The Impact of Active Sensing on Visual Processing in the Mouse

by

Shelby Lynn Sharp

A dissertation accepted and approved in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in Biology

Dissertation Committee:

Dr. Santiago Jaramillo, Chair

Dr. Cris Niell, Advisor

Dr. Caitlin Fausey, Core Member

Dr. Mike Wehr, Core Member

Dr. Matt Smear, Institutional Representative

University of Oregon
Fall 2024

DISSERTATION ABSTRACT

Shelby Lynn Sharp

Doctor of Philosophy in Biology

Title: The Impact of Active Sensing on Visual Processing in the Mouse

Vision is an active process that relies on one's ability to sample the visual scene through interactions with the environment. Traditionally, visual neuroscience has been investigated through restrained experimental conditions that limit animals to head-fixation and minimal locomotion on a spherical treadmill. To fully understand visual processing it is imperative that we utilize technological advances to study vision in more ethologically relevant experimental conditions. This dissertation lays out novel experimental methods and data analysis that enable the study of visual processing in freely moving mice during behavior.

Chapter II of the dissertation explores how mice determine distance information during an ethologically inspired jumping task through eye and whole body video tracking, optogenetic shutdown of primary visual cortex (V1), and monocular occlusion. The study finds that mice use vision to accurately jump across gaps of varying distances. Additionally, mice are able to accurately perform this task with monocular occlusion suggesting that proper distance estimation is not dependent on binocular disparity or stereo vision. Further, optogenetic shut down of both monocular and binocular V1 impaired task performance indicating that integration of visual information in primary visual cortex is necessary to reliably estimate distance. Importantly this study acts as a cornerstone for future experiments investigating neural circuitry during ethologically relevant tasks.

Chapter III explores the impact of head and eye movements on visual processing in V1, focusing specifically on the role of gaze shifts during free movement. During head-fixation mice rarely make saccadic eye movements, and in fact most mouse eye movements occur during head movements. To overcome this limitation we designed an experimental paradigm to investigate gaze shifts during free movement. By combining neural data, eye and world view cameras, and

an inertial measurement unit we evaluated the impact of gaze shifting head and eye movements on visual processing. This study uncovers a dynamic temporal sequence that occurs in V1 following gaze shifts, which is visually driven and is elicited by the influx of new visual information that occurs during saccadic movements. Further, these gaze shift dynamics are consistent with the computational principle of coarse-to-fine visual processing.

Chapter IV expands upon the previous freely moving gaze shift experimental methods to investigate the impact of head and eye movements on visual processing in a different brain region involved in visual processing, the superior colliculus (SC). This study explores the head and eye movement responses in SC across depth revealing differing gaze shift dynamics between superficial SC (sSC) and deep SC (dSC). Specifically, in sSC we see gaze shift responses that, like V1, correspond to coarse-to-fine processing and are visually driven; however, in dSC gaze shift responses reflect head movement signals that persist in the absence of visual information. We believe that this work shows for the first time the impact of free moving gaze shift responses across the layers of SC.

This dissertation consists of previously published and unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Shelby L. Sharp

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon
Westminster University, Salt Lake City, Utah

DEGREES AWARDED:

Doctor of Philosophy, Biology, 2024, University of Oregon
Bachelor of Science, Neuroscience, 2018, Westminster University

AREAS OF SPECIAL INTEREST:

Systems Neuroscience
Behavioral Neuroscience
Electrophysiology

PROFESSIONAL EXPERIENCE:

Teaching Assistant, Department of Biology, University of Oregon, Eugene
2019-2020

GRANTS, AWARDS, AND HONORS:

Ruth L. Kirschstein National Research Service Award (NRSA) Individual Predoctoral Fellowship to Promote Diversity in Health-Related Research, 2023

Westminster University Outstanding Senior in Neuroscience, 2019

The Great Salt Lake Summer Research Fellowship, 2018

Milton S. Sprague Science Scholarship, Westminster University 2016, 2017, 2018

National Youth Science Camp: Idaho Delegate, 2015

PUBLICATIONS:

Parker, P. R., Abe, E. T., Beatie, N. T., Leonard, E. S., Martins, D. M., **Sharp, S. L**., ... & Niell, C. M. (2022). Distance estimation from monocular cues in an ethological visuomotor task. *Elife*, *11*, e74708.

Parker, P. R., Martins, D. M., Leonard, E. S., Casey, N. M., **Sharp, S. L**., Abe, E. T., ... & Niell, C. M. (2023). A dynamic sequence of visual processing initiated by gaze shifts. *Nature Neuroscience*, *26*(12), 2192-2202.

Sharp, S. L., Shin, J., Martins, D. M., Jones K., Niell, C. M. (2024). The impact of active vision on mouse superior colliculus. *In preparation*

ACKNOWLEDGEMENTS

My journey in science has been enriched by the support of amazing mentors. I would first like to express my gratitude to my mentor, Dr. Cris Niell. When I was first thinking of applying for graduate school I came across the University of Oregon and was immediately excited by the work in the Niell lab, and in my naivety as an undergraduate I emailed Cris and asked if he would be willing to have a phone call with me to talk about the program. In what I would learn to be true Cris fashion he took time out of his very busy schedule to answer my questions and importantly encourage me in my scientific endeavors. I look back now and realize Cris started mentoring me before I even joined his lab. His encouragement early on in my PhD process was invaluable and I soon learned his unwavering support would be an integral part to my success as a scientist. Cris' enthusiasm for science and drive push the status quo of systems neuroscience is infectious and it has been a joy to work beside him during my doctoral program.

I also want to express my thanks to Dr. Phillip Parker, who spent countless hours showing me how to do surgeries, experiments, and fielding my never ending list of PhD student questions. Phil's mentorship gave me such strong footing to become an independent scientist and his approach to visual neuroscience focused on ethological relevance has reshaped my approach to how we should study the brain.

I extend my gratitude to Dr. Krista Todd who took me under her wing and showed me for the first time what it's like to be a scientist. She pushed me to be the best student and researcher I could be and her investment in me means the world. Some of the fondest memories in my scientific career thus far were hearing neurons fire for the first time after hours of troubleshooting the electrophysiology rig. Krista is a fantastic scientist and educator and I'm so thankful for our shared laughs and learning experiences.

To my DAC committee- Dr. Santiago Jaramillo, Dr. Mike Wehr, Dr. Caitlin Fausey, and Dr. Matt Smear- thank you for the excitement you shared for my project and for my success as a scientist.

To the members of the Niell lab, past and present, it has been such a blessing working beside such great scientists and even better people. Dr. Jhoseph Shin, Dr. Rolf Skyberg, Dr. Isaac Rhim, Dr. Luis Franco, Dr. Judit Pungor, Dr. Anne Liu, Dr. Des Ramirez, Angelique Allen, Chris Fields, Keaton Jones, Abbi Koenigsmark, Michael Siddikpramana, as well as former members, Dr. Phil Parker, Dr. Elliott Abe, Dr. Denise Piscopo, Dr. Mea Casey-Songco, Dr. Angie Michaiel, Dylan Martins, Emmalyn Leonard, and Nate Casey- you have all made the Niell lab such an incredible environment to work in over the years. Your encouragement, thought-provoking questions, mentorship, and care do not go unrecognized and I am thankful for you all.

