple object tracking (MOT) tasks, where the N2pc reliably increases with the number of tracked targets. Because the N2pc in MOT appears within ~250 ms of stimulus onset—well before objects begin to move—tracking cannot explain this early modulation. Instead, a key difference between MOT and standard VWM tasks is that in MOT, the target location at test differs from the object's original position, whereas in VWM tasks, the probed item generally reappears at its initial location. To test whether expectations about future spatial transformations drive N2pc load sensitivity, subjects performed a VWM task with two block-wise conditions. In the baseline condition, the probed item appeared at the same location in which it was encoded. In the spatial-shift condition, the probe was always displaced inward within the same hemifield. Crucially, the encoding displays were physically identical between conditions, and the N2pc was measured prior to any spatial shift. Behaviorally, memory performance was equivalent across conditions. Replicating prior work, the N2pc was insensitive to set size in the baseline condition. Strikingly, in the spatial-shift condition, the N2pc showed a reliable set-size effect, with larger amplitudes for four versus two items. These findings demonstrate that the anticipatory structure of the task—specifically, whether object positions are expected to change—alters how items are individuated during selection. Expectations about future spatial uncertainty appear to recruit enhanced target individuation at encoding, revealing that attentional selection mechanisms are more flexible and forward-looking than previously assumed.

This research was supported by the funding from National Institute of Mental Health (grant ROIMH087214); Office of Naval Research (grant N00014-12-1-0972).

56.404 ALLOCENTRIC VISUOSPATIAL WORKING MEMORY IS MORE ROBUST THAN EGOCENTRIC TO AGE-RELATED DECLINE

Stella Wernicke¹ (saiw2@cam.ac.uk), Paul M Bays¹; ¹University of Cambridge

In visuospatial working memory (VSWM), two kinds of location information can contribute to recall: egocentric, defined relative to the observer, and allocentric, defined by the spatial relationships between objects. Previous work (Aagten-Murphy & Bays, 2019) found that resources for storing information from these two sources are limited, but independent. In adults, VSWM declines with age, but whether the two memory systems deteriorate equally is not known. To disentangle the contributions of ego- and allocentric processing to age-related VSWM decline, 40 young (18–30 years) and 40 older (60–70 years) neurotypical healthy adults completed an online VSWM task. In each trial, three colored disks were presented at random locations. After a brief delay, one stimulus was cued by its color, and participants reported its remembered location. Additionally, a larger gray disk, serving as a visual landmark, was either present throughout the entire trial (LM-PRESENT; 96 trials) or only during stimulus presentation (LM-ENCODING; 48 trials). We fit the data with a model in which only egocentric location information is available in LM-ENCODING, but this is optimally combined with allocentric information about location relative to the landmark for LM-PRESENT. Allocentric precision is assumed to peak at the location of the landmark and decline with increasing distance from it, while egocentric precision is independent of stimulus-landmark distance. This model fit the data better than models assuming allocentric information was integrated in both conditions or neither. While egocentric memory fidelity was lower in older than in younger adults (BF10 = 325), there was evidence against age-related differences in peak allocentric precision (BF10 = 0.359) and its decline with distance (BF10 = 0.530). These results suggest that allocentric VSWM is relatively preserved with age, whereas egocentric memory degrades more strongly. Future research could investigate the neural basis of this dissociation and contrast healthy aging to neurodegenerative disorders.

SW received a PhD scholarship from the Foundation of German Businesses, financed by the German Federal Ministry of Research, Technology, and Space. Additional support was provided through a research prize for research expenses from the Wilhelm Woort Foundation for Aging Research awarded to SW. The authors declare no conflicts of interest.

56.405 SPATIOTOPIC WORKING MEMORY MEASURED WITH OCULOMOTOR RESPONSES

Sylvie R. Bass¹, Amelia H. Harrison¹, Thomas C. Sprague¹; ¹University of California, Santa Barbara, Santa Barbara, CA

Every saccade shifts visual input, requiring the brain to update spatial representations to keep perception stable. Previous studies have shown that spatiotopic memory – memory for object locations relative

to the world – degrades with each saccade away from the encoding gaze position (Golomb & Kanwisher, 2012; Shafer-Skelton & Golomb, 2018). However, using a mouse click to report a remembered object location may incur performance costs due to the need to convert spatiotopic representations across oculomotor movements into motor plans. To investigate whether the previously-seen degradation in recall precision reflects fundamental limitations in spatiotopic updating rather than modality conversion costs, we conducted an experiment in which participants reported a remembered spatiotopic location with an eye movement, rather than mouse click or touchscreen reach, after zero through three saccades, with the final gaze position matching or differing from the initial gaze position at encoding (where retinotopic and spatiotopic locations match). Our results replicate previous studies showing that spatiotopic memory error increases systematically with more saccades. However, there are observed benefits to spatial memory accuracy by returning to the initial encoding location with the final saccade endpoint. This suggests that the degradation arises from inherent limitations in spatial updating mechanisms, not from response modality constraints. Thus, the cumulative error may be attributed to the cost of converting spatial locations from our native retinotopic coordinate system into spatiotopic coordinates. When the two coordinate systems realign by returning gaze position to the encoding location, the preserved retinotopic memory trace may bypass remapping costs. Future work will seek to explore how both domain-general navigation processes and spatiotopic and retinotopic coordinate systems act in tandem to impact visual working memory performance.

UCSB Undergraduate Research and Creative Activities Grant, UCSB Academic Senate Research Grant

56.406 VISUAL CONTEXT MODULATES SYSTEMATIC BIASES IN HUMAN SPATIAL WORKING MEMORY

Danny Lim¹, Thomas Sprague; ¹UC Santa Barbara

Working memory (WM) maintains information to guide behavior, recruiting circuits across visual, parietal, and prefrontal cortex (Christophel et al., 2017; Curtis & Sprague, 2021; Kiyonaga & Serences, 2025). Yet spatial WM shows systematic recall biases (Jiang & Leung, 2021). Parallel research in humans and nonhuman primates shows that medial temporal lobe structures, including entorhinal and hippocampal cortex, use allocentric information to modulate gaze position (Killian et al., 2012; 2015; Meister & Buffalo, 2018; Nau et al., 2018; Julian et al., 2018; Wilming et al., 2018), suggesting that environmental structure influences spatial coding beyond retinotopic maps. We tested whether such geometry-anchored representations also shape human spatial WM. Participants performed a spatial WM task in which a stable screen aperture (circle, square, wide rectangle, tall rectangle) remained visible throughout each trial. They briefly viewed a target presented within the aperture (randomly chosen on each trial), and then reported its location either with a memory-guided saccade or a mouse click to the remembered position. If allocentric scaffolding contributes to WM, geometric context should

bias recall. Our results reveal interactions between aperture geometry and target position across response modalities. Circular and square apertures produced relatively small overall biases, whereas wide and tall rectangular apertures yielded pronounced anisotropies aligned with their long axis. Within the square aperture, however, responses to diagonal targets were systematically displaced toward the corners, indicating that even nominally symmetric geometries impose structured distortions. These effects demonstrate that spatial WM is not encoded solely in retinotopic coordinates but is shaped by allocentric structure in the surrounding visual environment. Together, these findings show that geometric context exerts a strong, systematic influence on spatial WM representations. They provide behavioral evidence that the visual system integrates allocentric cues to stabilize and organize memory, offering a principled framework for probing the neural mechanisms underlying reference-frame transformations.

Funding: UCSB Academic Senate Research Grant

Visual Working Memory: Objects, features

TUESDAY, MAY 19, 2:45 – 6:45 PM, PAVILION

56.407 BUY ONE, GET THE REST FREE: ACCESSING A SINGLE ELEMENT ACTIVATES THE WHOLE CHUNK EVEN WHEN IT HURTS

Nurullah Ertas¹, Eren Gunseli¹; ¹Sabanci University

Chunking helps us mitigate working memory (WM) capacity limits, but it may reduce flexibility in accessing individual elements. We investigated whether accessing a single element automatically activates the entire chunk even when doing so is unnecessary (Experiments 1 & 2) or harmful (Experiment 3). Participants first learned four chunks of three colors each. In a subsequent WM task, six colors were presented, followed by a retro-cue indicating the location of the to-be-tested item. In half the trials, a second cue followed, requiring reporting only the second-cued item. If chunks are accessed as a whole, response times (RT) should be faster when the second cue targets the same chunk as the first, compared to an item outside that chunk. In Experiment 1, each trial presented one studied chunk and three novel non-chunk items. RTs were faster when both cues targeted the chunk than when the first cue was a non-chunk. Experiment 2 ruled out a memory-type switch explanation by presenting two studied chunks, rather than one chunk and three non-chunk items: RTs were faster when the second cue remained within the same chunk vs switched to the other chunk. Drift diffusion modeling across experiments revealed that this RT benefit was driven by more efficient information processing (drift rate), rather than decision strategy or encoding delays. Experiment 3 manipulated the blockwise probability of within- vs across-chunk transitions between cues as 72% or 28%, making whole-chunk access advantageous or

disadvantageous, respectively. While within-chunk RT benefit was weakly attenuated when holistic access was disadvantageous, it remained robust in both block types. These results suggest that accessing an individual element triggers an automatic activation of the entire chunk. Although slightly reduced by rare probabilities, the robust persistence of the holistic benefit across all contexts highlights limited flexibility and strategic control over item-level access within chunks.

56.408 NO REAL-WORLD OBJECT BENEFIT FOR VISUAL WORKING MEMORY IN A WHOLE-REPORT TASK

Olga Kozlova¹, Kirsten Adam¹; ¹Rice University

Visual working memory (VWM) is a limited-capacity storage that allows individuals to actively maintain information in mind. However, previous research conflicts on whether the capacity of VWM differs between simplistic stimuli (e.g., colored squares) and real-world objects. Here, we sought to 1) replicate and extend the VWM object benefit in a whole-report VWM task, and 2) test whether proactive interference may modulate the difference between simplistic stimuli and real-world objects. In two experiments ($n=26$ each), we compared VWM performance for colored squares versus real-world objects. On each trial, participants remembered 6 items across a blank delay, and reported the correct binding of each item to its location at test (i.e., participants made 6 memory responses in a “whole-report” paradigm). In Experiment 1, the stimulus pools were matched between colored squares (9 possible colors) and objects (9 possible objects). In Experiment 2, the stimulus pool was smaller for colors (9 possible colors) than for objects (hundreds of trial-unique objects). In both experiments, we found no evidence for an object benefit in VWM. Instead, we found that participants remembered significantly more color-location bindings compared to object-location bindings ($p<.001$). Was this “color benefit” driven by more lapses (0-1 correct reports) for real-world objects? A between-subjects repeated measures ANOVA (Experiment $\times$ Stimuli) revealed a main effect of the stimuli type on the proportion of lapses ($p=.018$), but no between-subjects effect of the experiment ($p=.865$), nor interaction effect ($p=.110$). Thus, participants experienced more lapses for objects than simplistic stimuli overall, and this was not related to using trial-unique or repeating objects. Altogether, our study failed to replicate a real-world object benefit for VWM, and instead found more failures in VWM performance for real-world objects than simplistic stimuli.

56.409 FLEXIBLE GROUPING PROCESSES IN VISUAL WORKING MEMORY VIA ADAPTABLE POINTER ALLOCATION

Shachar Lando^{1,2} (shacharlando@mail.tau.ac.il), Roy Luria¹, Halely Balaban²; ¹Tel Aviv University, ²Open University Israel

Perceptual grouping enables effective information compression in visual working memory (VWM), but the precise computations that underlie grouping remain unclear. Recently, it has been suggested that VWM relies on a content-independent pointer system to establish unique representational instances, and to keep them in line with

environmental changes. In three experiments, we examined whether, and under what conditions, a group-specific pointer might be assigned to the product of the grouping process. We recorded EEG and monitored the contralateral delay activity (CDA), a neural marker that rises with the number of representations in VWM. Notably, the CDA also reflects pointer dynamics, exhibiting a sharp drop in amplitude when a pointer is invalidated. Participants performed change-detection tasks where three Pacmen moved either independently or as a Kanizsa triangle (movement was task-irrelevant). In all studies, the Kanizsa induced information-compression, reflected by lower CDA amplitudes and higher accuracy rates than for three independently-moving items. In 50% of trials, all Pacmen abruptly changed trajectory, which in the grouped condition broke the Kanizsa formation. We contrasted different tasks (holding the animation sequence constant), and examined the resulting pointer dynamics. When the test involved judging the orientation (grouping-relevant feature) of only a single Pacman, after the trajectory-change the CDA gradually rose (reflecting de-grouping) without a drop. This indicates that the pointers remained intact when the group broke, meaning the group was not allocated its own pointer. Similar results were found when the test required judging the (grouping-irrelevant) location of the entire Kanizsa. Yet, when the test involved judging the (grouping-relevant) orientation of the entire Kanizsa, the trajectory-change triggered a CDA-drop, suggesting that now there was a group-specific pointer. The results suggest that seemingly-similar VWM-grouping can be supported by different pointer dynamics in different contexts, depending on how perceptual factors and task demands join to signal the relevant unit.

56.410 PLANNING FAVORS GIST-LEVEL ENSEMBLES UNDER HIGHER VISUAL WORKING MEMORY DEMAND

Zhuojun Ying¹ (z5ying@ucsd.edu), Marcelo Mattar², Anastasia Kiyonaga¹; ¹UCSD, ²NYU

Multi-step planning is assumed to rely on working memory (WM), but the link between the functions is poorly formalized. In recent work, we developed an information-theoretic model to examine choice behavior and WM for reward information in a sequential planning context. In behavioral testing, participants were asked to choose the most valuable path in a decision tree where nodes varied in value and were revealed sequentially. After choosing, they were asked to recall a subset of node values. Both model and behavior showed better memory for nodes that were more informative to the path choice, suggesting that limited WM is dynamically allocated during planning based on choice-relevance. The model specifies that this resource-rational allocation arises from WM constraints, and predicts that higher WM demands would shift planning toward more gist-level strategies. Here, we manipulated WM demands in two sequential decision-making experiments to test this idea. In Experiment 1, we continuously varied node color to represent values, and we manipulated WM demands by varying tree size and color distinctiveness. Recall was biased toward local (path-level) averages when node colors were more distinct, but toward more global (branch-level) averages when colors

were less distinct, suggesting gist-level encoding under higher visual demands. Path choice also favored an all-or-none satisficing strategy in the latter case, consistent with the gist-level reward encoding. In Experiment 2, multiple features contributed to path quality: both color and angle varied across nodes, but only one feature was cued at the end of each path to determine the path value. Recall for choice-relevant features was biased toward the path average while recall for choice-irrelevant features was biased toward the branch average, suggesting more gist-level encoding of less informative content. Together, these findings show how visual WM demand and planning context interact to shape ensemble encoding and decision-making strategies.

56.411 RACE AND GENDER ARE SPONTANEOUSLY ENCODED IN VISUAL WORKING MEMORY (VWM)

Lauren Williams¹ (wlauren00@gmail.com), Xin Yang, Joshua Langfus², Yarrow Dunham³, Justin Halberda¹; ¹Johns Hopkins University, ²University of California, San Francisco, ³Yale University

How early in visual processing are faces marked for category membership like race and gender? We addressed this question using change detection paradigms adapted from visual working memory literature, where one face from an array of faces changes to another. Crucially, the two faces could come from the same or different gender or racial categories. We predicted that, if gender and race are coded in VWM, participants would be faster to find the changing face when it crosses one of these boundaries. With both accuracy (Study 1) and response time (Studies 2 and 3) measures, we found that people spontaneously categorize faces based on race and gender in VWM, detecting changes faster and more accurately when they cross category boundaries. This pattern persisted in both standard change detection (Study 1) and flicker change detection (Study 2) paradigms, and occurs even when participants complete only a single trial (preventing shifts to higher-level strategies; Study 3). To rule out the possibility that the positive effects that we observed can be explained by lower-level perceptual features or higher-level subjective features associated with faces, we incorporated a large set of such features into our analysis (luminance, feature shapes, affect, attractiveness, etc.). Including these features in our primary analysis did not diminish the predictive power of crossing a category-boundary; thus, the effects in working memory that we report are not reducible to these rich sets of features. Importantly, we find this in a working memory task that does not require conscious attention to categories (providing evidence of spontaneity), that does not appear to be reducible to effects of lower-level perceptual similarities (providing evidence of categorization over and above perceptual similarity), and in a task in which attention to categories is not task-relevant and cannot in general improve performance if employed as a deliberate strategy.