I wish to express my deepest appreciation to the members of Gender Inclusion in Neuroscience-past and present. Rubi Ruopp, Kana Suzuki, Matt Nardoci, Angelique Allen, Leah Blankenship, and former members Dr. Rachel Bedford, Dr. Mea Songco-Casey, Dr. Angie Michaiel, Dr. Emily Heckman, Isabella Salinas- you have been such an amazing source of community and constant support throughout my PhD journey. Not only has this group been such an inspiration to advocate for diversity in science but it has also been a safe space to learn and be my true self and for that I'm eternally grateful.

Lastly, I would like to thank the University of Oregon TeACS facility for outstanding animal care, the Institute of Neuroscience as a whole for their support, and recognize the NIH National Eye Institute and Brain Initiative for supporting this work.

DEDICATION

I dedicate this thesis to my parents, Allan and Kellye Sharp, and my husband, Sean Sharp-Speirs.

Mom and Dad, thank you for always being my biggest cheerleaders. You have both sacrificed so much for your daughters, and it does not go unnoticed. Throughout my life you taught me the value of hard work, supported my curiosity, and encouraged me to step beyond my comfort zone because you believed I could do whatever I set my mind to, even if I wasn't so sure. Your unconditional love, encouragement, frequent visits to Eugene, and support throughout this journey means the world to me.

Sean Alexander, you have been my rock throughout this chapter of life. You make me laugh on the longest days, you never fail to remind me of my capabilities, and you always remind me of what matters most in this world, love.

This has been possible in large part because of the unwavering support of my family, thank you from the bottom of my heart. I love you.

TABLE OF CONTENTS

LIST OF FIGURES

Page

CHAPTER I
INTRODUCTION

*This chapter contains previously published co-authored material.

Much of how we perceive the world is dependent on our sense of sight. As we interact with our surroundings our sense of vision often facilitates our behaviors, and informs our decisions as we navigate through constantly changing environments. Vision is a highly conserved sense, as the first known eyes were found in a now extinct marine arthropod, the trilobite, which lived 600 million years ago (Fortey et al., 1996). Through millions of years of evolution, and an estimated 40 different eyes later (Schwab, 2018), the visual system has become increasingly specialized, and 'fine-tuned' to the sensory needs of a given organism. Many species' visual needs rely on continuous incorporation of changing visual information during active sensing. Imagine searching for someone in a crowded room: our visual system allows us to find a specific face in the crowd, carefully navigating through people all while receiving constantly changing visual input. This is a complex visual task that requires the relay of information through thousands of synapses, yet we do this often and with ease. This is a prime example of how vision is an active process, one in which the visual scene is sampled via a series of eye, head, and body movements. The brain rapidly adjusts to changes in our visual environment, constantly updating the representation of the world while accounting for self-motion and encoding rapidly changing visual information. This phenomenon of active sensing as it pertains to visual processing is the main focus of this dissertation.

The output of the mammalian retina projects to two primary visual pathways: cortical and collicular. A large breadth of work in visual neuroscience has investigated the differences and similarities between these visual pathways. The cortical visual pathway encodes visual input from the retinal ganglion cells (RGCs) to the dorsal lateral geniculate nucleus (dLGN) to the primary visual cortex (V1). Whereas the collicular pathway relays visual information directly from RGCs to the superior colliculus (SC). While these two pathways function independently, they are known to be interconnected in early visual processing.

The cortical pathway is substantially expanded in primates for high precision visual processing and fewer RGCs (10%) project directly to colliculus in these animals (Perry & Cowey, 1984). Throughout evolution as the cortex has grown to accommodate more and more complex behaviors and thoughts perhaps collicular pathways share less of the load when it

comes to visual processing, especially in primates. However, in mice the majority of RGCs (~85%) project to superior colliculus (Ellis et al., 2016) suggesting this region may play a more substantial role in mouse vision. Numerous studies have investigated cortex and colliculus' role in mouse vision however, the role of the regions during active vision is largely understudied.

This dissertation consists of three projects focused on different aspects of active sensing during visual processing. In Chapter II, I will describe a project focused on implementing ethologically relevant experimental paradigms to investigate visual processing and active sensing during a jumping task. Chapter III is focused on the neural underpinnings of active vision in mouse V1 during spontaneous free moving exploration of a complex visual environment. In this study we record neural activity in V1 while analyzing responses to gaze-shifting head and eye movements. Lastly, in Chapter IV I explore the impact of the same gaze-shifting head and eye movements on SC during free movement, specifically focused on responses differences across the depth of SC.

1.1 Mice as model organism for visual processing

These studies were completed using a well established model organism in mammalian visual neuroscience, the mouse. Mice are a very popular model system for vision due to their well studied and established genetics which can be easily manipulated and used to dissect the neural circuitry of visual processing. While mice and primates have distinct visual processing niches, many of the same principles found throughout vision research in primates are also found in mouse vision. Much of the neural organization of the visual system and the neural computations that occur during visual processing such as orientation selectivity and the spatial structure of receptive fields are maintained between mouse and primate (Niell & Stryker, 2008; Huberman & Niell, 2011; Niell & Scanziani, 2021). While often thought of as blind, mice do in fact use their vision for a wide variety of tasks such as prey capture, escape, object discrimination, and decision making (Hoy et al., 2019; Campagner et al., 2023; Yu et al., 2018; Odoemene et al., 2018; Huberman & Niell, 2011). Together, this supports the use of mice as a model organism for studying visual processing.

1.2 Chapter II: Distance estimation from monocular cues in an ethological visuomotor task

Vision is an active process, one in which we consistently move our head, eyes, and body to continuously sample the environment around us. Often we use our visual cues to guide us through the world around us. Active vision is key for depth discrimination, determining distance

between observer and objects in the environment. A breadth of research has highlighted the importance of distance discrimination in active vision, discovering different factors that contribute to depth estimation such as retinal image size, motion parallax, and binocular disparity (Sousa et al., 2011; Rogers & Graham, 1979; Durgin et al., 1995). However, many studies looking into depth estimation explore stereo vision using head-fixed or otherwise restrained experimental paradigms. Given the influence of motion parallax, which incorporates self movement, on distance discrimination it is imperative to study distance estimation in unrestrained ethological contexts that allow for self motion. To investigate how distance is estimated during an ethologically relevant behavior we designed a novel visual distance estimation task in mice. We trained mice to jump across gaps of varying distances to variably sized platforms while measuring head, body, and eye movements along with optogenetic inactivation of binocular visual cortex (V1). We found that binocular vision is not required for accurate distance estimation in this task, and that optogenetic shutdown of V1 negatively affected task performance, suggesting V1 plays an important role for distance estimation. Additionally, we found that mice perform compensatory eye movements during this task to stabilize gaze prior to jumping, and that the eyes do not move to create larger binocular overlap prior to the jump. Finally, this work highlights the importance of studying visual processing during naturalistic behaviors. This task paves the way for future studies beyond distance estimation and provides a framework for future studies exploring goal-oriented behaviors, motor control, decision-making, and multimodal stimulus integration within an ethological context.

1.3 Chapter III: A dynamic sequence of visual processing initiated by gaze shifts

Visual processing has often been studied using head-fixed, and by consequence gaze-restrained, experimental paradigms limiting our understanding of how naturalistic movements, specifically head/eye movements, influence visual processing. To determine the role of head and eye movement signals on visual processing in V1 we recorded neural activity with simultaneous head and eye tracking in mice during spontaneous exploration in a visually enriched arena. We found that most V1 neurons respond to gaze-shifting head and eye movements but not to gaze-stabilizing compensatory movements. Neurons in V1 respond after gaze shift initiation with a continuous temporal sequence: early positive, late positive, biphasic, and negative. Additionally, we found this gaze shift initiated temporal dynamic to be visually dependent, as in the absence of visual input (complete darkness) the response dynamic was not

maintained and most neurons showed suppression around the time of gaze shifts. Further, we found that during the presentation of flashed visual stimuli neurons in V1 mimic their gaze shift response. This finding suggests that the temporal dynamic is driven by the abrupt change in visual information that occurs during a gaze-shifting head/eye movement rather than the motor efference signals. We also found V1 responses to gaze shift spatial frequency (SF) preferences are consistent with the coarse-to-fine visual processing model, in which the visual system first responds to coarse features of the visual scene and responds to finer features later. Finally, our observed responses are surprisingly similar across mice and marmoset, specifically as it relates to temporal response profiles and SF preferences. This suggests that some aspects of natural vision are shared across mammals. This work highlights the importance of measuring the dynamic of visual input that occur during naturalistic behavior to better understand the neural underpinnings of the visual system.

1.4 Chapter IV: The impact of active vision on mouse superior colliculus

The SC is a highly conserved midbrain region heavily implicated in visuo-motor tasks, multi-sensory integration, and the generation of head/eye movements. SC is organized topographically with superficial SC (sSC) responding to visuosensory input and deep SC (dSC) responding to multimodal sensory input and generating motor output. While a large breadth of research has explored visual responsiveness in sSC and motor output in dSC, little research has looked at the activity of both regions simultaneously. Additionally, the vast majority of research in mouse SC has been conducted using head-fixation, limiting our understanding of the interactions between the layers of SC during natural movement. We therefore investigated the impact of head and eye movements on SC across depth during free movement utilizing simultaneous neural recording and head/eye tracking. We find that neurons in mouse SC respond differently to gaze shifts based on depth. Neurons in sSC respond with an increase or decrease in response amplitude at gaze shift onset but not during compensatory movements, whereas dSC neurons respond prior to gaze shift onset and persist beyond the duration of gaze shift which we categorize as a motor response. Additionally, dSC neurons also respond to compensatory movements. We further quantify response differences by comparing modulation indices between sSC and dSC. Neurons in sSC are heavily modulated by contrast but not by head movement modulation. While dSC neurons are not modulated by contrast and are heavily modulated by head movement. This suggests that dSC neurons are not responding to the abrupt change in the

visual information that occurs during a gaze-shifting movement but rather to the movements themselves, whereas sSC neurons are responding to the change in visual input during a gaze shift, as further evidenced by a lack of response from sSC neurons during gaze shifts in complete darkness. We also compared gaze shift response dynamics in sSC to known responses in V1. We find that although sSC has less diverse response profiles as compared to V1 they share response sequences that correlate with coarse-to-fine visual processing. This work emphasizes the importance of investigating active vision in unrestrained conditions to better understand the computations performed by the visual system.

CHAPTER II
DISTANCE ESTIMATION FROM MONOCULAR CUES IN AN ETHOLOGICAL VISUOMOTOR TASK

*This chapter contains previously published co-authored material.

2.1 JOURNAL STYLE INFORMATION

Originally published as Parker PRL, Abe ETT, Beatie NT, Leonard ESP, Martins DM, **Sharp SL**, Wyrick DG, Mazzucato L, Niell CM. (2022). Reproduced from *eLife*, 11, 1-18. https://doi.org/10.7554/eLife.74708

2.2 AUTHOR CONTRIBUTIONS:

PRL Parker, Conceptualization, Data curation, Formal analysis, Supervision, Funding acquisition, Writing - original draft, Writing - review and editing; ETT Abe, Software, Formal Analysis, Visualization, Methodology; NT Beatie, Data curation, Investigation; ESP Leonard, Data curation, Investigation, Methodology; DM Martins, Data curation, Software, Formal analysis; SL Sharp, Investigation; DG Wyrick, Software, Formal analysis, Investigation; L Mazzucato, Formal analysis, Supervision, Methodology, CM Niell, Conceptualization, Resources, Supervision, Funding acquisition, Project administration, Writing - review and editing

2.3 INTRODUCTION

Vision is an active process – we continuously move our eyes, head, and body to gain information about the world around us. One core function of active vision is to determine the distance between the observer and objects in its environment. This ability is so critical that many species have evolved to use multiple distinct cues to estimate depth, including retinal image size, motion and position parallax, and binocular disparity (Kral, 2003; Shinkman, 1962). In particular, depth perception through stereo vision has been heavily studied, but other cues that provide important complements are less well understood. Furthermore, some of these monocular cues, such as motion parallax and loom, are closely integrated with movement. How does the brain make use of these diverse cues to guide different behaviors? For instance, is distance explicitly computed and represented in neural activity for some behaviors and implicitly encoded for others? Furthermore, how is this sensory representation converted into the appropriate motor

output? Neurophysiological studies are often performed on head-fixed subjects, limiting the range of depth cues and behaviors that can be studied. Addressing these questions requires behaviors where experimental subjects amenable to neural circuit interrogation can engage in distance estimation behaviors unrestrained (Leopold & Park, 2020; Parker et al.,2020).

The mouse is an important model for vision, yet relatively few behavioral paradigms exist for studying natural, active vision in the mouse (Boone et al., 2021; Hoy et al., 2016; Yilmaz & Meister, 2013).Previous work in other rodent models, including rats and gerbils, showed that animals will accurately jump to distant platforms for a reward, and that changing experimental conditions can bias animals toward the use of certain depth cues, including monocular ones (Carey et al., 1990; Ellard et al.,1984; Goodale et al., 1990). Here, we report that mice are capable of using vision to estimate the distance across a variable gap and execute an accurate ballistic jump. Using this behavior, we show that mice can use monocular vision to judge distance, and suppressing the activity of primary visual cortex (V1) disrupts task performance. Furthermore, this paradigm provides a foundation for studying various visual computations related to depth, and the corresponding motor output, in a species amenable to measurement and manipulation of neural activity in genetically defined populations.