Object Recognition: Models

TUESDAY, MAY 19, 2:45 – 6:45 PM, PAVILION

56.412 IDENTIFYING THE VISUAL FEATURES OF EUROPEAN PALEOLITHIC CAVE PAINTINGS THAT ARE DIAGNOSTIC OF CATEGORY, AGE, AND LOCATION

David Tomz¹ (dtomz@stanford.edu), Kushin Mukherjee¹, Isobel Wisher^{2,3,4}, Frederieke Wulff³, Laerke Braedder³, Chuan Yan⁶, Barbara Tversky^{1,5}, Riccardo Fusaroli^{3,4}, Kristian Tylén^{3,4}, Judith E Fan^{1,6,7}; ¹Department of Psychology, Stanford University, ²Department of Archaeology and Heritage Studies, Aarhus University, ³Department of Linguistics, Cognitive Science and Semiotics, Aarhus University, ⁴Interacting Minds Center, Aarhus University, ⁵Teachers College, Columbia University, ⁶Department of Computer Science, Stanford University, ⁷Graduate School of Education, Stanford University

Humans living during the Upper Paleolithic period (45,000 to 12,000 years cal.BP) and located in Europe were among the earliest humans to produce figurative art. Analyzing the way these early depictions carry meaning could advance understanding of human visual abstraction—the capacity to produce a set of visual marks to represent an abstract idea. In this study, we analyzed 140 Paleolithic cave paintings of animals from four cave sites in Monte Castillo, Spain. These depictions were photographed, then traced by archaeologists specializing in cave art from this period. Based on all available evidence, archaeologists have attributed each depiction as representing one of four animal species (i.e., Horse, Bison, Hind, Ibex) and being a particular age (Gravettian: 31,500-24,000 cal.BP; Solutrean: 24,000-20,000 cal.BP; Magdalenian: 20,000-13,000 cal.BP). Because these experts had access to a wide array of information beyond the visual properties of each depiction, it was unknown what information directly available in these images is associated with their age, location, and which animal species they depict. To address this question, we leveraged state-of-the-art computer vision algorithms (ResNet, ViT, SigLIP) to measure how reliably age, animal species, and location information could be decoded from only the visual features of these cave depictions. We found that all of these properties could be decoded with above-chance accuracy (age: 74.7%; location: 68.0%; species: 63.4%, $p < .01$), and that some regions were more diagnostic than others for predicting age (head>legs: $t=4.398$, $p<.001$) and location (head>legs: $t=3.716$, $p<.001$). Taken together, these findings highlight the promise of developing computational models that capture how observable variation in early cave depictions is linked to major cultural changes that unfolded throughout the Upper Paleolithic period.

EU ERC consolidator grant #101044626 to KT; U.S. NSF CAREER Award #2436199, U.S. NSF DRL #2400471, and a Hoffman-Yee grant from the Stanford Human-Centered AI Institute (HAI) to J.E.F.

56.413 THREE-DIMENSIONAL SHAPE CUES AFFECT HUMAN AND ARTIFICIAL RECOGNITION SYSTEMS DIFFERENTLY

Nicholas Baker¹, Mikayla Cutler, William Friebe, Luke Baumel, Joseph Tocco, George Thiruvathukal; ¹Loyola University of Chicago

Humans and neural networks use shape and texture information differently. While shape is the primary cue in human object recognition, neural networks are more biased towards texture cues. Many tests of shape vs. texture bias have focused on shape recognition from an object's external contour. However, shape information is also conveyed through internal contours, shading, and attached shadows, especially when an object is viewed from noncanonical perspectives. Using models from ShapeNet, we created datasets of 120,000 texture-substituted images of objects from many viewpoints with and without shading and attached shadows. We tested humans' and several neural networks' ability to classify these objects by both their shape and their texture. Humans were much better at classifying texture-substituted objects by their shape than any network, although these differences were greater when shape was defined only by the external contour than when 3D cues were included. Our findings suggest that networks' texture bias is reduced when 3D cues are included in images. We next tested whether the inclusion of 3D cues benefitted humans and neural networks more for images of objects viewed from canonical or noncanonical perspectives. Consistent with earlier research, we found that 3D cues primarily benefitted humans for noncanonical images. For neural networks, the greatest performance gains were for canonical images. These findings suggest fundamental differences in how humans and networks use shading and attached shadows for object recognition. We argue that humans use these cues to infer objects' 3D structures while neural networks use them as another surface-level cue like texture.

56.414 A DEEPER LOOK INTO OCCLUSION TYPES AND THEIR IMPACT ON OBJECT RECOGNITION MODELS

Courtney M. King¹ (cking74@fordham.edu), Elissa Aminoff¹, Daniel Leeds¹; ¹Fordham University

Modern computer vision offers a wide variety of models for recognition and object detection, with some being similar in design to biological vision. Partial occlusions have been shown to have a negative impact on performance in both cases. Similarly, color can be used to create illusions in images and instances of camouflage, which can also have an impact on these two tasks. This work explores several types of occlusions that can make objects appear illusory due to their color with respect to the greater image context. Specifically, we study cases where the occluder blends in with an object or when the occluder blends in with the background. These cases are interesting as they find commonality with optical illusions in human perception. In experiments, these artificial occluders are applied to 80 distinct objects as horizontal lines, with an occlusion ratio of approximately 0.25. Current experiments explore the impact of these types of occlusions on object recognition for several state-of-the-art computer vision models, ranging from simple convolutional neural networks (CNNs) to transformers to more recent "transformer-like" CNNs. This work also offers a look inside these diverse models to visualize how the occlusion-based errors propagate through the networks to the final layer. Current results find that transformer-based models are far more

robust to these types of occlusions when compared to standard CNN models, like ResNet, but that more recent CNNs like ConvNeXt, show similar robustness to transformers. A second more surprising observation is that these color-based occlusions show significantly higher performance when compared to black baseline occlusions in the CNN models, but similar performance in both the transformer and more recent CNN models. These results give more knowledge about the internal nature of neural networks as well as their robustness to specific types of partial occlusions which are not unseen in real-world scenes.

56.415 BETTER MODELS THROUGH WORSE IMAGES: DEGRADATION TRAINING HELPS ALIGN CNNs WITH HUMANS

Connor Parde¹, Zelin Zhao¹, Frank Tong¹; ¹Vanderbilt University, Department of Psychology

Human object recognition is remarkably robust to blur, noise, and adverse environmental conditions. A central question in vision science is how the visual system develops such robustness, and whether artificial systems can acquire similar tolerance through targeted training. Convolutional neural networks (CNNs), by contrast, often fail when presented with degraded input, raising the question of which training regimes yield robustness patterns that enhance performance and align CNNs with human perception. Here, we directly compared human blur and noise thresholds with the robustness profiles of CNNs trained on degraded images. Human participants (n=20 per condition) performed a progressive-revelation task using 800 images from 16 ImageNet categories, in which stimuli began as extreme blur or noise and gradually became clearer. Recognition thresholds were defined as the degradation level at which the correct category was identified. AlexNet and VGG-19 models were trained on the same categories under multiple regimes: standard images; blur; Gaussian noise; uniform noise; and adverse-weather corruption (50% or 75% snow/fog/rain). CNN thresholds were computed analogously by identifying the highest degradation level that permitted accurate classification. Blur training produced the broadest robustness gains, generalizing across both blur and noise. Noise training strongly enhanced noise robustness and modestly improved blur tolerance. Adverse-weather training conferred noticeably weaker gains. Models trained on Gaussian or uniform noise outperformed those trained on naturalistic weather corruptions. This disparity appears to reflect the severity of the degradation, with stronger distortions producing more robust, human-aligned representations. Further, networks trained on milder noise levels showed substantially reduced robustness, comparable to the weather-trained models. Critically, all degradation-trained networks showed stronger correlations with human thresholds than standard-trained models, with noise-trained networks achieving the highest correspondence and blur-trained networks also substantially improving alignment. These results suggest that access to strongly degraded input may play a key role in shaping robust, human-like visual systems.

This research was supported by NIH grants R01EY035157 and R01CA240274 to FT, and P30EY008126 to the Vanderbilt Vision Research Center.

56.416 CRITICAL VIEWING DISTANCE FOR OBJECT RECOGNITION UNDER DEGRADED VISION

Rui Jin¹ (jery18656908019@gmail.com), Gordon Legge², Roberto Manduchi³, Yingzi Xiong¹; ¹Wilmer Eye Institute, Johns Hopkins University, ²Department of Psychology, University of Minnesota, ³Department of Computer Science and Engineering, University of California, Santa Cruz

Low vision individuals usually require closer viewing distance than those with normal vision for object recognition. We define critical viewing distance (CVD) as the minimum distance at which an individual can correctly recognize an object, a crucial determinant for safe and effective navigation. In real-world navigation, objects are often viewed dynamically with gradually decreasing distances due to self or object motion, rather than statically at fixed viewing distances. Whether such dynamic viewing aids object recognition compared to static viewing, especially with degraded vision, is not well understood. Ten participants were asked to recognize common navigation icons (e.g., no crossing, no parking) on a computer display. The icons were chosen through a pilot study to ensure similar difficulty and filtered digitally to simulate low vision with 0.8 logMAR and 0.3 logCS. To measure CVD, the icons expanded in size, mimicking a gradual decrease in viewing distance from 7.2m to 1m by a constant speed (0.7m/s). Participants were asked to stop the trial as soon as they could recognize an icon, and CVD was computed as the equivalent viewing distances. Participants also performed two static viewing conditions: brief viewing (500ms) and unlimited viewing, at five equivalent viewing distances from 5m to 1m, with recognition accuracy obtained at each distance to derive psychometric curves. Across participants, CVD under dynamic viewing averaged 2.93m with a recognition accuracy of 96.5%, which was significantly higher than both static conditions (65.4% and 90.6%, $p < .01$) at each participant's CVD level. Modeling results showed that such benefits of dynamic viewing could not be explained by probability accumulation of brief viewing samples ($p < .001$). Our findings provide initial evidence on the importance of dynamic viewing for optimizing object recognition. Ongoing studies are expanding to complex objects, under self-motion, and with actual low vision in real world navigation.

56.417 A GEOMETRIC FRAMEWORK FOR TESTING EUCLIDEAN AND HYPERBOLIC STRUCTURE IN NEURAL VISUAL REPRESENTATIONS

Yifei E. Chen^{1,2} (evchen@uchicago.edu), Andrew J. Stier³, Marc G. Berman^{1,4}, Wilma A. Bainbridge^{1,2,4}; ¹University of Chicago, ²Institute for Mind and Biology, ³Santa Fe Institute, ⁴Neuroscience Institute

Understanding the geometry underlying neural representational spaces for complex visual knowledge is essential for explaining how the brain organizes information. Most representational frameworks

assume a Euclidean geometry, inherited from linear encoding models and classical multidimensional scaling (MDS). Yet, recent work (Lee et al., preprint) showed that human similarity judgments exhibit hyperbolic-like curvature, motivating the hypothesis that neural representations themselves may be better captured in non-Euclidean spaces. Here, we present the first fMRI pipeline for systematically evaluating Euclidean versus hyperbolic geometry within a region-of-interest (ROI) parcellation scheme. Using the THINGS-fMRI dataset (Hebart et al., 2023), we constructed concept-level neural representational dissimilarity matrices (RDMs) for each Harvard–Oxford region by averaging voxel-wise responses across stimuli belonging to the same concept. We then applied ALBATROSS (Stier et al., preprint), a persistent-homology method that infers each model's optimal dimensionality by minimizing topological divergence. For each ROI, Euclidean and hyperbolic candidate models were compared using a χ^2 divergence metric integrating three measures: Integrated Betti Value (IBV), L1 distance, and Wasserstein-1 (W1). Next, we generated embeddings using geometry-appropriate MDS and assessed model fidelity with stress (global metric preservation) and Spearman rank fidelity (relative-distance preservation). Hyperbolic advantages were evident across posterior temporal and parietal association cortices involved in high-level visual processing and semantic integration—regions where object representations become relational and conceptually structured. More broadly, hyperbolic embeddings yielded lower stress across nearly all ROIs, indicating a superior global geometric fit to empirical neural dissimilarities. In contrast, Euclidean and hyperbolic embeddings exhibited similar Spearman rank fidelity, suggesting that both spaces preserve local ordering of representational distances to a comparable degree. Together, this pipeline provides a principled, geometry-aware framework for testing whether neural representations exhibit Euclidean or hyperbolic structure, opening new possibilities for mechanistic modeling and the discovery of representational principles that may be invisible under traditional Euclidean assumptions.

56.418 WHEN MODELS SEE WHOLE: A MECHANISTIC ACCOUNT OF HOLISTIC PROCESSING IN DEEP VISION MODELS

Fenil Doshi^{1,2}, Thomas Fel^{1,2}, Talia Konkle^{1,2}, George A. Alvarez^{1,2}; ¹Harvard University, ²Kempner Institute for the Study of Natural and Artificial Intelligence

Human object recognition relies on configural processing, the ability to integrate local parts into coherent global shapes. In contrast, contemporary deep neural networks often rely heavily on local texture cues, leaving open whether any current architectures support human-like holistic shape perception. Here we introduce (i) a new measure of configural sensitivity using visual anagrams and (ii) a mechanistic account of the representational transformations that give rise to holistic processing in models. To isolate configural processing, we constructed visual anagrams: image pairs that depict different objects but are assembled from the exact same set of local patches

rearranged into different global layouts. We quantify Configurational Shape Score (CSS) as the ability to classify both members of each anagram pair. Because local patches are matched, classifying both images requires sensitivity to global configuration. Across 86 pretrained models, we find a continuum of holistic competence: self-supervised and language-aligned ViTs show strong configurational sensitivity, whereas supervised CNNs and supervised ViTs fail, even when matched in accuracy. CSS also predicts performance on other tasks including noise robustness, foreground bias, and phase dependence. To examine how high-CSS models implement configurational processing, we developed a mechanistic interpretability method that disentangles positional structure from semantic content, and tracked how these signals evolve across layers. In high-CSS models, positional structure remains geometrically intact into deeper layers, preserving the spatial scaffold needed for part-whole reasoning, while content representations become contextually enriched through long-range interactions. The contextual enrichment is supported by self-attention mechanisms, with the intermediate blocks forming a causally indispensable stage: removing contextual influences in these blocks collapses CSS. Overall, our findings suggest that holistic perception arises when a system maintains precise spatial relationships while contextually enriching local feature representations from long-range interactions, offering a mechanistic account of the mid-level computations that assemble local parts into structured global percepts.

1. NSF PAC COMP-COG 1946308 to GAA 2. Kempner Institute Graduate Fellowship to FD

56.419 ADOPTING A HUMAN DEVELOPMENTAL VISUAL DIET YIELDS ROBUST AND SHAPE-BASED AI VISION

Zejin Lu^{1,2}, Sushrut Thorat², Radoslaw Cichy¹, Tim Kietzmann²; ¹Freie Universität Berlin, ²Osnabrück University

Despite years of research and the dramatic scaling of artificial intelligence (AI) systems, a striking misalignment between artificial and human vision persists. Contrary to humans, AI relies heavily on texture-features rather than shape information, lacks robustness to image distortions, remains highly vulnerable to adversarial attacks, and struggles to recognise simple abstract shapes within complex backgrounds. To close this gap, here we take inspiration from how human vision develops from early infancy into adulthood. We quantified visual maturation by synthesising decades of research into a novel developmental visual diet (DVD) for AI vision. Guiding AI systems through this human-inspired curriculum, which considers the development of visual acuity, contrast sensitivity, and colour, produces models that better align with human behaviour on every hallmark of robust vision tested, yielding the strongest reported reliance on shape information to date, abstract shape recognition beyond the state of the art, and higher resilience to image corruptions and adversarial attacks. Our results thus demonstrate that robust AI vision can be achieved by guiding how a model learns, not merely how much it learns, offering a resource-efficient route toward safer and more human-like artificial visual systems.

56.420 EXPLORING INDIVIDUAL DIFFERENCES IN DNN REPRESENTATIONS

Yinuo Peng¹, Jason Chow¹, Thomas Palmeri¹; ¹Vanderbilt University

Deep neural networks (DNNs) are used to explore individual differences in visual cognition. We quantitatively explore differences in representational dissimilarity matrices (RDMs) for (1) the same objects through different model architectures (19 different AlexNet-VGG, ResNet, Transformer, InceptionNet, CLIP, and CORnet models), (2) different images sampled from the same object categories (from ImageNet), and (3) the same objects subject to the kinds of transformations used as augmentation during training. We previously explored differences in DNN representations caused by random differences in initial weights and training image order in (small) trained DNNs, arguing that meaningful individual differences in representations attributable to what might be considered trait differences needed to exceed a baseline comparable to what might be considered state differences. We chose image transformations (augmentations) that maintained >95% network classification performance. We first examined robustness of object representations (penultimate layer representations) to transformations, testing N=19 models from different architectural families on 50 image sets and 100 image transformation sets, comparing RDMs of the original image sets with transformations of each image set. We observed large differences in robustness of representations to image transformation across different model architectures, with linear mixed-effects modeling showing that variance in RDM distances was largely attributed to model architecture (96.5%) with little variance from image sampling (1.3%) or transformation set (2.3%). We then relate differences in object representations (RDMs) relative to these baseline transformations between different image samples from the same category and between different model architectures. Representational differences across image samples exceed those caused by modest image transformation (akin to state differences). Representational differences across models diverge and depend on architectural differences between models; for most model pairs compared, especially those differing vastly in architecture, representational differences exceed those caused by image transformation and by differences in image sampling.

56.421 NATURAL SHAPE FEATURES FACILITATE OBJECT REPRESENTATION IN CORTICAL AREA V4 AND ARTIFICIAL NEURAL NETWORKS

Dagmawi N. Wube¹ (dnw698@uregina.ca), Najib J. Majaj², Timothy D. Oleskiw^{1,2}; ¹Department of Computer Science, University of Regina, ²Center for Neural Science, New York University

Planar shape, or the silhouette contour of a solid body, carries rich information important for object recognition, including both local (curvature) and global shape cues. While neurons selective for shape are found within the intermediate cortical area V4 of primates, it remains unknown how populations of these neurons contribute to our perception of objects in natural environments. Recently, we have used

a unique array of shape stimuli that dissociates local and global cues to identify V4 neurons that preferentially respond to natural shape features (Oleskiw et al., 2023). To investigate the role of this tuning toward a population-level object code, we analyzed simultaneous activity recorded from a 96-channel 'Utah' array implanted in area V4 of a juvenile macaca nemestrina observing natural and synthetic shapes. Linear classifiers trained to identify these stimuli decode natural shapes from our V4 population twice as accurately (Mann-Whitney, $p < .05$) than synthetic shapes that lack natural cues. Furthermore, a correlation analysis suggests that the V4 population activity is robust to trial-to-trial variability: the decoding of synthetic shapes, including those with identical local curvature statistics, is significantly more sensitive to shared neural noise. To explore the computations underlying a population shape code for natural objects, we repeated our analysis in silico with a variety of artificial neural networks (ANNs) trained for object recognition. Interestingly, while many convolutional networks, including AlexNet and VGG, contain intermediate layers from which natural shapes are decoded somewhat more effectively than synthetic controls, modern transformer architectures (i.e., ConvNeXT) readily utilize natural shape features, demonstrating a 43.5% and 63.1% ($p < .01$) increase in classification accuracy for local and global cues, respectively. Together, our findings reinforce the notion that modern ANNs are compelling models for understanding the integration of natural shape features seen in V4 neurons.

Faculty of Graduate Studies and Research, University of Regina

56.422 TASK-DRIVEN RECURRENT DEMANDS REDUCE ANN ALIGNMENT WITH PRIMATE IT

Ezgi Fide¹ (ezgifide@yorku.ca), Sabine Muzellec², R. Shayna Rosenbaum¹, Kohitij Kar²; ¹York University, Department of Psychology and Centre for Vision Research, Centre for Integrative and Applied Neuroscience, Toronto, Canada, ²York University, Department of Biology and Centre for Vision Research, Centre for Integrative and Applied Neuroscience, Toronto, Canada

Are we overlooking crucial insights by evaluating current artificial neural network (ANN) models of primate vision primarily based on neural responses recorded during passive stimulus viewing? Much of the field's recent progress relies on widely used benchmarks, such as Brain-Score, Natural Scenes Dataset, and THINGS, which assess model-brain similarity using data collected while subjects passively view images. Yet real-world vision is goal-directed, suggesting that current benchmarking practices may overlook critical task-dependent computations. To address this gap, we recorded neural activity from the inferior temporal (IT) cortex of two macaque monkeys ($n=155$ and 109 reliable sites) during both passive viewing and an active object discrimination task, using the same set of 1320 images. We then compared the accuracy of 38 state-of-the-art ANNs in predicting IT responses across task conditions. Given these models' optimization for object recognition, we expected neural predictivity to be higher during active task engagement. However, cross-predictivity, ANNs

predicting IT (forward predictivity, FP), and IT predicting ANNs (reverse predictivity, RP) were modest but significantly higher during passive than active viewing (FP: $t(37)=15.8$, $p<0.001$; RP: $t(37)=-0.196$, $p=0.577$). This reduction in predictivity was substantially greater for late (170-200ms) than for early (70-100ms) phase responses (FP: $t(37)=61.6$, $p<0.001$; RP: $t(37)=15.1$, $p<0.001$), consistent with models' known limitations in capturing recurrent computations. In contrast, the IT responses measured during active task engagement better predicted behavioral accuracy patterns across images. Moreover, transforming ANN activations using the learned mapping between passive and active IT responses significantly improved the ANN's behavioral predictivity, suggesting that ANN representations lack task-dependent components present in IT. The modest discrepancies between active and passive conditions affirm the value of large-scale passive datasets, which capture core feedforward IT structure for high-throughput benchmarking. Rather than rejecting this approach, our results highlight task-related gaps and the need to model task-driven recurrent computations in next-generation vision models.

KK is supported by funds from the Canada Research Chair Program (CRC-2021-00326), NSERC (RGPIN-2024-06223). EF is supported by the TBS Grant (York University). RSR acknowledges support from a York Research Chair, and NSERC. The CFREF VISTA and Connected Minds programs support KK, EF, SM, and RSR.

56.423 THE IMPACT OF RECURRENT CIRCUITRY ON EMERGENT ORTHOGONAL CATEGORY STRUCTURE IN DEEP VISION MODELS

Kexin Cindy Luo^{1,2} (kexinluo@g.harvard.edu), George Alvarez^{1,2}, Talia Konkle^{1,2}; ¹Harvard University, ²Kempner Institute for the Study of Natural and Artificial Intelligence

The human visual system contains rich feedforward, feedback, and bypass pathways that shape how category information is represented across the ventral stream. Motivated by these biological motifs, many deep neural networks now incorporate recurrent circuitry, but it remains unclear how these architectural choices reshape category representations. Here, we use representational geometry analyses to ask how models with different feedback motifs change the global layout of categories in deep networks, relative to purely feedforward architectures. We first examined a standard feedforward AlexNet backbone and two variants with multiplicative or additive long-range top-down connections (LRM/LRA; Konkle & Alvarez, 2023). Comparing the distribution of pairwise cosine distances between category prototypes, the purely feedforward AlexNet showed a spread-out, flat distribution of between-category distances ($M=.81$, $SD=.37$). In contrast, both LRM and LRA models exhibited sharply peaked distributions tightly clustered near 1 ($M=.96$, $SD=.12$), approximating a near-orthogonal simplex in decision space. This divergence was not present at the penultimate layer, where all three models showed similar distributions with high cross-model RSA ($\sim r=.96$). To determine whether near-orthogonality is a signature

general to any model with recurrent circuitry, we examined a broader set of networks, including CORnet-RT (local additive feedback), ResNet-50 (skip forward connections), and VGG16 (pure feedforward). Contrary to our initial expectation, we found that only ResNet-50 showed a near-orthogonal category geometry at the logit layer, whereas CORnet-RT and VGG16 resembled the flatter feedforward AlexNet profile. Interestingly, both Resnet50 and the LRM/LRA models have a channel-to-channel bypass motif that mixes information across non-adjacent hierarchical layers. Taken together, these results raise the possibility that bypass connections provide pressure on the decision stage to form nearly orthogonal category prototypes. This may have benefits for simplified downstream readout while still allowing earlier layers to maintain richer similarity structure. The mechanism underlying this potential bypass-geometry effect is currently unknown.

This work was supported by the Kempner Institute Graduate Fellowship (to K.C.L.), CRCNS-2309041 (to T.K., G.A., & H.P.), and funding from the Kempner Institute (to T.K.).

56.424 OBJECT AND SCENE RECOGNITION ABILITIES PREDICT THE CONTENT AND QUALITY OF IMAGE DESCRIPTIONS

Melina O. Mueller¹ (melina.o.mueller@vanderbilt.edu), Ian G. Dobbins², Isabel Gauthier¹; ¹Vanderbilt University, ²Washington University in Saint Louis

Do non-verbal visual abilities predict how people describe what they see? We measured novel object recognition (o; simultaneous matching to silhouettes, sequential matching, and oddball tasks), scene recognition (sequential matching and oddball tasks with prepared food and outdoor images), and intelligence (g) in 101 adults, then asked them to describe 10 complex images (e.g., market, coffee shop, abstract sculpture), from memory or during perception. We examined the quality of descriptions via semantic similarity scores to a large language model's descriptions and the content of descriptions via feature extraction (e.g., feature descriptors, hedging, spatial language). With hierarchical regression, we controlled for age, gender, vocabulary, description length, g (and o for scenes as a predictor). People who were better at recognizing scenes provided higher quality descriptions of images while viewing them ($\Delta R^2 = .093$, $p = .003$) or from memory ($\Delta R^2 = .125$, $p < .001$). Additionally, visual abilities predicted how people described images. Those with a higher o made more references to shape ($\Delta R^2 = .033$, $p = .048$) and spatial relations ($\Delta R^2 = .040$, $p = .014$), and used less hedging ($\Delta R^2 = .052$, $p = .016$), in memory-based descriptions. Those with higher scene recognition made more references to texture ($\Delta R^2 = .062$, $p = .022$) and to spatial relations ($\Delta R^2 = .035$, $p = .046$) in memory-based descriptions, and used more adjectives in perception-based descriptions ($\Delta R^2 = .062$, $p = .012$). Beyond intelligence and vocabulary, non-verbal perceptual visual abilities predict how we write about images, and the quality of these descriptions. One interpretation is that o supports encoding of richer visual representations that can be accessed during linguistic

output. These results have theoretical implications for both the study of vision and that of language, and practical implications for what we can learn about people using natural language processing.

This work was supported by the David K. Wilson Chair Research Fund and NSF BCS Award 2316474

56.425 THE SHINIER THE BETTER: AN ADAPTATION OF THE SHINE TOOLBOX ON PYTHON

Mathias Salvas-Hébert¹, Nicolas Dupuis-Roy², Catherine Landry¹, Ian Charest¹, Frédéric Gosselin¹; ¹Cerebrum, Université de Montréal, ²Elephant Scientific Consulting

The SHINE toolbox (Willenbockel et al., 2010; >1,350 citations) has played a central role in vision science by enabling researchers to control accidental low-level image properties that might otherwise confound behavioural or neural responses. SHINE provides tools to match or specify Fourier amplitude spectra, enforce rotational averages in Fourier space, specify luminance histograms, and specify a mean and a variance for luminance distributions. These controls remain essential not only for psychophysics but also for modern computer vision mostly developed in Python, where the training data of deep neural networks often contain unintended statistical regularities. We introduce SHINIER, a Python-based re-implementation of SHINE that preserves the full feature set of the original toolbox while extending its capabilities. SHINIER adds support for colour images, improved memory efficiency, multiple histogram-specification algorithms, dithering operations, high-precision numerical routines, and enhanced visualisation tools. It is available on GitHub and PyPI (pip install shinier) and provides legacy-compatible functions to streamline migration from MATLAB. To illustrate its utility for computer vision, we trained two convolutional neural networks (CNNs) on a cat-dog classification task while manipulating luminance histogram statistics. Images followed one of three luminance profiles: negatively skewed (A), positively skewed (B), or their average (C). One CNN was trained on images normalized to C, eliminating luminance-histogram cues; the other was trained with cats mapped to A and dogs to B, introducing a spurious correlation. At test, luminance histogram was decorrelated from category by applying a 50/50 mixture of A and B. The histogram-normalised model generalized well (77% accuracy), whereas the histogram-contaminated model collapsed to chance (50% accuracy), demonstrating how uncontrolled low-level statistics can drive model behaviour. SHINIER provides a robust, accessible solution for eliminating such confounds in both human and machine vision research.

Perceptual Organization: Neural mechanisms, models

TUESDAY, MAY 19, 2:45 – 6:45 PM, PAVILION

56.426 RECURRENT NETWORK CONTRIBUTIONS TO VISUAL RESPONSE DYNAMICS IN MACAQUE IT REVEALED BY TARGETED OPTOGENETICS AND MODELING

Alvin Dinh¹, Behnam Karami^{1,2}, Reza Azadi¹, Arash Afraz¹; ¹National Institute of Mental Health, ²Donders Institute for Brain, Cognition and Behavior

Visual processing is the result of complex interactions among neurons within and across visual areas. In inferotemporal (IT) cortex, neurons show a characteristic temporal response to visual stimuli: spiking activity rises, then declines, and settles into a plateau, typically described by a post-stimulus time histogram (PSTH). Two hypotheses have been proposed for the declining phase: (1) an intrinsic mechanism, where higher activation leads to fatigue and a drop in firing even in the presence of constant input; and (2) a network mechanism, where suppressive contributions from other neurons stabilize activity at a lower, more energy-efficient level. To directly test these hypotheses, we delivered 200 ms optogenetic pulses to a small region of macaque IT cortex transduced with pAAV-CaMKIIa-C1V1(t/t)-TS-EYFP to express an excitatory opsin in pyramidal neurons. We first measured each neuron's typical PSTH in response to visual stimuli. We then trained an adaptive deep neural network model developed by Vinken et al., 2020, fitting its adaptation parameters in the Conv5 layer to match each neuron's visual PSTH. Next, we tested whether these neuron-specific adaptation parameters would also reproduce the optogenetically evoked IT response curves. To do this, we stimulated Conv5 units with square-wave current injections matched to the magnitude of firing-rate modulation produced by optogenetic stimulation and computed the residual difference between the model's prediction and the empirical optogenetically driven dynamics. We find that although at lower stimulation powers, neurons' intrinsic adaptation parameters are sufficient to capture the artificial activation dynamics, they fail as the modulation level increases. These results indicate that intrinsic adaptation alone is not sufficient to explain the visual response dynamics; network interactions are also required, and their effects become increasingly apparent under stronger optogenetic drive.

This research was supported by the Intramural Research Program of the NIMH ZIAMH002958.

56.427 NEURAL BASIS OF PERCEIVED POSITION IN THE FRAME EFFECT

Eunhye Choe¹ (eunhye.choe.gr@dartmouth.edu), Wentao Si¹, Yong Hoon Chung¹, Viola Störmer¹, Patrick Cavanagh^{2,1}, Peter U. Tse¹; ¹Dartmouth College, ²Glendon College

Where in the brain are perceived object positions represented? We used fMRI and the frame-induced position shift (FIPS) illusion that allows dissociating (mis)perceived positions from physical positions of objects to test this. A pair of probe dots were flashed in succession in the upper and then lower visual fields, displacing vertically up and

down with each alternation. At the same time, a surrounding frame moved leftward and rightward in synchrony with the probe presentations. Rather than seeing the probes displaced vertically, they appeared shifted along a slanted trajectory, demonstrating that perceived position is coded relative to the position of the anchoring frame. We collected 3T fMRI data while participants viewed this FIPS stimulus that induced illusory leftward- or rightward-slanted shifts in perceived probe positions. The probes were also shown without the frame, either at the vertically aligned physical positions or shifted corresponding to each participant's reports of the illusory shift determined in a pre-scan psychophysical experiment. Participants performed a fixation brightness task or a change detection task on the probes, ensuring that probe positions were always task-irrelevant during the fMRI scan. Multivoxel pattern analysis was applied to assess whether neural activity elicited by illusory positions generalized to activity patterns evoked by matched physical positions. Matched physical positions were robustly decoded in early visual cortex, but cross-decoding of illusory position and matched physical position only emerged in dorsal occipital and occipitoparietal regions, suggesting shared representations between illusory and physical locations arise in higher-order visual areas. Moreover, using a searchlight analysis, strongest cross-decoding was found in dorsal anterior cingulate, DLPFC and the anterior temporal pole similar to Liu et al (2019). These findings provide further support for the unorthodox view that perceived position is represented at a much higher stage of the visual hierarchy, well beyond early retinotopic cortex.

56.428 THE NEURAL CORRELATES OF METACONTRAST MASKING ARE STIMULUS SPECIFIC.

Giulio Contemori¹ (giulio.contemori@unipd.it), Marianna Musa¹, Martina Passaggi¹, Carolina Maria Oletto¹, Stefano Vicentin², Luca Battaglini¹, Giorgia Cona^{1,3}, Marco Bertamini¹; ¹University of Padova, ²University of Osnabrück, ³Padua Neuroscience Center

Metaccontrast masking reduces target visibility when a spatially adjacent, non-overlapping mask follows the target after a brief delay, yet its functional locus remains debated: does masking mainly disrupt early feedforward encoding, or later feedback-mediated integration? Aydın et al. (2021) isolated nonlinear target-mask interactions using a synthetic-minus-natural ERP measure. In contour discrimination with same-polarity target and mask, they reported a largely monotonic SOA dependence of the interaction signal in an early posterior cluster (160–300 ms), whereas a later centro-parietal cluster (300–550 ms) showed a U-shaped SOA profile paralleling Type-B behavioral masking, suggesting that behaviorally relevant suppression emerges in later, recurrent stages. We asked whether these signatures generalize to stimuli requiring global integration: visual symmetry. First, we confirmed that brief (20 ms) octagonal shapes elicit a canonical Sustained Posterior Negativity (SPN), indexing regularity processing. The SPN is a well-established posterior ERP marker, typically observed as a sustained negative deflection for regular relative to irregular patterns. Next, we presented the same stimuli followed by a

same-polarity, spatially non-overlapping annular mask at SOAs of 10, 50, or 200 ms (plus an unmasked baseline). Behaviorally, symmetry discrimination showed a Type-B (U-shaped) masking function with maximal impairment at 50 ms. We quantified synthetic-minus-natural interaction waveforms within Aydın's predefined windows (160–300 ms; 300–550 ms) and within an SPN-like cluster (276–424 ms) functionally defined from unmasked symmetry–asymmetry differences via cluster-based permutation. The predefined windows showed only modest modulation and did not capture the behavioral trough. In contrast, SPN-like attenuation varied systematically across SOAs, including a robust quadratic (U-shaped) component tracking performance; exploratory brain–behavior analyses linked stronger SPN suppression to lower accuracy. These findings suggest that the electrophysiological “window” of metacontrast suppression is stimulus- and computation-dependent, motivating task-sensitive, functionally defined neural markers when adjudicating feedforward versus feedback accounts of masking.