2.4 RESULTS

2.4.1 Mouse distance estimation (jumping) task

In order to establish a framework for studying distance estimation in the mouse, we adapted a gerbil/rat jumping task (Ellard et al., 1984; Legg & Lambert, 1990; Richardson, 1909), where animals were rewarded for successfully jumping across a variable gap (Figure 1A). Mice were free to roam around the arena, then initiated trials by mounting a take-off platform. An occluding barrier was introduced to block the mouse's view while the experimenter randomly placed one of three landing platforms at one of seven distances from the take-off platform (Figure 1). We used landing platforms of variable size to minimize the use of retinal image size cues, which may not require visual cortex for accurate distance estimation after learning (Carey et al., 1990). The trial began as soon as the occluding barrier was removed, and the decision period comprised the time between barrier removal and the last video frame before the mouse executed one of three outcomes (Figure 1C). On 'success' trials, the mouse jumped and landed on the landing platform, and received a reward. On 'failure' trials, the mouse jumped and missed the landing platform, landing on the arena floor, and received no reward. On 'abort' trials, mice

dismounted the take-off platform onto the arena floor and received a mild air puff. Training, which usually took one to two weeks, was complete when mice successfully jumped to each of the three landing platforms at the maximum distance (22 cm). To quantify behavior, markerless pose estimation was performed on side- and top-view video with DeepLabCut (DLC; Mathis et al., 2018).

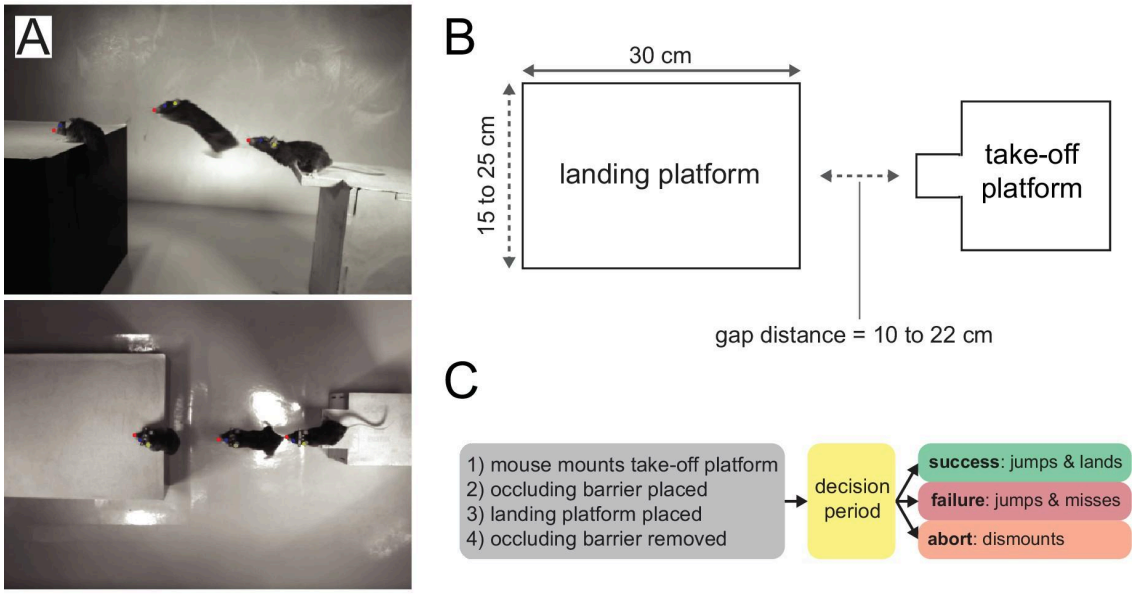


Figure 1. Mouse distance estimation (jumping) task.
(A) Example side and top-down video frames (three overlaid) from single trial. **(B)** A random combination of landing platform width (three sizes) and gap distance (seven distances) is chosen for each trial. **(C)** Trial logic.

2.4.2 Mice accurately estimate distance under binocular and monocular conditions

Mice successfully jumped to all three sizes of platforms at all gap distances (Figure 2A, example, blue lines in B top; N = 3580 trials in eight mice), with only a minor effect of gap distance on success rate (ANOVA, $F = 2.316$, p=0.048). On success trials, the distance jumped increased as a function of gap distance (Figure 2B, blue line in bottom panel; ANOVA, distance $F = 12.845$, p=1.17e-8), showing that mice accurately jumped rather than adapting an alternative strategy (e.g., picking one of two jump forces across the five distances). The gap was too large for mice to reach across with their whiskers, preventing the use of somatosensation to judge the distance. Furthermore, mice did not perform any jumps in the dark (n = 4 mice, four sessions), suggesting that they relied on vision.

A number of depth cues are available in natural contexts. To test the need for stereopsis, we performed monocular eyelid suture (N = 1613 trials in eight mice), after which mice performed equally well at the task with no significant difference in the fraction of success, failure, and abort trials (Figure 2B, magenta lines, and Figure 2—figure supplement 1; ANOVA binocular vs. monocular, failure $F = 0.002$, $p=0.965$, success $F = 0.101$, $p=0.752$, abort $F = 0.275$, $p=0.601$). There was no effect of gap distance on success (ANOVA, $F = 1.593$, $p=0.169$), and mice similarly increased distance jumped as a function of gap distance under monocular conditions (ANOVA, $F = 5.623$, $p=1.68e-4$). These data suggest that binocular vision is not required for accurate distance estimation under these conditions and demonstrate that mice can use monocular cues to accurately judge distance. We also tested for a role of retinal image size by analyzing performance across the three different landing platforms (Figure 2C). Mice performed similarly across all three sizes, suggesting that they did not rely primarily on retinal image size, although distance jumped was influenced by platform size under monocular conditions (ANOVA; success, binocular $F = 2.345$, $p=0.099$, monocular $F = 0.774$, $p=0.463$; distance jumped, binocular $F = 4.436$, $p=0.013$, monocular $F = 3.261$, $p=0.041$). Finally, to determine whether accuracy or precision were altered after monocular occlusion, we calculated the change in the mean landing position between the two conditions, and the standard deviation of the landing position in each condition (Figure 2D). The standard deviation of landing position was similar across conditions and was greater than the difference in the mean landing position, suggesting that mice were equally accurate and precise after monocular occlusion.

2.4.3 Mice perform more head movements under monocular conditions

To quantify the fine-scale structure of behavior leading up to the jump, we analyzed both the movement and position of the mouse during the decision period (Figure 3A) as differences between binocular and monocular conditions could indicate a change in the use of visual cues. To analyze movements, we identified zero crossings in the velocity of eye position from the side-view camera data, then took a 500 ms window around these time points, discarding any with vertical amplitudes less than 1 cm (Figure 3B). We then performed principal component analysis on the concatenated x/y traces, and *k*-means clustering on the reduced data ($k = 10$, see 'Materials and methods' for details). The resulting movement clusters (Figure 3C; ordered by total variance, high to low) showed a diversity of trajectories that together capture most of the head movements that the mice made leading up to the jump (example clusters in Figure 3A and

B). The average trajectories of these movement clusters were highly similar between the binocular and monocular conditions (Figure 3C). The frequency of movements per trial was significantly increased under monocular conditions across clusters (Figure 3D; ANOVA; binocular vs. monocular, $F = 16.633$, p=7.58e-5), though no individual cluster showed a significant increase after accounting for repeated measures (p>0.005). Both the amplitude (Figure 3E; ANOVA; binocular vs. monocular $F = 1.349$, p=0.247) and relative timing (Figure 3F) of movement clusters were unchanged. We confirmed that movement frequency per trial was increased in an additional dataset by performing autoregressive hidden Markov modeling on the nose, eye, and ear positions, and found that binocular and monocular conditions could be differentiated by a simple decoder using the transition probabilities between movement states (Figure 3— figure supplement 1).

To determine whether the increased frequency of head movements reflected an increase in sampling rate, or an increase in temporal integration, we compared the decision period duration (Figure 3G) and total distance moved (Figure 3H). Monocular animals spent more time making the decision to jump (Wilcoxon signed-rank test; binocular 2.48 ± 0.27 s vs. monocular 5.56 ± 1.31 s, p=0.008) and moved a greater distance overall than binocular animals (Wilcoxon signed-rank test; binocular 14.65 ± 1.27 cm vs. monocular 28.73 ± 5.01 cm, p=0.016). These results suggest that mice use a temporal integration strategy when binocular cues are unavailable.

Finally, to determine whether animals changed their head position under monocular conditions, we calculated head angle from the side (pitch) and top (yaw) camera data. On average, mice decreased the pitch of their head (downward tilt) under monocular conditions (Figure 3I; Wilcoxon signed-rank test; binocular –19.76 ± 2.39° vs. monocular –26.84 ± 3.28°, p=0.008) without changing the range (standard deviation) of pitch values (data not shown; Wilcoxon signed-rank test; binocular 15.25 ± 0.81° vs. monocular 16.79 ± 0.72°, p=0.109). Interestingly, the mean change in pitch was inversely correlated with overall change in cluster frequency (Figure 3J; Spearman correlation, p=0.047), suggesting that mice either increase the frequency of vertical head movements or change the positioning of their head. Yaw was not significantly changed (data not shown; Wilcoxon signed-rank test; binocular 17.15 ± 2.44° vs. monocular 17.45 ± 2.93°, p=0.641) while the range of yaw values was significantly increased (data not shown; Wilcoxon signed-rank test; binocular 8.33 ± 1.45° vs. monocular 16.24 ± 4.16°, p=0.039), suggesting an overall increase in the range of side-to-side head movements. These

changes were not associated with changes in movement cluster frequency (data not shown; Spearman correlation, p=0.320).

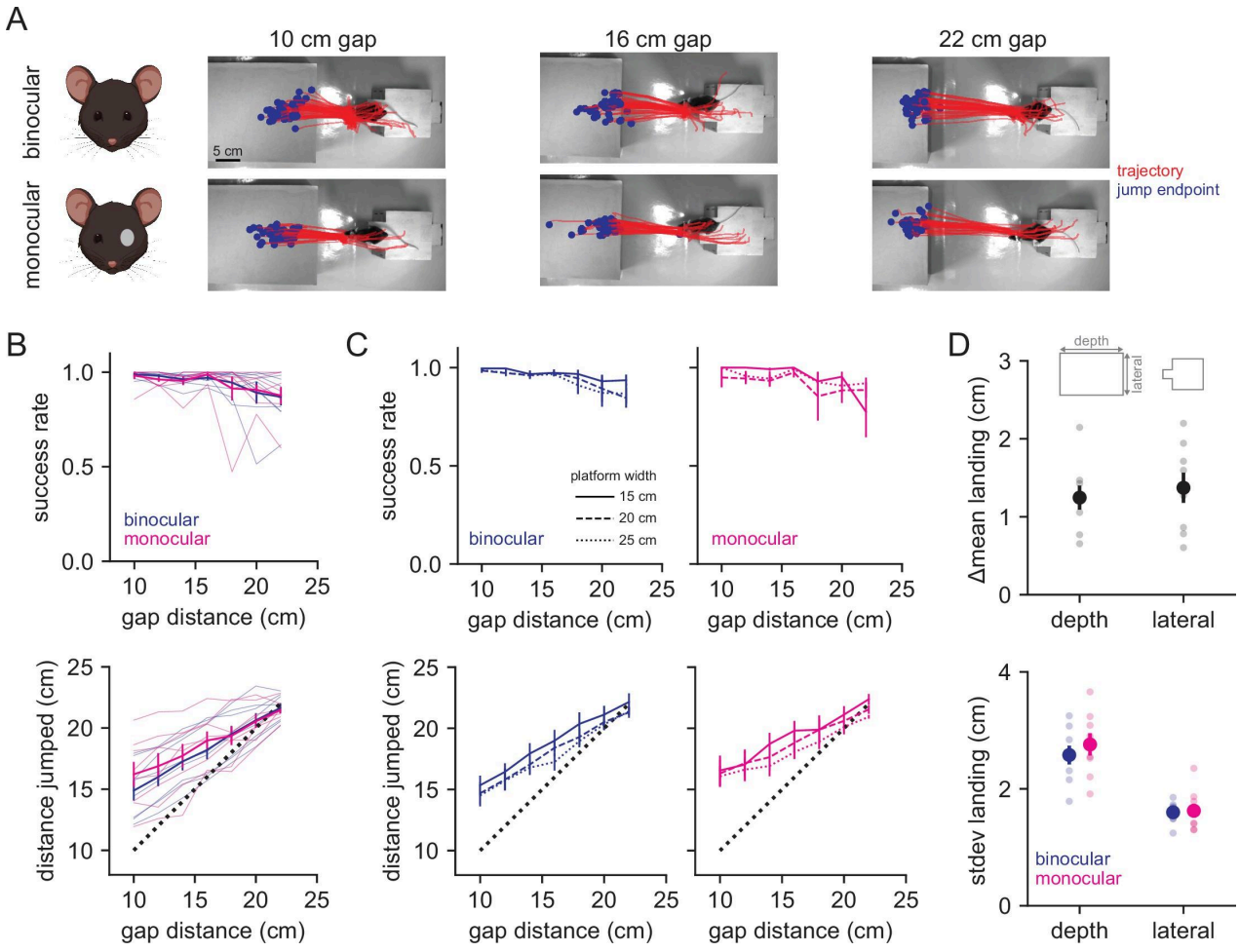


Figure 2. Mice accurately judge distance under binocular and monocular conditions.
(A) Example jump trajectories from a single mouse (red line is trajectory of left ear tracked by DeepLabCut, blue dot is end point of jump) at three distances for binocular (top row) and monocular (bottom row) trials. **(B)** Performance (top) and accuracy (bottom) in binocular (blue, n = 8 mice) and monocular (magenta, n = 8 mice) conditions averaged across landing platform widths. Thin lines are individual animal data. **(C)** Performance (top) and distance jumped (bottom) for bi/monocular conditions by landing platform width (indicated by line style). **(D)** Change in the mean landing position (top) and standard deviation of landing position (bottom) for binocular vs. monocular conditions. Smaller points are individual animal data.