PRIN Grant No. 2022HEKCWH; PRID initiative (CoPeSy), Department of General Psychology, University of Padova.

56.429 SYMMETRY RESPONSES IN MARMOSSET VISUAL CORTEX MEASURED WITH SSVEPS

Peter J. Kohler^{1,4} (pjkohler@yorku.ca), Amirhossein Asadian^{1,2,4}, Dirk B. Walther^{3,4}, Liya Ma^{1,4}; ¹Department of Psychology, York University, ²Department of Biology, York University, ³Department of Psychology, University of Toronto, ⁴Centre for Vision Research, York University

Symmetry is a biologically significant mid-level visual feature that has been studied for more than a century. Brain imaging has revealed reliable responses to many symmetry types and provided an emerging understanding of the temporal dynamics and brain regions involved. This work has mostly been done in humans, but macaque monkeys have similar responses in homologous brain regions (Audurier et al., 2022). Here we present the first study of symmetry responses in another non-human primate species, the common marmoset. Marmosets have structurally and functionally similar visual cortices to those of humans and macaques, but comparatively small brains with smooth cerebral cortices that enable layer-specific intracranial recordings within homologues of human visual areas that would be hidden in sulci in macaque brains. As a first step in that direction, we recorded EEG data from 32-channel surface grids chronically attached to the skulls of 3 marmosets, and measured symmetry responses using a Steady-State Visual Evoked Potentials (SSVEPs) paradigm well-documented with human participants (Norcia et al., 2002). Our stimuli were a class of regular textures known as wallpaper groups that each contain unique combinations of symmetries. We generated large (24°) textures from two groups that produce robust responses in both humans and macaques: PMM, which contains reflection symmetry, and P4, which contains four-fold rotation symmetry. Our paradigm made it possible to isolate and identify both symmetry-specific brain responses and low-level image-update responses, in the odd and even harmonics of the stimulation frequency (Kohler et al., 2016). Both

groups produced robust image-update responses, as well as reliable symmetry responses within posterior electrodes, with similar temporal dynamics to those found in humans. Our findings show that marmosets produce responses to reflection and rotation symmetry that can be measured with SSVEPs, and open the door to intracranial recordings of spiking activity and local field potentials.

This research was undertaken thanks in part to funding from the Vision: Science to Applications, supported by the Canada First Research Excellence Fund. Additional support was provided by a Natural Sciences and Engineering Research Council of Canada Discovery Grant awarded to PJK.

56.430 THE INFLUENCE OF FAMILIARITY ON SYMMETRY PERCEPTION

Chi T. K. Dao^{1,4} (khchi31@my.yorku.ca), Nihan Alp², Anthony M. Norcia³, Peter J. Kohler^{1,4}; ¹Department of Psychology, York University, ²Psychology Program, Sabanci University, Istanbul, ³Department of Psychology, Stanford University, ⁴Centre for Vision Research, York University

Symmetry and familiarity share important roles in visual perception. Both are detected rapidly and help the visual system resolve the structure of complex scenes. Yet it remains unclear how familiarity interacts with symmetry. Addressing this gap is essential for understanding how symmetry and familiarity jointly shape perceptual organization. Here, we investigate how familiarity influences symmetry responses, using high-density EEG (Study 1) and a psychophysical symmetry-detection task (Study 2). Object recognition depends strongly on configural shape—relationships between local curvatures around the contour. We used three classes of shape silhouettes that vary in configural shape and recognizability, but are well-matched in terms of local features: Intact animal silhouettes, silhouettes with disrupted configural shape (Frankensteins; Baker & Elder, 2022), and curvature-matched controls (CMCs; Elder et al., 2018), which share local curvature statistics with animal silhouettes but lack familiar configural shape. We tiled the shapes to create patterns with and without reflection symmetry. In Study 1 (N = 25), we recorded Steady-Visual Evoked Potentials (SSVEPs) while participants viewed a series of exemplars in which the two pattern types alternated, with each shape class shown on separate trials. The amplitude of symmetry-specific brain responses decreased systematically with familiarity: Lowest for intact, a bit higher for Frankensteins, and highest for CMCs. In Study 2 (N = 14), participants completed a 2IFC task with an adaptive staircase to measure symmetry-detection duration thresholds. Based on the SSVEP results, we predicted longer duration thresholds for more familiar shapes. Frankensteins and CMCs did show the expected pattern, with CMCs having significantly shorter durations ($p=0.029$), but thresholds for intact stimuli were intermediate, not significantly different from the other two conditions. Taken together, our results show that symmetry processing can be influenced by familiarity, but at a suprathreshold level that is not captured by our psychophysical measurements.

This research was undertaken thanks in part to funding from the Connected Minds Program and from the VISTA Program, both supported by the Canada First Research Excellence Fund. Additional support was provided by a Natural Sciences and Engineering Research Council of Canada Discovery Grant awarded to PJK.

56.431 LOW-VARIANCE DIMENSIONS OF CORTICAL ACTIVITY CARRY BEHAVIORALLY RELEVANT INFORMATION

Chihye Han¹ (chan21@jhu.edu), Michael Bonner¹; ¹Johns Hopkins University

A central question in visual neuroscience is how neural population activity supports perceptual behavior. Traditional approaches have focused on interpretable dimensions that explain large portions of neural variance, such as those corresponding to object categories or semantic properties (Huth et al., 2012; Tarhan & Konkle, 2020; Hebart et al., 2020). Here, we tested whether perceptual structure is concentrated in these leading dimensions or extends throughout the full spectrum of neural population activity. Using fMRI responses to natural scenes from the midventral visual stream (Natural Scenes Dataset; Allen et al., 2022), we applied PCA to identify latent dimensions of neural population activity. For each dimension, we sampled images with extreme values to create target clusters and paired them with randomly sampled foil images. We confirmed that early dimensions (e.g., PCs 1-5) contained interpretable categories like faces, while later dimensions (e.g., PCs beyond rank 10) showed no obvious semantic interpretation. We then asked participants (N=80) to view target and foil image clusters sampled from different dimension ranges (1-10, 10-100, 100-1000) and judge which cluster was more similar to a reference image sampled from the respective ranges. We found that behavioral relevance was just as strong for dimensions 10-100 as for dimensions 1-10. To test whether this extended to individual dimensions and a more challenging task paradigm, we asked participants (N=74) to judge which of two image clusters sampled from individual dimensions was more coherent, without presenting a reference image. Dimensions showing significant behavioral relevance ($p < 0.05$, FDR-corrected) were distributed throughout the entire range tested. These findings demonstrate that behaviorally relevant information extends well beyond interpretable, high-variance dimensions. The striking dissociation between variance and behavioral relevance across individual dimensions suggests that the neural code utilizes information distributed across the full spectrum of representational dimensions, not just a subset of leading components.

56.432 CONTOUR INTEGRATION IN HUMANS AND CNNs

Michael Herzog¹, Ben Lonnqvist¹, Elsa Scialom¹, Abdulkadir Gökce¹, Zehra Merchant¹, Martin Schrimpf¹; ¹EPFL

Contour integration plays a central role in human object recognition, a fact reflected in both ancient cave paintings and modern cartoons. But does the same hold for deep neural networks (DNNs)? To address this question, 50 human participants and 1,038 DNNs performed an object

recognition task using images from 12 everyday categories, such as bananas and shovels. Participants were first shown highly fragmented black-and-white versions of the images, and the degree of fragmentation was gradually reduced until they could identify the object in a 12-alternative forced-choice task. Fragmentation was implemented using either directional line segments or non-directional phosphene-like elements. Both humans and DNNs also viewed the full contour images as well as the original full-color images. While humans and DNNs performed perfectly on the full-color images, some DNNs struggled even with the full-contour stimuli, and all DNNs performed far below human levels on the fragmented images. Most models achieved only slightly above chance performance at all fragmentation levels, whereas human accuracy ranged from roughly 20% to 80% depending on the level of fragmentation. Both humans and some DNNs performed significantly better with line-segment fragmentation than with phosphene fragmentation. These results show, first, that object recognition is possible without contour integration; and second, that DNNs exhibit human-like contour integration only at coarse scales—and in a manner seemingly independent of their overall object recognition ability. This raises the question of whether humans, too, recognize objects primarily through their contours, or whether contour recognition and object recognition are in fact two distinct processes.

56.433 INTERNAL NOISE MAY NOT BE SO INTERNAL

Cherlyn J Ng^{1,2} (cherlyn.j.ng@uci.edu), Jeannie Pham², Marvin Le², Andrew Browne², Jan Skerswetat²; ¹School of Social Sciences, University of California, Irvine, ²Center for Translational Vision Research, University of California, Irvine

Background: Perceptual sensitivity is limited by neural fluctuations called internal noise. Even when signals exceed the noise level, sampling efficiency determines whether the brain uses information in the signal. Accurate measurements of both are important for understanding cognitive pathologies. However, uncorrected refractive errors present in every eye are often overlooked. These degrade the signal quality even before neural processing begins. When accounted for, the brain may be less noisy, more efficient and more similar across individuals than previously thought. Methods: Residual lower- and higher-order refractive errors in both eyes of 16 healthy 18-45yo participants were measured using the iTrace aberrometer in an ongoing study. All had normal or corrected-to-normal visual acuity. Internal noise and sampling efficiency were determined using the AIM equivalent noise tasks for color and motion (Skerswetat et al., 2024) representing mechanistically distinct aspects of vision. Participants wore their habitual prescription eyewear to perform the task and assess symptoms of visual discomfort (Vinkers et al., 2024). Linear regression was used to correlate optical, perceptual, and discomfort measures. Results: Uncorrected refractive errors increased internal noise for both color and motion perceptions. However, different types of refractive errors distinctly affected each domain. Lower-order aberrations (sphere and astigmatism) impaired color vision. Being large and noticeable, they also prompted complaints about clarity ($r=0.66$). In contrast, higher-order aberrations that are present in

minute amounts affected motion perception without reducing clarity ($r=0.80$). This finding was paralleled by sampling efficiency, which was also associated only with color but not motion perception ($r=0.72$). Conclusions: Internal noise and sampling efficiency are often attributed to neural processing. Our findings suggest that focusing errors in the eye are also significant bottlenecks. Effects from disparate types of refractive errors on clarity may reveal neural pathways that use different sources of information for color and motion perceptions.

JS was supported by the Alcon Research Foundation Award awarded to AB. JP and ML were supported by the Undergraduate Research Opportunities Program (UROP), University of California, Irvine. AB was supported by NIH/NEI 1K08EY034912 - 01.

56.434 FROM NEURAL UNITS TO CONSTANT CURVATURE REPRESENTATIONS OF CONTOUR SHAPE

Philip Kellman¹, Austin Phillips¹, Nicholas Baker², Patrick Garrigan³; ¹University of California, Los Angeles, ²Loyola University Chicago, ³St. Joseph's University

How does visual processing transform early encodings—transient responses of orientation-sensitive units—into durable, symbolic representations of object boundaries and shape? Recent research provides clues about how this transition occurs in contour and two-dimensional shape perception. Evidence suggests that an initial symbolic representation of contour shape may consist of segments of constant curvature (Baker, Garrigan, & Kellman, 2021). We hypothesize that these representations arise from banks of constant-curvature filters ("arclets") built from oriented units linked by constant turning angles (Kellman, et al., 2013; c.f., Poirier & Wilson, 2006). Arclets span multiple turning angles and scales; collectively, they force any smooth contours into a representation of constant curvature segments. Baker & Kellman (2021) developed a computational model supporting the plausibility of the general approach and its agreement with psychophysical data. Here, we describe a neurally plausible model that operates on simple images and produces a symbolic encoding of open contours and 2D shapes in terms of constant curvature segments. Each arclet is composed of three co-circular, odd-symmetric, Gabor filters. We use 6 scales, with 6 positive and 6 negative turning angles, plus a zero turning angle, per scale. Filled 2D shapes on homogeneous backgrounds are convolved with each filter type. Activations along the contour, normalized in terms of maximum possible activation, are measured for the best fitting arclet of each type. Arclets of the various turning angle and scale combinations compete to "capture" local contour segments. Model tests show that the resulting representations agree well with human perception, even for arbitrary shapes lacking constant-curvature parts. We further describe how encoding with curvature filters yields a scale-invariant shape code that may underlie the ease with which perceivers detect shape identity and similarity across in absolute size.

NIH R01CA236791 to PK

56.435 DIVERGENT DEVELOPMENTAL TRAJECTORIES OF VISUAL ACUITY AND CROWDING REVEALED THROUGH BIOMIMETIC DEEP NEURAL NETWORK TRAINING

Suayb Arslan^{1,2,4} (suayb.arslan@bogazici.edu.tr), Fazilet Yildirim-Keleş³, Hannan Toprak²; ¹Department of Computer Engineering, Boğaziçi University, ²Institute for Data Science and Artificial Intelligence, Boğaziçi University, ³Department of Psychology, Boğaziçi University, ⁴Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

Visual acuity and crowding develop along different timelines in childhood. Acuity reaches adult-like levels by approximately 5–6 years, while crowding—the deleterious influence of nearby elements on object recognition—continues to improve beyond this age, suggesting that these functions rely on different stages of visual cortical development. To investigate the mechanisms underlying this dissociation, we implemented a biomimetic deep neural network that mimics hierarchical visual processing, from the retina through primary (V1) and intermediate (V2–V4) visual areas. The network incorporated foveated retinal sampling, V1-like oriented filters, and higher-level convolutional layers whose pooling, lateral interactions, and cortical magnification could be selectively adjusted to simulate developmental progression. We evaluated network performance at three simulated developmental stages. Acuity was measured by the model's ability to identify isolated targets, while crowding was measured using targets flanked by surrounding stimuli, quantifying critical spacing. At the early stage, the network's retinal and V1-like layers were fully mature, whereas higher-level layers remained immature. In this stage, the network achieved adult-like acuity but exhibited large critical spacing, consistent with prolonged crowding deficits in children. At intermediate stages, partial maturation of higher layers improved crowding performance, and at the final stage, fully matured higher layers yielded both adult-like acuity and crowding. Systematic manipulations revealed that increasing pooling size, strengthening lateral inhibition, or refining cortical magnification in higher layers reduced critical spacing without affecting acuity. Networks lacking hierarchical constraints failed to reproduce the developmental dissociation. These results indicate that early maturation of retinal and V1-like processing explains the rapid development of acuity, whereas delayed refinement of spatial integration in higher visual areas accounts for prolonged crowding. Biomimetic neural networks provide a precise framework to link neural maturation to behavioral outcomes and generate testable predictions for human visual development.

This work is supported by TUBITAK with grant number 125E068.

56.436 REPRESENTATIONAL MOMENTUM IN 2D VISUAL FEATURE SPACE FOLLOWS A FEATURE-SELECTION STRATEGY

Sedthapong Chunamchai^{1,2,3,4} (sedthapongc@gmail.com), Nithit Singtokum^{2,5}, Pakapon Suesatchapong³, Anthipa

Chokesuwattanaskul^{2,3,4}, Chaipat Chunharas^{2,3,4}; ¹Medical sciences program, Division of Neuroscience, Faculty of Medicine, Chulalongkorn university, Bangkok, Thailand, ²Cognitive Clinical and Computational Neuroscience Center of Excellence, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand, ³Division of Neurology, Department of Medicine, Faculty of Medicine, King Chulalongkorn Memorial Hospital, Thai Red Cross Society, Bangkok, Thailand, ⁴Chula Neuroscience Center, King Chulalongkorn Memorial Hospital, Bangkok, Thailand, ⁵Department of Physiology, Faculty of Medicine, Chulalongkorn university, Bangkok, Thailand

Representational momentum (RM) describes a robust perceptual bias in which memory for a changing stimulus is displaced forward along its trajectory. Although well established for single-feature representations, less is known about how RM operates in multidimensional feature spaces and whether the effect reflects integrating information across features or reliance on a dominant one. We examined RM in tasks where visual stimuli changed either along one dimension—size (log scale) or orientation (absolute degrees)—or simultaneously along both. Participants completed 288 one-dimensional (1D) trials and 384 two-dimensional (2D) trials. Stimulus trajectories were created as linear paths in latent 1D or 2D space. On each trial, participants viewed three sequential stimuli sampled along the trajectory, with step length manipulated across low, medium, and high levels and predicted a fourth stimulus using two-AFC response. Foils followed a 2x2 design manipulating deviation direction (Overestimation vs. Underestimation) and target distance (Near vs. Far). Twenty-three adults participated. In 1D trials, accuracy was higher for orientation than size changes ($t = 9.0$, $p < 0.001$), and accuracy on 2D trials exceeded mean 1D accuracy ($t = 6.3$, $p < 0.001$). Across conditions, near target foils elicited more errors than far foils ($F = 158.2$, $p < 0.001$). RM was robust in both 1D and 2D conditions ($F = 30.7$, $p < 0.001$). Notably, RM magnitude decreased as step length increased ($F = 22.8$, $p < 0.001$), suggesting that forward displacement reflects an absolute representational mechanism rather than a proportional scaling with trajectory size. To test whether 2D decisions reflect feature integration, we modeled 2D accuracy using 1D performance. A Feature Selection model—predicting performance based on the more reliable single feature—significantly outperformed a Probability Summation model ($t = 28.1$, $p < 0.001$). This indicates that participants predominantly relied on the most diagnostic feature rather than combining information across dimensions.

56.437 THE MULTISCALE ENTROPY-REGULARIZED SYMMETRY (MERSYMM) ALGORITHM FOR THE GENERAL EXTRACTION OF GESTALT PRINCIPLES AND SYMBOLIC KNOWLEDGE FROM IMAGES

Brody Kutt¹, Zygmunt Pizlo¹; ¹University of California, Irvine

How humans form structured perceptions of visual stimuli and judge objectness are long-studied questions in psychophysics. According to the Gestalt laws of perception, the visual system ignores low-level details. It extracts global, simplified symbolic knowledge, which is

critical to higher-level cognition, such as concept formation and reasoning. To predict what humans are likely to perceive, prior work cast the problem as a search task, often using complex, hand-crafted scoring procedures that were limited to line drawings. In contrast, this work introduces a closed-form algorithm founded on domain-general visual information processing operations without requiring optimization, search, or training data. MERSymm is an algorithm for assessing the likelihood of a percept grounded in the idea that human judgments of objectness reflect compressible regularities. MERSymm takes as input a mathematical structure, which is a candidate percept of an image. MERSymm detects Euclidean symmetries as indicators of information redundancy using a parallelizable multiscale pyramid and produces a prior probability of the percept for Bayesian inference. We evaluate MERSymm and off-the-shelf compression algorithms on a custom benchmark. Algorithms must identify which perceptual object in a pair exhibits properties associated with objectness (organized into eight Gestalt principles: symmetry, closure, proximity, similarity, continuity, uniform connectedness, common region, and figure-ground). MERSymm detects these eight Gestalt principles nearly perfectly in their basic presentations (currently $99.28\% \pm 0.011\%$ detection accuracy), unifying them for the first time as arising from a symmetry-based procedure. MERSymm also remains robust under different noise conditions. It captures global compressible regularities that existing compression algorithms miss. Existing algorithms achieve 56%-74% average accuracy. MERSymm is an open-domain, massively parallelizable, and generally applicable algorithm for mirroring human judgements of objectness based only on fundamental visual processing operations. This work has implications for Bayesian models of perception and developing human-like, neurosymbolic AI and computer vision systems.

56.438 COMPARING BRAIN, HUMAN, AND MACHINE PERCEPTUAL SIMILARITY OF VISUAL IMAGES

Daniel Chong¹ (djc473@cornell.edu), Raja Marjeh², Ilia Sucholutsky³, Nori Jacoby¹, Amy Kuceyeski¹; ¹Cornell University, ²Princeton University, ³New York University

Understanding how humans judge the similarity of visual images is of interest in computer vision, neuroscience, and cognitive science. While some research has shed light on how brain activity representations in response to images are related to subjective human similarity judgements, a full picture is yet to be realized. In this study we use a brain voxel-level encoding model to quantify the shared variance of human similarity judgements and the brain's visual regions activity similarity. We analyzed 100 images from the BOLD5000 dataset (4,950 unique image pairs). For each pair, we had three sources of similarity: human ratings (0-6 scale, averaged across 3-6 raters), AI models (573 vision models, 64 language models), and predicted brain activity patterns derived from a voxel-wise encoding model. The encoding model allowed us to predict voxel-level fMRI responses and compute pairwise brain similarity within 7 early visual regions (V1d, V1v, V2d, V2v, V3d, V3v, hV4) and 12 late visual regions

(FFA1, FFA2, OFA, EBA, FBA2, PPA, RSC, OPA, OWFA, VWFA1, VWFA2, mfswords). We correlated each brain region's similarity structure with human judgments, vision models, and language models. Late visual regions were significantly more correlated with human judgments compared to early visual regions (mean $r=0.36$ vs. 0.28 ; $t(17)=4.04$; $p<.001$), with mid-fusiform sulcus word area, parahippocampal place area, and ventral word form area 2 showing the strongest correlations ($r=0.406$, 0.401 , 0.391 ; $p<.001$). Correlations with vision AI models were not significantly different between early and late regions (mean $r=0.179$ vs 0.184 ; $t(17)=0.505$ $p=0.620$), while language AI models showed stronger correlations with late semantic regions compared to early visual regions (mean $r=0.27$ vs 0.19 ; $t(17)=5.61$ $p<.001$). This suggests that human similarity perception is grounded in semantic representations that are also found in higher level visual regions, and that language models better mimic this semantic structure compared to vision models.

56.439 ATTENTION MODULATES BOTH SERIAL DEPENDENCE AND CENTRAL TENDENCY BIAS IN SPATIAL JUDGMENTS

Saija M. Niemi¹ (saija.m.niemi@helsinki.fi), Toni P. Saarela¹, Maria Olkkonen¹; ¹University of Helsinki

Visual judgments are biased toward stimuli from preceding trials (serial dependence) and toward the overall stimulus mean (central tendency bias). Whether these two sequential effects reflect a shared mechanism remains unclear. We investigated whether they are similarly affected by attention in spatial judgments by manipulating task and stimulus distributions. In two separate sessions, sixteen participants estimated the width or the hue of colored ellipses in alternating trials using the method of adjustment. Each trial began with a cue to attend to either width or hue, followed by a stimulus for 500 ms and a 1250 ms delay. Participants then adjusted a probe to match the cued feature. In width-attended trials in both sessions, ellipse widths were uniformly distributed between 4.5 and 7.5 degrees of visual angle. Depending on the session, the widths in hue-attended trials were randomized from different distributions. If attention gates stimulus integration into the internal reference underlying central tendency bias, the bias should always be towards the mean of the width-attended stimuli. If attention has no effect on central tendency bias, it should be towards the mean of all widths in a given session and thus be different between the sessions. A linear mixed-effects model was used to quantify biases in width judgements from the cumulative mean of stimuli (central tendency bias) and four preceding stimuli (serial dependence) and their interactions with attention. Serial dependence in width judgments was stronger when width was attended in the previous trial compared to when hue was attended. Independently of serial dependence, we also found a central tendency bias which was toward the mean of the width-attended stimuli in both sessions. To conclude, serial dependence and central tendency bias have independent effects on judgments but both are modulated by attentional top-down processes.

This work was supported by funding from the University of Helsinki Doctoral School to SMN.

Attention: Capture 2

TUESDAY, MAY 19, 2:45 – 6:45 PM, PAVILION

56.440 WHICH DELAY IS MOST EFFECTIVE? CHARACTERIZING TEMPORAL SENSITIVITY TO ABRUPT ONSET SALIENT CUES ACROSS DELAYS AND INDIVIDUALS

Derrek T. Montalvo¹ (montalvo.46@osu.edu), Blaire Dube², Julie D. Golomb¹; ¹The Ohio State University, ²Memorial University of Newfoundland

In dynamic visual environments, abrupt salient events can influence performance, but the time period in which the visual system is maximally sensitive to such events remains unclear. A central question is whether a critical time window exists, after search onset but before a decision is made, during which abrupt onset salient events exert their strongest influence, and whether this temporal sensitivity varies based on the spatial validity of the salient event (e.g., capture by a distractor vs. facilitation by a valid cue), and/or across individuals. In the present study, we directly examined this post-stimulus, pre-decision period by manipulating the onset delay of an abrupt cue (50, 150, 350, 425, and 500 ms post search onset), that appeared around either the search target (salient target) or a non-target (salient distractor). Salient target and salient distractor cues were tested in different versions of the experiment (all participants performed both versions, order counterbalanced), and onset delays within each version were randomly intermixed with cue-absent trials. Across both versions of the task, cue timing strongly shaped performance. The earliest onset (50 ms) produced the largest effects: salient target cues generated the strongest facilitation, while salient distractor cues produced the greatest slowing. Sensitivity to the salient cue declined progressively as the delay increased, with later delays exerting a progressively weaker influence on performance. Moreover, reliable individual differences were present at the earliest delay (but not later). At 50 ms, the magnitude of the distractor driven slowing was correlated with the magnitude of target driven speeding, such that individuals who were strongly slowed by a salient distractor also showed the strongest facilitation when the cue appeared at the target location, suggesting that both early capture and facilitation may be driven by a shared salience-susceptibility mechanism and that priority maps are maximally unstable early after search onset.

NIH R01-EY025648 (JG), Discovery Grant from the Natural Sciences and Engineering Research Council of Canada

56.441 THE INFLUENCE OF MOTIVATION ON ATTENTIONAL CAPTURE AND FEATURE PERCEPTION

Attention capture episodes represent dynamic attentional states during which irrelevant feature information can interfere with and distort the perception of visual search targets. Are such distortions automatic, or can individuals exert control over the extent of the distortions? We approach this question by using performance-contingent reward to manipulate the degree to which individuals are motivated to accurately perform a continuous feature report task. Participants were presented with an array of four colored squares for 50 ms and asked to report the color of a target square marked by a thick white border. On two-thirds of the trials, a pre-cue (four white dots surrounding one placeholder) was presented 150 ms before the color array. When the pre-cue was present, it could highlight either the upcoming target location (valid condition) or a non-target location (invalid condition). At the end of the trial, continuous report color wheels were presented in each of the four display locations, and participants were asked to report both target location and color by clicking on the wheel at the target location. Critically, we manipulated motivation by varying the amount of monetary reward between blocks in two levels: a low-reward baseline and a high-reward (motivation boost) condition with increased incentives for accurate responses and penalties for inaccurate responses. Probabilistic mixture modeling analysis showed that both attention cues and reward modulated feature perception. Notably, on invalid-cue trials, probability of reporting the target color was increased with higher reward. Interestingly, we also observed more feature-binding errors in the higher reward condition, where participants accurately localized the target but swapped the invalidly cued color. These results suggest a potential trade-off between the orienting of spatial attention and updating of the location's bound feature. When sufficiently motivated, individuals may be able to reduce distractor-evoked spatial distractions at the cost of unbound feature perception.

NIH R01-EY025648 (JG)

56.442 ATTENTIONAL CAPTURE BY AN ABRUPT ONSET IMPAIRS TARGET PERCEPTION UNDER SERIAL SEARCH, BUT NOT UNDER PARALLEL SEARCH

Seonbeom Heo (tjsqja1011@gmail.com), Jonga Kwon, Suk Won Han, Sohyun Kim; ¹Department of psychology, Chungnam National University

Abrupt onsets are known to capture attention automatically, but the processing stage at which they interfere with task performance remains unclear. This study examined whether interference from task-irrelevant abrupt onsets arises at perceptual or post-perceptual stages, and whether this depends on search mode. Each participant performed one of two visual search tasks: an orientation search (a tilted line among vertical lines), which tends to proceed in a parallel manner, and a T/L search (a rotated T among rotated Ls with offset junctions), which requires serial shifting of attention among the search

stimuli. In both tasks, the search display was briefly presented (75 ms), straining the perceptual processing of search stimuli. Importantly, in half of all trials, a green outlined square was presented at a non-target location prior to the search display, while in the other half, no such onset distractor was presented. Under this perceptually challenging condition, changes in response accuracy and speed were measured as indicators of perceptual and non-perceptual modulations, respectively. As results, In the parallel search task, the presentation of an abrupt onset significantly increased RT ($p < .001$), but did not affect accuracy ($p = .826$), indicating post-perceptual interference. By contrast, in the serial search task, the abrupt onset significantly reduced accuracy (in baseline around 80%) and increased RT ($p < .001$), indicating disruption at the perceptual stage. These findings demonstrate that the locus of attentional capture by abrupt onsets depends on task demands; interference is confined to response-level processes during parallel search, but extends to perceptual encoding during serial search. This task-dependent dissociation helps reconcile mixed results in the literature and underscores the importance of search mode in shaping the impact of exogenous attention.

This research was supported by the National Research Foundation grant (NRF) funded by the Korea government (No. NRF-2024S1A5A2A01019821)

56.443 WHERE ATTENTION IS CAPTURED: LOCALIZING THE PROCESSING LOCUS OF INTERFERENCE FROM SALIENT SINGLETON DISTRACTORS.

Jonga Kwon¹ (seedsproutkwon@gmail.com), Suk Won Han¹, Seonbeom Heo¹, Sohyun Kim¹; ¹Chungnam National University

Stimuli that perceptually stand out from their surroundings often capture attention involuntarily, potentially disrupting goal-directed visual search. The present study examined the processing locus of this interference caused by a salient singleton distractor, focusing on whether the disruption occurs at an early perceptual stage or during post-perceptual processes such as decision-making and response selection. To dissociate these stages, we measured both accuracy and reaction time (RT): perceptual interference should impair both measures, whereas non-perceptual interference should selectively affect RT. Participants performed two types of visual search tasks—a line-orientation search thought to rely on parallel processing, and a rotated T/L search that requires serial attentional shifts. On half of the trials, a salient color singleton distractor was presented. In the initial experiments, singleton distractors increased RT only in the line-orientation task under longer display durations ($p < .05$), with no accuracy costs in either task ($ps > .37$), suggesting primarily non-perceptual interference. However, because these RT effects were obtained under perceptually demanding conditions, they are also consistent with the possibility that singleton distractors delay the accrual of sensory information rather than leaving perception untouched. To directly test this, we conducted additional experiments in which a local mask was presented at the target location immediately after the search display, interrupting ongoing sensory encoding. Under

these masked conditions, singleton distractors significantly reduced both accuracy ($ps < .001$) and RT ($ps < .01$). Across experiments, the presence of a singleton distractor impaired accuracy only when sensory encoding was truncated by masking, indicating that attentional capture delays the build-up of perceptual information. These findings suggest that distractor-driven interference originates at the perceptual stage, but becomes observable in accuracy only when the encoding interval is limited.

This research was supported by the Chungnam National University research grant and the National Research Foundation grant (NRF) funded by the Korea government (No. NRF-2024S1A5A2A01019821)

56.444 SEEING AN EVENT ISN'T ENOUGH: PHENOMENAL CAUSALITY MODULATES ATTENTIONAL CAPTURE BUT NOT AUTOMATICALLY

Wanna He¹, Hui Chen¹, Mowei Shen¹; ¹*Department of Psychology and Behavioral Sciences, Zhejiang University*

Phenomenal causality is perceived solely from physical cues, reflecting the perceived necessity between two motions. It is perceived automatically and independently of higher-level reasoning. Some researchers suggest that phenomenal causality may function like a Gestalt cue, integrating two motions into a unified whole in a nature way, but this hypothesis has not been empirically tested. Prior work showing that Gestalt-based perceptual integration carries over into visual working memory (VWM) representation provides a basis for examining whether phenomenal causality yields a similar integrative effect. Across three experiments, we investigated whether phenomenal causality shapes how VWM interacts with attention. Experiment 1 used a classical VWM-guided attention paradigm and manipulated causality through time interval. Participants memorized two disc colors and their causal relation before completing a visual search and a memory tasks. VWM-guided attention emerged only in the causality condition. Experiment 2 replicated this results by manipulating motion-path, again demonstrating that causally linked motions were represented as a single event in VWM that guided attention, whereas non-causal motions were not. Experiment 3 tested whether such integration occurs automatically. Participants memorized only one disc's color (i.e. target color) and completed a search task without intentionally encoding the discs' relation. If integration were automatic, both colors would be encoded and guide attention in the causality condition. However, results revealed only the target color guided attention, indicating that VWM does not automatically encode the non-target color even when it is perceptually organized by causality. Together, these findings show that phenomenal causality can restructure how multiple motions are represented and used to guide attention, but only when explicitly encoded. This work highlight event structure as a critical determinant of how VWM shapes attentional selection offering new insight into how dynamic event structure shapes cognition.