Together, these results show that movement and position of the head during distance estimation differ when binocular vision is no longer available, consistent with a switch from the use of binocular cues to other cues such as motion and/or position parallax that require temporal integration.

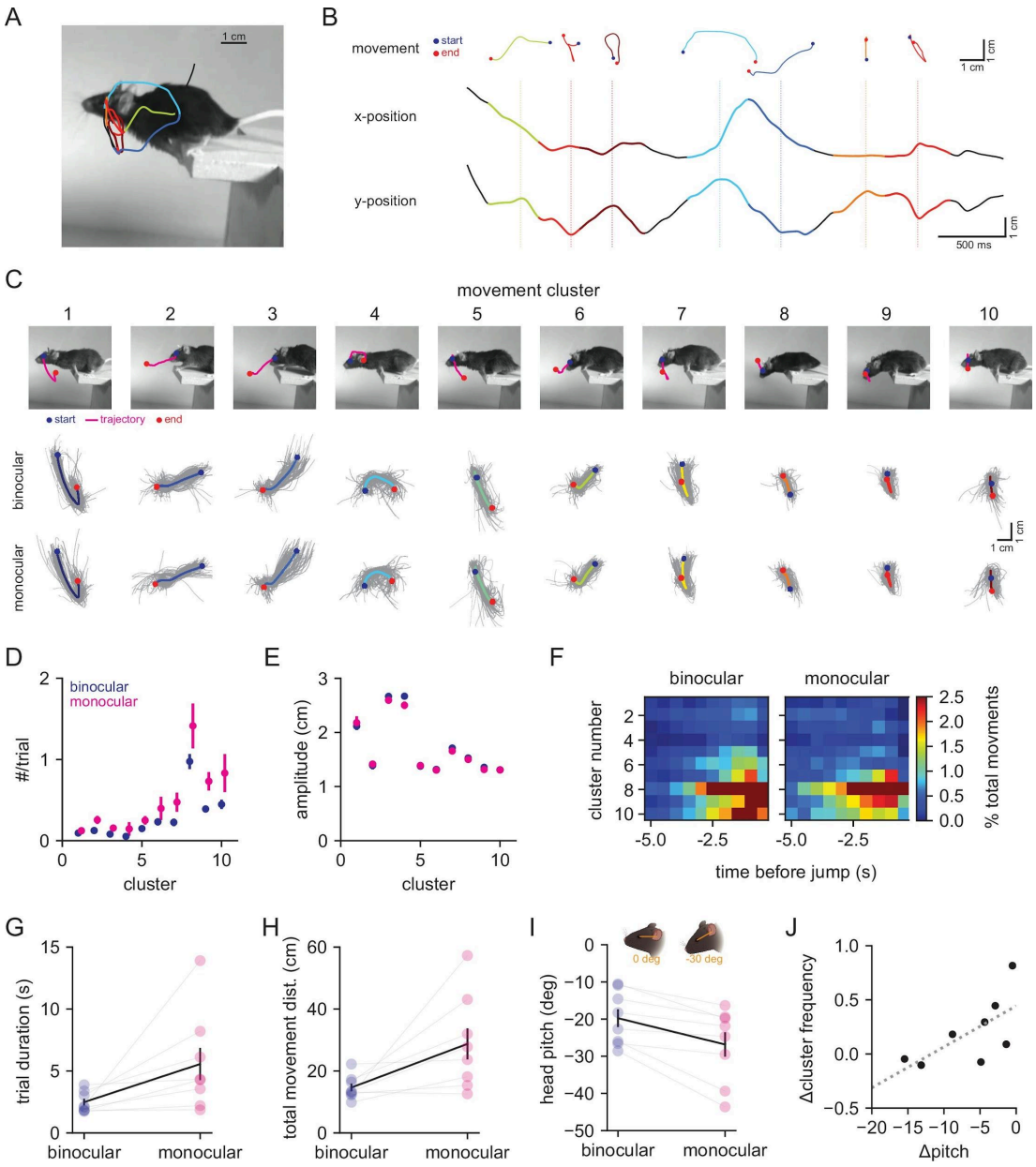


Figure 3. Mice perform more head movements during the decision period under monocular conditions.
(A) Example decision period trajectory of the left eye position overlaid with movements identified through velocity zero crossings. Color corresponds to movement cluster identity in (**B ,C**). Image is the last time point in the decision period. **(B)** Horizontal (x) and vertical (y) positions of eye across time from the trace in **(A)**. Individual movements are plotted above as x/y traces, with dotted lines corresponding to the middle time point, and blue and red points indicating the start and end, respectively. Colors correspond to clusters in (**C**). **(C)** Top: example individual movements from 10 k-means clusters; magenta is the trajectory, blue and red are start and end points, respectively. Bottom: individual movement clusters for binocular (top row) and monocular (bottom row) conditions, with means

Figure 3 continues on next page

Figure 3 continued

plotted over 100 individual examples in gray. **(D)** Mean number of movements per trial for each cluster in binocular (blue) vs. monocular (magenta) conditions. **(E)** Mean amplitude of movement clusters for binocular (blue) and monocular (magenta) conditions. **(F)** Normalized movement frequency as a function of time before the jump for all clusters. **(G)** Mean trial duration (decision period only) for the two conditions. **(H)** Mean of the total distance traveled by the eye during the decision period for the two conditions. **(I)** Mean head pitch, measured as the angle

between the eye and ear, across the decision period for the two conditions. **(J)** Relationship between the change in head pitch and change in cluster frequency between the binocular and monocular conditions. Dotted line is fit from linear regression.