56.445 DISTRACTION UNDER HIGH PERCEPTUAL LOAD: ATTENTIONAL CAPTURE OCCURS EVEN WITHOUT BEHAVIORAL COSTS.

Greta Manini¹ (greta.manini@univr.it), Francisco Garre-Frutos², Fabiano Botta², Elisa Martín-Arévalo², Juan Lupiáñez²; ¹*University of Verona*, ²*University of Granada*

A central claim of the perceptual load theory (Forster & Lavie, 2008) is that high perceptual load prevents the processing of irrelevant distractors. In our recent work (Manini et al., 2025), we replicated the classic finding that distractors do not impair performance under high-load. However, we also observed an unexpected pattern: reaction times (RTs) were significantly faster on distractor-present than distractor-absent trials. Given that, in this task, the distractor is positioned outside the relevant search array, the absence of distractor processing should lead to no effect, rather than facilitation. Furthermore, these observations were accompanied by increased error rates, leading us to hypothesize that this reversed effect reflects spatially controlled attentional capture. That is, the distractor may act as an alerting stimulus that speeds responses, while attentional shifting toward its location is controlled early, since the distractor never corresponds to the target. To test this hypothesis, we first selected all the studies employing the same Forster and Lavie (2008) task, meta-analyzing the effect of the distractor presence. This revealed a significant and robust facilitation in high-load across the perceptual load literature. We then re-analyzed a subset of studies with open-access datasets available to test whether the target-distractor spatial proximity could be modulating search performance. This revealed that speed-ups were primarily driven by distractors appearing near the target, whereas distractors far from the target were associated with higher error rates. Overall, our findings demonstrate that distractors bias attention even in the absence of clear performance costs. This challenges the perceptual load theory's claim that irrelevant distractors are not processed under high load, and supports our proposal that facilitation effects may instead reflect a form of controlled attentional capture.

56.446 EFFECTS OF INTER-TRIAL PRIMING ON CONTINGENT ATTENTIONAL CAPTURE AND SET-SPECIFIC CAPTURE

Katherine S. Moore¹ (moorek@arcadia.edu), Victoria Fedorko², Gabriel Conn¹, Zoe Kempf¹, Zephyr Markley¹; ¹*Arcadia University*, ²*University of Chicago*

When conducting a visual search (e.g., for green letters), distractors that share features with targets (e.g., green digits) automatically capture attention, a phenomenon known as contingent attentional capture. A related phenomenon, set-specific capture, occurs when people search for more than one target type simultaneously (e.g., green letters and orange letters). Set-specific capture costs are greater than contingent capture costs when a distractor (e.g., a green digit) matches one attentional set, but the target on that trial is from a

different attentional set (e.g., an orange letter). Target identification is also affected by selection history. In particular, inter-trial priming refers to the phenomenon in which the presence of a target or a distractor sharing target features on the previous trial affects performance on the current trial. Attentional selection is superior when targets and distractors have similar characteristics from one trial to the next. Here, we investigate the effects of inter-trial priming on both forms of attentional capture. In five experiments, participants searched for green and orange letters appearing in a heterogeneously colored rapid serial visual presentation (RSVP). Distractors matching the target on that trial or the other target color appeared before the targets in the RSVP, at varying time lags. Trials containing both targets and distractors were always preceded by a trial with only a target or only a distractor item. We replicated both kinds of capture. Moreover, we found that performance was improved by a target-color matching prime on the previous trial, regardless of whether the prime was a distractor or a target. These primes had a profound effect on performance, even though prior studies have shown capture effects to be shorter-lived. These findings suggest that the timing target set enhancement from attentional capture may last longer than previously thought.

56.447 THE INCIDENCE OF VISUAL ALERTING IS NOT BLACK AND WHITE.

Rachel J. Yapp¹ (rachel.yapp@sfu.ca), Nadja Jankovic¹, Aaron A. N. Richardson¹, Thomas M. Spalek¹; ¹Simon Fraser University

Alerting is the facilitation of a response to a target when it is preceded by a brief non-informative cue. Although auditory tones are often used as the cue in the literature, a brief luminance increase has been shown to produce similar facilitation. While this visual cue reliably speeds responses, it is unknown whether both increases and decreases in display intensity produce alerting. Three experiments are presented to explore this question. In Experiment 1, a brief (33 ms) white flash preceded the presentation of a visual search display at one of four stimulus onset asynchronies (SOAs: 50, 100, 150, 200 ms). Robust alerting was observed at the three longer SOAs. Experiment 2 was a replication of Experiment 1, except that the white flash was replaced by a black flash. Alerting was observed at the 150 and 200 ms SOAs, but not at 100 ms. Although this finding is consistent with the idea that luminance offset-transients are processed slower than luminance onset-transients, another possibility is that the alerting was not due to the offset-transient but rather was due to the onset-transient produced as the black screen returned to the grey background luminance 33 ms later. Experiment 3 tested these two possibilities by extending the black-flash duration to 83 ms while keeping everything else constant. If alerting was in response to the offset-transient, then that occurs at the same point as in Experiment 2, and so the same results should be observed. If, however, the alerting is due to the onset-transient, then because this transient occurs 50 ms later in Experiment 3 than in Experiment 2, we would expect to see a further shift of 50 ms in the pattern of results. The results revealed alerting only at 200 ms, which

is consistent with the notion that only onset-transients, not offset-transients, produce alerting.

56.448 IMPACT OF TARGET- AND DISTRACTOR-DEFINING STIMULUS FEATURES ON DISTRACTOR-INDUCED ATTENTIONAL CAPTURE AT TASK-IRRELEVANT LOCATIONS

Sarah E. Pumell¹ (sarahepumell@ucsb.edu), Daniel D. Thayer¹, Thomas C. Sprague¹; ¹University of California, Santa Barbara

Foreknowledge of task-relevant features and locations guides attention during visual search, whereby information within relevant locations is strengthened while extraneous information at irrelevant locations is minimized (Kastner et al., 1999; Theeuwes, 2004). It is believed that this, in part, is due to interactions between top-down 'relevance maps' and 'feature dimension maps' in feature-selective regions encoding salient locations (Thayer & Sprague, 2023; 2025). Based on this account, we would predict that different visual features (e.g., color, motion, shape) defining task-relevant and salient stimuli, act equivalently, however, spatial attention may differentially affect interference from color- and motion-defined distractors (Leonard et al., 2015; Liu & Mance, 2011). We tested this by having participants covertly search an 8-item array for a unique target item in a prespecified feature dimension (e.g., color, motion, shape) and indicate the orientation of a line (horizontal or vertical) inside the target. Most trials were preceded by a valid, central cue indicating the task-relevant hemifield of the search array (top, left, bottom, or right; 4 of 8 items), where the target always appeared. On half of all trials, a singleton distractor in a task-irrelevant dimension (e.g., color distractor in a motion-defined search) could appear within or outside the cued hemifield. Singleton distractors were operationalized as the opposite color, motion direction, or shape (Li et al., 2019) of all other items in the array along a 360-degree continuous feature space. Cueing spatial relevance reduced distractor interference at task-irrelevant locations, but these effects differed substantially across variations in target- and distractor-defined feature dimensions. These results indicate that while top-down goals may attenuate distractor-induced capture to guide visual search, this attenuation is not uniform across all types of distraction. This motivates future studies on how the neural mechanisms mediating task relevance and salience may sculpt neural representations in feature-selective regions of the cortex.

Funding: National Eye Institute R01-EY035300

56.449 TASK STRUCTURE SHAPES BEHAVIORAL ESTIMATES OF FOCUSED AND DISTRIBUTED SPATIAL ATTENTION

Amelia Harrison¹, Thomas Sprague¹; ¹UC Santa Barbara

A large body of work reports that perceptual sensitivity declines when attention is distributed across multiple spatial locations compared to when it is focused on a single location (e.g., Ling & Carrasco, 2006; Popovkina et al., 2021). Neural responses often show similar

attenuation of stimulus representations under distributed attention (McMains & Somers, 2005; Toffanin et al., 2009). However, these effects are not universal: performance costs can be minimized by changes in decision demands (Harrison et al., 2023), and recent findings demonstrate that distributed spatial attention does not always reduce neural responses (Harrison et al., VSS 2024; Harrison & Sprague, SfN 2025). These observations suggest that measured 'capacity limits' may arise from task-specific factors rather than fixed constraints in attentional modulations of sensory processing. To characterize how task structure shapes behavioral impacts of focused versus distributed attention, we developed a framework in which stimulus input remains constant while attentional and decisional demands vary. On each trial, observers viewed two iso-eccentric stimuli and were cued to attend to one or both peripheral locations. Across task variants, participants either monitored for a target and immediately reported its identity at the pre-cued location(s), or used a post-cue after a brief delay to determine which stimulus should be reported, requiring brief maintenance in visual working memory. These manipulations dissociate sensory selection, target monitoring, working-memory storage, and decision readout while preserving visual stimulation across conditions. Across tasks, aspects of behavioral performance (RT, discriminability) depended on the number of attended locations, though the magnitude of these differences varied substantially with task factors. These results indicate that behavioral limits in spatial attention are influenced by specific features such as cueing scheme and report requirements. By isolating these components within a common stimulus framework, this approach identifies when distributed-attention costs reflect sensory selection, monitoring demands, or post-perceptual decision processes.

Research supported by a University of California, Santa Barbara Academic Senate Research Grant and an Alfred P Sloan Research Fellowship.

56.450 RAPID INVERSION OF SINGLETON DISTRACTOR REPRESENTATIONS UNDERLIES LEARNED ATTENTIONAL SUPPRESSION

Ziyao Zhang¹ (ziyaopsy@gmail.com), Jarrod Lewis-Peacock¹; ¹The University of Texas at Austin

In visually complex and dynamically changing environments, humans must often filter out salient but task-irrelevant stimuli. Prior work shows that with repeated exposure to search arrays containing color singleton distractors, individuals can learn to divert attention away from these salient items even before they capture attention. However, the neural mechanisms supporting such attentional suppression remain unclear. The present study examined the temporal trajectories of singleton distractor representations during visual search to address this gap. Using multivariate pattern analyses of EEG data (N = 40), we identified two clusters of decodable singleton distractor representations: an early cluster from 100-200 ms and a later cluster from 200-400 ms. Temporal generalization analyses showed that the later representations were transformed versions of early ones. Training

classifiers on early activity yielded negative decoding evidence when tested on later activity, indicating an inversion of representational format. Importantly, stronger late, but not early representations, predicted faster search responses, suggesting that the later signals support distractor suppression rather than attentional capture. We hypothesized that this transformation facilitates suppressing singleton distractors in the spatial priority map. To test this, we trained decoders on singleton-absent trials, where target locations are enhanced, and applied these decoders to singleton-present trials. Comparing decoding evidence across locations revealed that target locations were consistently enhanced relative to non-singleton distractors, whereas singleton distractor locations were suppressed. Moreover, comparing the neural coding of locations revealed that the spatial organization in the singleton distractor neural space was inverted relative to that in the target neural space. Together, these findings reveal a rapid representational transformation underlying salient distractor suppression at the onset of visual search. This rapid inversion of singleton distractor signals was likely driven by top-down control mechanisms that transform bottom-up saliency signals, producing an inverted arrangement of target and distractor information within a shared parietal-occipital neural space.

Attention: Temporal

TUESDAY, MAY 19, 2:45 – 6:45 PM, PAVILION

56.451 MODULATIONS OF OCULAR DRIFT DURING A COVERT ATTENTION TASK

Riccardo Brandolani¹, Claudio Galletti², Patrizia Fattori^{2,3}, Rossella Breveglieri^{2,3}, Martina Poletti¹; ¹Department of Brain & Cognitive Sciences, University of Rochester, ²Department of Biomedical and Neuromotor Sciences, University of Bologna, ³Interdepartmental center for industrial research-Aerospace, University of Bologna

It is well established that during gaze fixation, microsaccade directionality is influenced by covert shifts of spatial attention. However, the eye still moves between microsaccades due to ocular drift. Although drift is a slow movement resembling a random walk, recent research has highlighted not only its important contribution to fine spatial vision, but also the fact that the oculomotor system can exert some degree of control over this behavior. Here we examined whether drift directionality is impacted by covert shifts of attention and the temporal dynamics of ocular drift control in response to attentional cues. Participants (n = 7) performed a spatial cueing task while their right eye movements were recorded with a high-resolution Dual Purkinje Image eye tracker. Subjects fixated a central square while two colored circles appeared 5° to the left and right (Encoding window, 500 ms). Then the peripheral circles turned gray, and the central square changed into one of the two circles' color cueing covert attention to one of the two sides (Direction window, 700–1000 ms). A low-contrast target then appeared in one of the circles (80% valid trials), and participants released a button upon its detection. Trials without

microsaccades were analyzed (≈ 571 per participant). Reaction times to button release in these trials were modulated by cue validity confirming that microsaccades are not essential for attentional shifts. Trials were classified as toward, away, or other based on drift direction relative to the cue in the Directional cue window. Whereas we previously showed that microsaccades were biased toward the cue 240–340 ms post cue-onset, here we report that also drifts showed a bias but later in time around 470–550 ms. These findings show that ocular drift is modulated by covert attention, although with a slower dynamic compared to microsaccades.

This work was funded by NIH Grant EY029788 to Martina Poletti and NIH grant EY001319 to the Center for Visual Science. EU Marie Skłodowska-Curie grant agreement No 101086206 - PLACES.

56.452 CONCURRENT PHYSICAL EFFORT FACILITATES TEMPORAL ATTENTION

Li Yang¹, Weiwei Zhang; ¹University of California, Riverside

Attention fluctuates over time, creating moments when critical information is missed, as in the attentional blink (AB). Although physical exertion is common in daily life, little is known about how bodily effort shapes temporal attention. This study therefore examined how concurrent physical exertion influences visual temporal attention. Participants performed an AB task, reporting two targets (T1 and T2) embedded in a rapid serial visual presentation while exerting varied levels of isometric handgrip force. Performance on T2 given correct T1 response (T2|T1) was lower at short intertarget lags (lags 2–4) than at other lags, replicating the AB effect. The Arousal hypothesis proposes that concurrent physical effort can induce arousal, subsequently facilitate perceptual and working memory encoding of the targets, and attenuate AB. Alternatively, the Competition hypothesis proposes that the competition for cognitive control between the concurrent physical exertion and attention re-engagement process of the AB task may amplify AB. Results showed that T2|T1 accuracy in the blink window improved as physical effort increased, with no significant changes at other lags, supporting the arousal hypothesis that physical effort exertion selectively facilitates target processing in temporal attention task.

56.453 MECHANISMS OF TEMPORAL INDIVIDUATION AND AVERAGING: THE ROLE OF PRIOR KNOWLEDGE

Roy Ramati¹, Amit Yashar¹, Yaffa Yeshurun¹; ¹University of Haifa

Previous studies suggest that temporal individuation and temporal averaging are mediated by distinct processes and that individuation over long timescales is shaped by temporal interference—temporal crowding—which differs qualitatively from visual masking. To further elucidate the underlying mechanisms, we examined how prior knowledge modulates temporal individuation and averaging with central presentation. Participants ($n=36$) viewed three-items sequences with varying SOAs (200–400ms). Two conditions required individuation—reproducing the orientation of a single item. In one

condition, the target position was known in advance (always the second item). In the other, the target position (first, second, or third) was cued only after the sequence. Thus, in the former condition only a single item was ever relevant, whereas in the latter condition, all items had to be maintained in working memory until the cue appeared. In a third condition, all items were task-relevant because participants reported the items' average orientation. Mixture-modeling analysis of individuation errors, revealed the hallmark of temporal crowding: regardless of prior knowledge, SOA affected the precision of target encoding. A robust temporal-position effect emerged—precision increased for items appearing later in the sequence. Critically, prior knowledge eliminated this effect: when the target position was known, the second item's precision matched the third item's precision in the no-prior-knowledge condition. Prior knowledge also reduced substitution errors, with the magnitude of the reduction depending on SOA and on whether the distractor preceded or followed the target. In the averaging condition, unlike individuation, SOA had no effect, supporting the dissociation between averaging and individuation. This pattern resembles our findings in peripheral temporal averaging. However, unlike in peripheral averaging, the third item was assigned the highest weight. Together, these results suggest that advance knowledge can dynamically narrow the temporal window of interference, possibly by advance filtering of information based on its temporal position.

ISF, grant number 1780/19 to Yaffa Yeshurun

56.454 AN ELECTROPHYSIOLOGICAL STUDY OF RHYTHMIC ATTENTIONAL SAMPLING DURING TEMPORAL ORIENTING

Travis Powell¹ (tgppowell@ucdavis.edu), Megan Wong¹, Atticus Mackey¹, Mingzhou Ding², George R. Mangun¹; ¹University of California, Davis, ²University of Florida, Gainesville

Task performance fluctuates in the theta-band (4–8 Hz) when the deployment of visuospatial attention is sampled across a variable range of cue-target intervals. This behavioral periodicity bears evidence that attentional sampling of external stimuli is discrete. A conventional model of this phenomenon implicates the dorsal attention network (DAN) utilizing a common theta rhythm to coordinate oculomotor potentiation and visual sensitivity in an antiphasic relationship. This facilitates an orderly separation of shifting and sensing when scanning visual space. Recent evidence is emerging that extends these rhythmic behavioral findings to other domains, pointing toward an expanded model of rhythmic cognition. We contribute to this effort with a modified Posner paradigm, cueing participants to two temporally distinct variable ranges of target onsets, while acquiring scalp electroencephalography (EEG). Temporal attention recruits a cortical network distinguished from the DAN — notably by its exclusion of the FEF and reliance on the left intraparietal sulcus (LIPS) — thereby rendering it a useful candidate to extend the study of rhythmic attention beyond visual space. Mean reaction time oscillated predominantly in the theta band. Cue condition was

decodable via support vector machines (SVM) trained on RT time series data, indicating a systematic effect of temporal cue on behavioral rhythms and suggesting temporal orienting modulates attentional rhythms. Topographical distribution of EEG, consistent with extant literature, provided confirmatory evidence that the temporal orienting network was recruited for the task. The relationship between behavioral rhythms and theta (4–8) and alpha (9–13 Hz) EEG bands in relevant electrode sites suggested RT oscillations covaried with activity in the temporal orienting network. These results indicate that temporal attention also utilizes oscillations to coordinate functions that share a common neural resource.

56.455 DYNAMIC PROFILES OF TEMPORAL ATTENTION TO WORKING-MEMORY CONTENTS

Irene Echeverria-Altuna^{1,3}, Jewel Cox¹, Kia Nobre^{1,2}; ¹Department of Psychology, Yale University, ²Wu Tsai Institute, Yale University, ³Basque Center on Cognition, Brain and Language (BCBL; Spain)

Attention can be directed to sensory events (external attention) and to memory contents (internal attention). Predictable temporal structures guide external attention dynamically and proactively to anticipated onsets of relevant sensory events with good temporal resolution. Temporal expectations also modulate access to working-memory (WM) representations, but much less is known about the dynamic profile of these internal attention functions. We investigated how we use time to select and shift between WM contents. Participants encoded the angles of two tilted gratings into WM. Gratings appeared within separate circular, colored placeholders on the left and right of the screen. At the end of the trial, participants reproduced the tilt of one of the WM items. Which item participants reported was predicted by the combination of a retrocue and the subsequent delay. A colored retrocue during the WM delay predicted that participants would be probed on the item matching the cue color after a short (1 sec) delay or on the other item after a long (2.5 sec) delay. Using variations of this task, we investigated the time course of the behavioural advantages conferred by internal temporal attention. In Experiment 1 (n=30), the cue validly predicted the timing of the item to be reported on most trials (80%), but, on other trials, an item could be probed along several intervals spanning (0.5-2.7 s after cue offset). Responses at validly and invalidly cued times revealed a dynamic profile of attentional benefit, with progressive response speeding toward the expected timing of the probe. Interestingly, accuracy-related benefits were more pronounced for items that were expected to be probed early. Ongoing studies are investigating whether temporal orienting in WM yields behavioral costs in addition to benefits. The findings provide insights into how temporal regularities shape the dynamics of selecting WM contents in the service of adaptive behavior.

56.456 THE POST-TARGET DIP: DETECTING TARGETS IN A CONTINUOUS STREAM BOOSTS MEMORY FOR TARGET-PAIRED IMAGES BUT IMPAIRS MEMORY FOR THE NEXT IMAGE

Teresa P. Pham¹ (pham0738@umn.edu), Vanessa G. Lee¹; ¹University of Minnesota

Visual attention is rarely constant over time. In the attentional boost effect (ABE), participants encode background images while monitoring a sequence of colors for target colors. Memory is better for target-paired than distractor-paired images, reflecting a transient temporal orienting response. But is this response confined to the target moment, or does it facilitate (akin to the “lag-1 sparing” effect) or impair the next image, relative to later temporal positions? In four experiments, participants encoded a continuous stream of objects (800ms/item) for a later memory test while pressing the spacebar for a pre-specified target tone (Experiments 1 and 2) or a pre-specified target colored square (Experiments 3 and 4) in a separate stream. We examined five nontarget temporal positions after the target, followed by 1-4 additional trials to disrupt target regularity, and varied the nature of nontarget stimuli to create either a continuous dual-task (nontargets were distractor tones or colors, “distractors” task) or an occasional dual-task context (nontarget trials were blank, “no distractors” task). We found an attentional boost effect—better memory for target-paired than nontarget-paired images. The ABE was attenuated when nontargets were blank rather than distractors. Importantly, memory varied systematically for the five nontarget positions immediately following the target: memory was worst at the T+1 position and gradually improved at later positions, revealing a post-target dip followed by gradual recovery. These findings suggest that detecting and responding to behaviorally relevant stimuli produces a temporal ripple, facilitating target-paired images while pulling resources away from the next image. The gradual recovery at later positions may reflect a shift of resources to the background images. These results uncover a new temporal dynamic in continuous tasks and have implications for everyday multitasking, such as driving or receiving occasional text messages.

McKnight Foundation

56.457 CONCEPTUAL RELATEDNESS FAILS AS A SEARCH TEMPLATE UNDER HIGH-SPEED PRESENTATION

Moussa Kousa¹ (mak6934@psu.edu), Brad Wyble¹; ¹The Pennsylvania State University

A picture flashed for only a few milliseconds can still be clearly perceived if nothing follows it. However, in a rapid serial visual presentation (RSVP) of photographs, each picture masks the one before it, yet even under such rapid masking, people can still detect a briefly presented target scene when given only a conceptual cue, such as “picnic” (Potter 1976). To explore the limits of this selection, we tested whether searching for “relatedness” allows participants to detect a pair of semantically related objects (e.g., a golf club and a golf ball; depicted as line drawings) embedded within an RSVP of unrelated pairs (e.g., a shoe and a plant), as assessed by a memory probe. We also tested the reverse: detecting an unrelated pair within

a stream of related pairs. These tasks were assessed with four RSVP SOAs: 800, 400, 200, and 106 ms. At longer SOAs, performance in both conditions was near ceiling. As SOA decreased, accuracy declined drastically for both conditions, although detecting related pairs was slightly easier than detecting unrelated pairs. This poor performance at short SOAs contrasts with excellent performance for using conceptual cues with natural images at SOAs of 100ms in prior work, suggesting that “relatedness” is not an effective cue for selecting targets in the same way that a broad scene label is. To verify that subjects were able to connect our stimuli to conceptual labels even during high-speed RSVP, we conducted a variant of the experiment in which participants were given a conceptual label (e.g., chopping for axe–tree) instead of searching for related or unrelated pairs. At SOAs of 400, 200, and 106 ms, performance at the lower SOAs was much higher than in the first experiment. These results suggest a specific inability to select based on relatedness as a general property.

56.458 EXPERIENCE DEPENDENT MODULATION OF AROUSAL DURING DYNAMIC GAME PERFORMANCE

Jordan Kulwicki¹, Shawn Akridge², Matthew Peterson²; ¹George Mason University

Understanding how visual attention and arousal scale with task demands is central to models of perceptual expertise. Rhythm action video games provide a naturalistic yet controlled environment for testing how visual strategies adapt to increasing workload. Here we present a secondary analysis of an existing Clone Hero dataset, focusing on how experience level modulates confidence, workload, performance, and pupillary arousal under varying speed demands. In the parent study, forty-five participants played repeated trials at base speed (9) and fast speed (18) while eye movements and pupil size were recorded with an EyeLink 1000. Participants spanned novice, moderate, and expert experience levels. The dataset includes visual attention metrics, NASA-TLX workload subscales, confidence ratings, performance accuracy, and pupil area. We analyzed these measures using ANOVAs and linear mixed-effects models, with an emphasis on expertise and speed related changes in arousal. Replicating core findings, experts showed reduced fixation dispersion, higher accuracy ($p < .001$), and higher confidence ($p < .001$) than less experienced players. Speed increased subjective workload across groups, but workload did not differ by experience. Critically, overall pupil size did not differ across experience levels; instead, a significant Speed x Experience interaction emerged, $F(2, 222) = 3.36$, $p = .037$. Mixed effects modeling revealed that only experts exhibited reliable increases in pupil dilation with faster speeds ($\beta = 32.56$, $p = .016$), indicating adaptive, context sensitive arousal modulation rather than globally elevated arousal. These results suggest that expertise is characterized not only by more efficient visual attention strategies, but also by the ability to selectively upregulate arousal when task demands increase, supporting accounts that treat arousal as a learned control signal for tuning visual processing efficiency under uncertainty.

56.459 AGING AND VISUAL ATTENTION: MINIMAL IMPACT OR MAJOR DECLINE?

Alex Holcombe¹, Viplav Tuladhar¹, Loretta Duffy¹, Styliani Katsoulis¹, Joshua Pham¹, Vince Tafea¹, Rachel Wong¹, Sarah Cronje¹, Yuenchen Lee¹; ¹The University of Sydney

Based on possibly the largest published study of visual attention in older adults, Verissimo et al. (2022) concluded that attentional orienting and executive inhibitory efficiency improved with age until people are in their mid-to-late 70s. Re-analysing their data, we reveal that the orienting improvement finding is based on a coding error. We also question the validity of their inhibitory efficiency and orienting measures. To address the orienting or attentional shifting issue ourselves with a new empirical study, we conducted a multiple object tracking study based on that of Roudaia & Faubert (2017). In that paradigm, with objects widely spaced to avoid spatial interference, speed thresholds are limited by temporal interference between distractors and targets (Holcombe & Chen, 2013; Roudaia & Faubert, 2017). After implementing changes to avoid spatial interference and ensure the older adults understood the task, we replicated the finding of Roudaia & Faubert (2017) that older adults' multiple object tracking temporal thresholds are substantially worse than that of undergraduates, despite them performing similarly on a matrix reasoning task. These findings further undermine the conclusions of Verissimo et al. (2022) and suggest that older adults have particular difficulty with avoiding long-lasting activation of distracting stimuli.

Attention: Reward

TUESDAY, MAY 19, 2:45 – 6:45 PM, PAVILION

56.460 SPATIAL GRADIENTS OF REWARD AND THREAT INTERACT TO GUIDE HUMAN FORAGING PATTERNS

EBRU ECEM TAVACIOGLU¹ (ebrueecem@gmail.com), MATTHIAS GAMER¹, JEREMY WOLFE^{2,3}; ¹University of Würzburg, Germany, ²Brigham and Women's Hospital, ³Harvard Medical School

Human foraging in natural environments requires continuously allocating attention across space while weighing potential gains against potential costs. Although visual foraging paradigms have revealed much about how observers guide search across extended scenes, far less is known about how people decide where to forage, given spatially varying rewards and threats having a different spatial layout. Here, we examined how distributed value and aversive sound jointly shaped attentional allocation and harvesting behavior during continuous visual foraging. Participants completed a grid-based foraging paradigm in which reward magnitude (1–5 points) and auditory threat level (0–4 intensity) followed orthogonal gradients across a 20×20 display. In a baseline block, rewards were uniform and no sound was presented; in a gradient block, both reward and sound values were spatially structured. Participants freely harvested targets across multiple patches. Click-level behavior was extracted from all

cell responses and aggregated into 5×5 spatial bins reflecting reward and sound levels. Relative to the baseline block, the gradient block led to higher click rate, more hits and false alarms (FA), reduced time per patch, and decreased positive predictive value (PPV) for each click. Critically, spatial heatmaps revealed that these effects were not uniformly distributed. Performance deteriorated most notably in the high-value regions, which showed a marked increase in FAs and a drop in PPV, whereas behavior in low-value/low-sound regions remained comparatively stable. These results visualize how distributed value and threat can jointly shape harvesting decisions, influencing both overall click behavior and the spatial distribution of hits and errors. Since behavioral adjustments were not uniform across the grid, we can reject the hypothesis that threat simply reduces overall performance. Together, these findings demonstrate that aversive sound impairs the efficient exploitation of high-value areas, reducing the efficiency of harvesting.

56.461 ATTENTIONAL CAPTURE BY SIGNALS OF REWARD INFLUENCES VALUE-BASED DECISION-MAKING

Daniel Pearson¹ (daniel.pearson@sydney.edu.au), Kirralee Harris¹; ¹The University of Sydney

Pairing a stimulus with reward produces a rapid, automatic attentional bias towards that stimulus that persists even when counterproductive – a phenomenon known as value-modulated attentional capture (VMAC). Beyond influencing performance in visual search tasks, individual differences in VMAC have been shown to predict real-world behaviour, particularly reward-related behaviour related to addiction and compulsivity. However, the mechanism that links VMAC and behaviour remains unclear. Evidence from the decision-making literature suggests that individuals are more likely to choose options that receive more attention during decision-making, though this work has focused primarily on goal-directed attention. The present experiments test whether automatic, reward-related attentional biases similarly influence value-based choices. Across two experiments, participants completed a visual search task where different coloured distractors predicted either high or low-value rewards. Then, they made choices between similarly-rated snack foods randomly paired with either the high- or low-value colour. Experiment 1 used a blocked design, which revealed a small choice bias towards the snack food paired with the high-value colour, but this bias did not persist when excluding participants who believed that the search and choice task were related. Experiment 2 used an intermixed design to reduce the influence of task context, and demonstrated a robust choice bias for snacks paired with the high-value colour, which persisted when excluding participants who believed that there was a relationship between the tasks. Moreover, individual differences in the magnitude of the VMAC effect predicted the magnitude of the choice bias, such that those who were more susceptible to having their attention captured by stimuli associated with high-value rewards were also more likely to choose options that were paired with those stimuli. These

findings suggest that reward-related attentional biases can influence value-based decision-making, providing a potential mechanism for the link between VMAC and real-world behaviour.

56.462 EMOTIONAL CLIMATE IMAGES CAPTURE ATTENTION

Erin Wallace¹, Zephyr Markley¹, Gabriel Conn¹, Katherine Moore¹; ¹Arcadia University

Emotional stimuli tend to capture attention and are prioritized for visual processing ahead of neutral stimuli. Marketing, social, and political campaigns may use this principle to capture users' attention on social media platforms, which are highly visual. Laboratory studies show this phenomenon using the attentional blink (AB) task—an emotional first target (T1) will cause a greater impairment in identifying a subsequent target (T2) as compared to a neutral T1, and an emotionally salient T2 will be protective against an AB. Here, we examined whether emotional nature images protect against an attentional blink when presented as a T2 in an AB task, and whether these effects are related to one's views on climate change and the environment. On each trial, participants were cued with a category (e.g. "trees") and were asked to identify how many times (0, 1, or 2) this item appeared within a 12-item rapid serial visual presentation (RSVP). Critical trials (50% of total) containing two targets included a neutral T1 and T2 with positive, neutral, or negative valence; images were rated on emotionality by a separate group of participants. We observed smaller ABs for positive and negative T2s than for neutral stimuli. This effect was not significantly modulated by a participant's environmental views. These findings have implications for the use of emotionality in image-based propaganda and social movements.

56.463 THE INFLUENCE OF EMOTION ON OBJECT-BASED ATTENTION USING REWARD

Sanweda Mahagabin¹, Sarah Shomstein¹; ¹The George Washington University

Object-based attention (OBA) enhances processing of features within attended objects beyond their spatial position. While reward history is known to bias attention, it remains unclear how the emotional valence of feedback—reward, punishment, or neutral—modulates space- and object-based selection. Using performance-contingent feedback to link each object with a positive or negative emotional value, we examined how these associations modulate OBA in a modified Egly-rectangle paradigm. Participants completed a valenced session where one object was consistently rewarded and the other punished, and a separate group completed a structurally identical no-reward session with neutral feedback. Each trial contained a cue at one end of a rectangle, followed by a target at the cued location (valid), the opposite end of the same object (same-object, SO), or the equidistant location on the uncued object (different-object, DO). In the valenced session, objects carried fixed emotional value: targets on the rewarded object always yielded gains, whereas targets on the punished object always yielded losses. The no-reward session provided non-contingent,

neutral feedback. Accuracy was uniformly high ($\geq 97\%$) with no group differences, ruling out speed-accuracy tradeoffs. Space-based attention was robust: responses were faster for valid than invalid trials in all conditions, with no modulation of this effect by valence. Within the valenced session, reward produced faster responses than punishment for both trial types ($ps < .01$). OBA was also reliable: invalid SO trials were faster than DO trials across groups (SO-DO: 22.62 ms reward, 37.35 ms punishment, 33.15 ms no-reward). Importantly, a significant Valence $\times$ Trial interaction emerged ($F(2,76)=3.73$, $p=.029$), indicating slower DO responses under punishment than under reward. In a set of follow-up experiments, we tested whether similar emotional influences arise when affective content, rather than monetary reward, is embedded in background scenes or used as targets themselves.

56.464 EMOTIONAL AROUSAL MODULATES RECEPTIVE FIELD PROPERTIES IN HUMAN EARLY VISUAL CORTICES

Chenanke Sun¹, Joseph McGuire¹, Janneke Jehee², Sam Ling¹; ¹Boston University, ²Donders Institute for Brain, Cognition and Behavior

Arousal significantly shapes human behavior and sensory processing. Its effects are believed to extend to early visual processing stages, such as contrast sensitivity and spatial frequency processing. Yet, the neural mechanisms underlying these effects remain poorly understood. In this study, we used fMRI to assess how arousal, specifically emotional arousal, modulates population receptive field (pRF) properties in human visual cortex. We combined classic pRF mapping paradigm with auditory cues that varied in its emotional valence (Neutral or Negative valence), drawn from the International Affective Digitized Sounds database (IADS). pRFs were estimated using classical moving bar, wedge, and ring apertures filled with images of objects and faces, while participants were asked to maintain fixation and detect color changes of a central fixation point. During the task, participants heard concatenated negative valence (e.g. people screaming) or neutral (e.g. footsteps) auditory clips from the IADS database, or no auditory stimulus. We estimated pRFs in early visual areas (V1–V3) and compared pRF sizes across auditory conditions, as a function of eccentricity. We found substantially smaller receptive field sizes when participants heard the negative valence stimuli compared to neutral stimuli, with the smallest receptive fields observed in the no-sound condition. This modulation of pRF size was robust across early visual areas V1-V3, and showed a clear eccentricity dependence: differences between conditions were most pronounced in the periphery, with little-to-no modulation at central eccentricities. Together, these results demonstrate that emotional arousal and the presence of auditory input sharpens spatial representations in periphery in early visual cortex, offering new insight into how affective states alter early stages of human visual processing.