2.4.4 Eye movements compensate for head movements to stabilize gaze

Previous work shows that the majority of eye movements in rodents, including mice, are compensatory for head movements, and that saccades occur primarily as a consequence of large-amplitude head movements (Meyer et al., 2020; Michaiel et al., 2020; Wallace et al., 2013). Some species make horizontal vergence eye movements to increase binocular overlap during behaviors such as prey capture (Bianco et al., 2011). To determine how mice target their gaze during binocular distance estimation, we performed bilateral eye tracking using miniature head-mounted cameras (Meyer et al., 2018; Michaiel et al., 2020; Sattler & Wehr, 2020), then used DLC to track the pupil in order to quantify horizontal and vertical eye movements (Figure 4A). Importantly, mice continued to accurately perform the task despite the head-mounted hardware and tether (~3 g weight; Figure 4B and C). It should be noted that these experiments were performed in a different set of animals with narrower landing platforms than the other experiments in this study, so performance is worse relative to the data in Figure 2; however, subjects showed no difference in their performance relative to their performance under control conditions (ANOVA; control vs. eye cameras $F = 0.373$, p=0.543). Head pitch (vertical head angle) was anticorrelated with both eye vergence (horizontal angle of the two eyes) and eye phi (vertical eye movements) both during the early portion of the decision period when mice were approaching the jump (start of trial to 2 s before jump; pitch vs. vergence $R^2 = 0.51$, phi $R^2 = 0.28$) and in the late portion of the decision period immediately prior to the jump (2 s prior to jump; pitch vs. vergence $R^2 = 0.70$, phi $R^2 = 0.54$; Figure 4D and E). Thus, upward head movements caused the eyes to move down and toward the nose, while downward head movements caused upward and outward eye movements, consistent with vestibulo-ocular reflex-mediated gaze maintenance throughout the decision period. Additionally, while there was a slight change in vergence between the early and late periods of the trial (vergence early 1.77 ± 0.21°, late –1.66 ± –0.29°; p=1.57e-4), this was explained by a similar difference in head pitch between these two periods (pitch early –45.45 ± 1.24°, late –38.47 ± 1.43°; p=4.97e-5), demonstrating that mice do not move their eyes to increase binocular overlap preceding the jump.

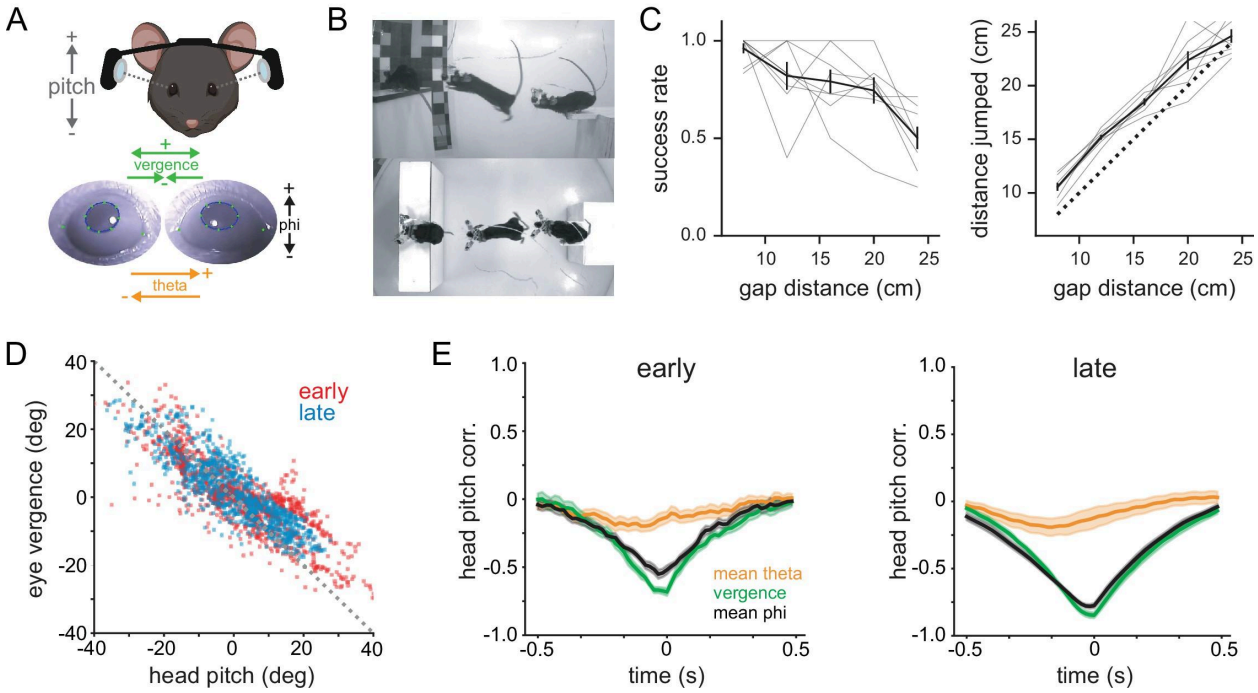


Figure 4. Eye movements compensate for head movements to stabilize gaze. (A) Schematic of experimental setup for measuring head and eye movements; bilateral eye tracking with miniature head-mounted cameras (top) and ellipse fitting of DLC-tracked pupil points (bottom). **(B)** Side and top view images of a mouse performing the task with the eye tracking system (three frames overlaid). **(C)** Performance (left) and distance jumped (right) for eye-tracking experiments. Gray lines are individual animal data. **(D)** Horizontal angle between the two eyes (eye theta divergence) as a function of head pitch during the decision period. 'Early' is from the start of the trial to 2 s before the jump, and 'late' is the 2 s preceding the jump. **(E)** Mean eye theta, eye theta vergence, and eye phi cross-correlations with head pitch angle for early (left) and late (right) portions of the decision period; n = 8 mice for all plots.

2.4.5 V1 optogenetic suppression disrupts distance estimation task performance

Finally, we tested whether distance estimation behavior requires visual cortex. We first asked whether suppressing the activity of binocular zone V1, which corresponds retinotopically to the central visual field in front of the mouse, affected task performance. Bilateral optic fibers were implanted at the surface of the cortex above binocular V1 in either control mice (PV-Cre) or mice expressing channelrhodopsin-2 (ChR2) in parvalbumin-expressing inhibitory interneurons (PV-Cre:Ai32, referred to as PV-ChR2 here; Hippenmeyer et al., 2005; Madisen et al., 2012), all of which had intact binocular vision (Figure 5A, left column; control n = 948 trials in four mice, PV-ChR2 n = 911 trials in four mice). On a third of trials, light was delivered through the implanted optic fibers during the decision period (470 nm, 5 mW/mm^2 , 40 Hz, 50%

duty cycle). Control animals showed no change in performance with the laser on (ANOVA; laser off vs. on, failure $F = 0.030$, p=0.864, success $F = 0.026$, p=0.872, abort $F = 0.070$, p=0.793), whereas PV-ChR2 animals showed a significant reduction in performance across distances (Figure 5B; ANOVA; laser off vs. on, failure $F = 7.836$, p=0.008, success $F = 15.252$, p=3.35e-4, abort $F = 10.876$, p=0.002; see Figure 5—figure supplement 1 for a breakdown of all three outcomes). On success trials, the mean landing position of PV-ChR2 mice was significantly changed compared to controls (Figure 5C; t-test, PV-ChR2 vs. control, p=0.019) while the standard deviation of landing position was not significantly different (t-test, PV-ChR2 vs. control, p=0.249). Interestingly, similar to mice that underwent monocular occlusion, PV-ChR2 mice showed decreased head pitch on success trials with the laser on (Figure 5D; t-test, PV-ChR2 vs. control, p=0.004), with no change in the standard deviation of pitch (t-test, PV-ChR2 vs. control, p=0.268) or in yaw (t-test, PV-ChR2 vs. control, mean p=0.116, SD p=0.164). There was no laser-associated change in trial duration, movement cluster frequency, or amplitude (data not shown; t-test, trial duration p=0.391; ANOVA, frequency p=0.106, amplitude p=0.930). Together, these data show that suppression of binocular zone V1 in animals with intact binocular vision significantly decreases distance estimation task performance, and that the ability to successfully perform the task is associated with changes in both pre-jump behavior and landing position.