This work was funded by the National Institutes of Health (NIH) Grand R01EY035640 to S. Ling.

Multisensory Processing: Audiovisual

TUESDAY, MAY 19, 2:45 – 6:45 PM, PAVILION

56.465 WHAT DOES VISION SOUND LIKE? VARIABILITY IN CROSS-MODAL PERCEPTION

Naomi Lytle¹, Sami Yousif²; ¹UNC Chapel Hill, ²Ohio State University

Imagine the shape of a star or a circle. Each of these shapes looks a certain way, but they each sound a certain way, too. Specifically: There are well-documented cross-modal perceptual effects whereby, for instance, jagged shapes sound more like the word "kiki" and rounded shapes sound more like the word "bouba". This is a particularly compelling instance of sound symbolism, in which people seem to universally agree on the cross-modal associations. But how robust are such associations? If presented with a wide range of cross-modal comparisons like these, do people always agree? Here participants ($N=250$) evaluated cross-modal associations for six hundred distinct item pairs across three item types: (1) visual images (e.g., abstract art), (2) made-up words, and (3) brief audio clips of everyday sounds. On each trial, we showed participants a pair of items (e.g., an image and a made-up word) and simply asked them to rate how well the two items fit together. For all item-type combinations, judgments of fit were reliable ($ps < .001$). Then, we asked whether certain participants were systematically typical/atypical compared to the rest of the population. Again, we found systematic variability ($ps < .001$). This systematicity was consistent across the various cross-modal conditions: We found that participants who were typical for one type of association (e.g., image/word) tended to also be typical for others (e.g., sound/word). Finally, we investigated whether people were aware of their level of typicality/atypicality in these judgments by asking participants to guess how other people would rate that pair. Participants' guesses of others' judgments were no more predictive of the aggregated responses. Collectively, we show that while individuals systematically vary in their 'typicality' for cross-modal associations, they seem generally unaware of how their cross-modal 'taste' compares to others'.

56.466 LOWER SUSCEPTIBILITY TO THE MCGURK ILLUSION IN MISOPHONIA

Ghazaleh Mahzouni¹, Nicolas Davidenko; ¹University of California, Santa Cruz

The McGurk effect is a multisensory illusion where hearing one sound (/ba/) while seeing incongruent lip movements (/ga/) can produce a new percept (/da/ or /tha/). Previous work has linked individual differences in McGurk susceptibility to gaze patterns (e.g., fewer mouth-directed fixations; Gurler et al., 2015), sensory weighting (Stacey et al., 2021), and individual differences in clinical traits such as autism (Zhang et al., 2019). In Study 1, we investigated whether people with misophonia (characterized as a strong aversion to orofacial sounds like chewing and slurping; Swedo et al., 2021) differ in their susceptibility to the McGurk illusion and to a new audio-visual

conflict illusion based on the Sound-Swapped Video database, where trigger sounds (e.g. mouth swishing water) are paired with plausible alternative videos (e.g. shaking a water bottle; Samermit et al., 2022; Mahzouni et al., 2024). We recruited 29 participants with misophonia and 49 controls based on the Duke Misophonia Questionnaire. Participants completed a McGurk task followed by a sound identification task using congruent (e.g., /swishing/, /swishing/) and incongruent (sound-swapped; /swishing/, /shaking bottle/) audiovisual stimuli. We found that misophonic participants were less susceptible than controls to visual influences in the McGurk task but not in the sound identification task. Further, there was no relationship between performance in the two tasks. In Study 2, we recruited 30 new participants and monitored their eye movements as they completed a McGurk task. We replicated our behavioral findings, but found no difference in the proportion of mouth-directed fixations between misophonia and control participants and no relationship between mouth-directed fixations and McGurk susceptibility. Together, our results suggest that lower susceptibility to the McGurk illusion in misophonia is not driven by differences in mouth-directed fixations but perhaps by differences in auditory vs. visual weighting in speech stimuli.

56.467 INDIVIDUALIZED ACOUSTIC NOISE ENHANCES VISION THROUGH CROSSMODAL STOCHASTIC RESONANCE

Mustafa Ozaydin^{1,2}, Simay Uner^{3,4}, Irem Akdogan^{3,4}, Hulusi Kafaligonul^{3,4,5}; ¹Department of Cognitive Science, Middle East Technical University, Ankara, Türkiye, ²Department of Psychology, Atilim University, Ankara, Türkiye, ³Department of Neuroscience, Aysel Sabuncu Brain Research Center, Bilkent University, Ankara, Türkiye, ⁴National Magnetic Resonance Research Center (UMRAM), Bilkent University, Ankara, Türkiye, ⁵Neuroscience and Neurotechnology Center of Excellence (NÖROM), Department of Anatomy, Faculty of Medicine, Gazi University, Ankara, Türkiye

Stochastic Resonance (SR) describes a counterintuitive phenomenon in which adding external noise helps subthreshold signals to surpass the system threshold, thereby increasing perceptual performance. In the crossmodal context, acoustic white noise can specifically enhance visual perception through this mechanism. However, recent evidence indicates that SR does not emerge at a single "optimal" noise level. Instead, the effective noise intensity is highly dependent on an individual's dynamic neurophysiological state. Therefore, traditional fixed-level stimulation risks masking the SR effect (Uner et al., 2025). To address this, we employed a fully individualized protocol designed to map the idiosyncratic nature of SR. Participants first completed auditory and visual threshold assessments and were then assigned to either subthreshold or suprathreshold visual contrast groups (targeting ~65% vs. ~85% accuracy, respectively). They performed a 4-AFC visual detection task while individualized acoustic noise (dB Sensation Level) was concurrently applied, allowing psychophysical matched comparisons across observers. Results showed that the magnitude of

the SR effect was significantly greater in the subthreshold than in the suprathreshold group, when assessed at optimal acoustic noise intensities. Additionally, SR gains were inversely related to baseline accuracy, suggesting that cross-modal noise preferentially boosts observers with weaker initial visual performance—a common pattern shown in SR studies. Together, these findings demonstrate a robust crossmodal SR effect, showing that task-irrelevant auditory noise can systematically enhance visual sensitivity when tuned to individual sensory characteristics. By accounting for inter-individual variability, our results indicate that acoustic noise interacts with visual processing to facilitate perception, likely by leveraging functional interactions across auditory-visual pathways or operating at multisensory convergence stages.

This work was supported by The Scientific and Technological Research Council of Türkiye (BİDEB 2211 Program) and the BAGEP Award of the Science Academy-Türkiye.

56.468 THE DEVELOPMENT OF CROSSMODAL INTUITIVE PHYSICS AND ITS RELATION TO CROSSMODAL CORRESPONDENCES

Mahan Abdollahinarenjbon¹, Lily Charbonneau¹, Vivian Ciaramitaro¹; ¹University of Massachusetts Boston

The bouba-kiki effect is a type of crossmodal correspondence, in which pseudowords, like "bouba", tend to be matched with round shapes, and "kiki" with spiky shapes. Evidence for such avBK associations is equivocal in children under 3 (Fort et al., 2017), and even 6-8-year-olds show weaker associations than older children (Chow et al., 2019). Auditory dimensions of spectral content and temporal continuity can correctly categorize BK pseudowords, as well as categorize the sounds expected from rolling objects based on their shape (round vs. spiky; Fort & Schwartz, 2022). Everyday perception and action rely on such crossmodal intuitive physics (avIP): the ability to infer a scene's structure and predict how objects will move and what sounds will result. We hypothesized that the same cognitive mechanism allowing us to anticipate the sound a rolling ball would make based on its shape may also bias our matching of more abstract audio-visual correspondences. We tested if avIP strength correlated with avBK strength. Each participant (6-8-, 9-11-, and 18-35-year-olds; N=198 total) completed 16 trials (2shape pairs*4sounds*2repeats) of avBK and avIP, with block order counter-balanced. On each trial, two shapes appeared, round and spiky, and a sound was played. avBK presented abstract round and spiky shapes and pseudowords, and avIP presented round and spiky balls about to roll (implied visual motion) and rolling sounds. We quantified matching strength as the proportion of congruent choices. All participants showed above-chance congruent matches for avIP and avBK, with effects increasing with age. Overall, children with stronger avBK matches showed stronger avIP matches (Spearman $\rho=.22, p=.012$), an association driven by 9–11-year-olds ($\rho=.28, p=.031$). Thus, we showed that children as young as 6 years of age have a crossmodal intuitive physics capacity, which strengthens with development. This capacity may help scaffold how

abstract shapes and nonsense pseudowords are associated, an association important for word-learning.

56.469 BOTTOM-UP AND TOP-DOWN FACTORS IN THE VENTRILOQUIST EFFECT

Zion Park¹ (gkr8448@gmail.com), Chai-Youn Kim¹; ¹*School of Psychology, Korea University*

Multisensory integration depends on bottom-up sensory reliability to infer a common cause across inputs from different modalities. While the predictive coding framework suggests that top-down signals can actively guide this process, it remains unresolved how such influence modulates multisensory integration with the presence/absence or reliability of sensory inputs. To address this, we manipulated both bottom-up and top-down aspects within a ventriloquist paradigm. Participants conducted an auditory localization task while the visual stimulus was either present or absent and its contrast varied when present (high- vs. low-visibility). Top-down influence induced by an arrow cue consistently indicated the upcoming visual location after a fixed fixation interval. The ventriloquist effect, the auditory localization bias toward the visual location, was analyzed to determine the relative influences of the bottom-up and top-down factors and their potential interaction. Results showed that a stimulus of high visibility significantly induced the ventriloquist effect ($10.14^\circ \pm 2.61^\circ$), whereas the magnitude of this effect decreased with a stimulus of low visibility ($3.95^\circ \pm 4.68^\circ$). Crucially, the presentation of a spatial cue in the absence of a visible stimulus failed to induce the ventriloquist effect ($1.16^\circ \pm 1.68^\circ$). Taken together, these findings suggest that bottom-up sensory reliability modulates multisensory integration within the ventriloquist paradigm, while the presence of sensory input serves as a prerequisite that gates top-down influence. To further examine the interaction of sensory reliability and top-down influence, ongoing work investigates whether prediction can modulate multisensory integration by enhancing the effective reliability of low-visibility visual inputs.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2023R1A2C2007289).

56.470 SYNTHETIC SCENE GENERATION FOR EVALUATING VISUAL FEATURE CONTRIBUTIONS TO SEGMENTATION DECISIONS

Joshua M. Martin¹ (joshua.martin@tu-darmstadt.de), Thomas S. A. Wallis; ¹*Centre for Cognitive Science and Institute for Psychology, Technical University of Darmstadt.*, ²*Center for Mind, Brain and Behavior (CMBB), Universities of Marburg, Giessen and Darmstadt.*

Neural network models have been widely applied to study the processes underlying high-level object recognition; however, mid-level visual processes, such as segmentation and grouping, remain relatively understudied. One reason is the lack of suitable datasets: natural images are difficult to annotate manually and offer limited experimental control over the availability of visual cues. This challenge

is well suited to synthetic data generation using physically based rendering, which can produce fully customizable 3D scenes while preserving key aspects of visual realism through physically accurate light transport. Here, we present a Blender-based rendering pipeline for quantifying how different visual features contribute to segmentation decisions. The pipeline draws inspiration from two main paradigms: (1) digital embryos (complex 3D geometric shapes based on simulations of embryonic development), and (2) dead leaves (scenes created by sequentially layering shapes to reproduce key statistical properties of natural images). Together, the resulting scenes are ideal for studying segmentation processes, as they include complex 3D geometrical shapes, while approximating natural scenes through the presence of depth cues and partial occlusions. A central advantage of our approach over existing rendering pipelines is precise control over the availability of visual cues. By independently manipulating features such as material, lighting, geometry, and motion, experimenters can generate scenes that are matched in their overall configuration but differ only in whether one or more cues are informative for a segmentation decision. Furthermore, the scalable and automated nature of the pipeline can generate large, varied datasets with pixel-accurate ground-truth segmentation masks, making it an ideal testbed for training and testing neural network models. Overall, this approach combines the benefits of controlled psychophysics-style stimulus design and large-scale computer vision frameworks, offering a new method for studying and comparing mid-level visual processing in biological and artificial systems.

Funded by the European Union (ERC, SEGMENT, 101086774). Views and opinions expressed are however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

56.471 SUSTAINED AUDITORY SPATIAL ATTENTION FACILITATES VISUAL PROCESSING

Yong Min Choi¹ (yong.min.choi@dartmouth.edu), Viola Störmer¹; ¹*Dartmouth College*

Real-world behavior rarely relies on a single sensory modality. Instead, multi-sensory inputs are concurrently processed and integrated to guide complex behavioral goals. For example, orienting spatial attention to a transient, peripheral sound enhances visual processing, as reflected in improved visual discrimination performance and amplified visual-cortical responses for stimuli at the sound's location (e.g., Keefe & Störmer, 2021). Meanwhile, auditory processing often requires sustained, rather than transient, attentional engagement (e.g., listening to a speech). Therefore, in the present study, we examined whether sustained auditory spatial attention enhances visual processing at spatially congruent locations. Participants ($N = 24$) performed a dual task combining a dichotic listening paradigm with a visual discrimination task. In the auditory task, participants listened to two streams of digits presented from speakers located at the left and right side of the monitor, with each digit randomly spoken in either a

male or female voice. Before each block, they were instructed to attend to the stream from (1) the left speaker, (2) the right speaker, or (3) a specific voice (male or female) regardless of location, and to detect immediate digit repetitions (auditory 1-back) while maintaining fixation. Concurrently, brief subthreshold Gabor patches appeared intermittently in either visual hemifield, and participants reported their orientation (clockwise vs. counterclockwise). We found significantly enhanced visual discrimination accuracy when directing auditory attention to the speaker located at the same hemifield, compared to when attending to the opposite location ($p_{\text{bonf}} = .008$) or to a non-spatial voice cue ($p_{\text{bonf}} = .048$). Exploratory analyses of eye-tracking data revealed small but systematic gaze shifts toward the attended hemifield, which did not explain the observed effect of auditory spatial attention on visual task performance. These findings demonstrate that sustained auditory spatial attention selectively enhances visual processing at corresponding locations, revealing robust cross-modal attentional facilitation under prolonged attentional demands.

This study was supported by National Science Foundation under BCS-2446115.

56.472 EMOTIONAL AROUSAL ALTERS POPULATION SPATIAL FREQUENCY TUNING

Luis D. Ramirez^{1,2}, Jasmine Pan¹, Janneke F. M. Jehee³, Joseph T. McGuire¹, Sam Ling¹; ¹Boston University, ²University of California San Diego, ³Donders Institute for Brain, Cognition, & Behavior, Radboud University

Emotions guide perception, influencing threat and opportunity recognition in daily life. Because spatial frequency (SF) processing defines the scale of detail encoded by the visual system, emotional modulation of SF sensitivity could have a substantial impact on daily vision. Indeed, psychophysical work has suggested that negative emotions may augment SF processing. However, the mechanisms by which emotion affects visuocortical SF tuning remains unknown. To address this, we used fMRI to measure voxel-wise population SF tuning (pSFT) in early visual cortex (V1–V3), while participants heard sounds that contained either Negative or Neutral valence. To measure pSFT, we presented bandpass-filtered noise stimuli (0.5–12 cpd; 40 log-spaced steps; 80% contrast; 40-s blocks). During each pSFT block, participants also heard either Negative or Neutral valence daisy-chained auditory clips (IADS database), followed by a 2-s arousal rating window. In a separate baseline condition meant to map SF preferences in the absence of sounds, participants performed a silent luminance change detection task at fixation during pSFT mapping. Importantly, visual statistics were identical across blocks, and the only difference was what participants heard and the task performed. Negative emotion was validated by behavioral ratings, pupil dilation, and amygdala activation. Interestingly, we discovered that Negative sounds induced a systematic increase in preferred SF, both as a function of eccentricity, as well as a function of baseline pSFT peak. Furthermore, this shift was accompanied by increased selectivity (narrower population-level bandwidths), particularly in extrastriate

cortex. Our results suggest that negative emotion may alter how we see by shifting population tuning preferences towards a higher visual resolution, allowing us to better discern the finer details of potential threats.

This research was funded by National Institutes of Health Grant EY028163 to S. Ling and supported by F99NS124144 and K00EY036804 to L.D. Ramirez.