We next asked whether mice with monocular vision would be affected by binocular zone V1 suppression (Figure 5A, middle column). Interestingly, PV-ChR2 mice showed no change in success rate (Figure 5B; ANOVA; laser off vs. on, failure $F = 2.388$, p=0.130, success $F = 0.389$, p=0.536, abort $F = 0.090$, p=0.766), and no change in landing position (Figure 5C; t-test; PV-ChR2 vs. control, mean p=0.189, SD p=0.955) or head position (Figure 5D and E; t-test; PV-ChR2 vs. control, pitch mean p=0.153, pitch SD p=0.117, yaw mean p=0.903, yaw SD p=0.849). There were also no laser associated changes in trial duration, movement cluster frequency, or amplitude (data not shown; t-test, trial duration p=0.712; ANOVA, frequency p=0.882, amplitude p=0.296). This suggests that once animals switch to using monocular cues, binocular zone V1 is no longer required for accurate task performance, and that therefore the peripheral visual field is engaged.

Lastly, we asked whether suppression of monocular zone V1, which corresponds retinotopically to the peripheral visual field, affected task performance of mice with monocular

vision (Figure 5A, right column). PV-ChR2 mice were less successful at the task with optogenetic suppression due to an increase in the number of abort trials, whereas the fraction of failure trials remained unchanged (Figure 5B; ANOVA; laser off vs. on, failure $F = 0.555$, $p=0.462$, success $F = 10.120$, $p=0.003$, abort $F = 4.663$, $p=0.039$). On success trials, PV-ChR2 mice showed no significant change in landing position (Figure 5C; t-test; PV-ChR2 vs. control, mean $p=0.189$, SD $p=0.200$); however, they showed changes in both pitch mean (Figure 5D; t-test; PV-ChR2 vs. control, mean $p=0.002$, SD $p=0.486$) and yaw standard deviation (Figure 5E; t-test; PV-ChR2 vs. control, mean $p=0.388$, SD $p=0.018$), consistent with shifting head position to utilize a portion of the visual field unaffected by optogenetic suppression. There was no laser-associated change in trial duration, movement cluster frequency, or amplitude (data not shown; t-test, trial duration $p=0.050$; ANOVA, frequency $p=0.476$, amplitude $p=0.344$). It should be noted these experiments were performed in a separate group of mice with narrower landing platforms, thus comparisons of success rate and landing position to the other experimental groups may not be insightful.

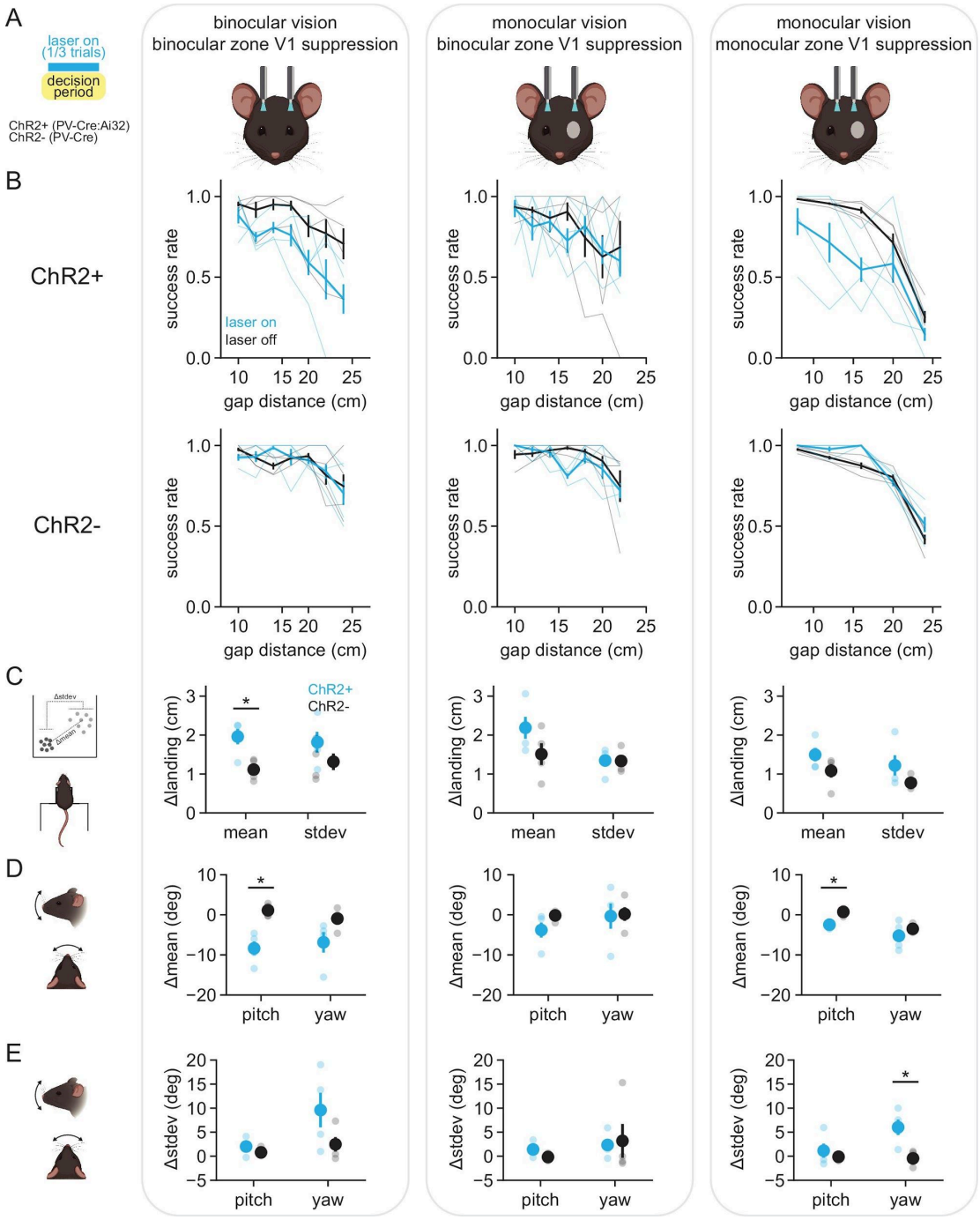


Figure 5. V1 optogenetic suppression disrupts distance estimation task performance. **(A)** Schematic of experimental setup for optogenetic experiments; bilateral illumination of either binocular or monocular zone V1 in either binocular or monocular animals during the decision period on one-third of trials. All plots within a column correspond to the schematized condition. **(B)** Performance curves for laser-off (black) and laser-on (cyan) conditions in mice expressing ChR2 in PV+ inhibitory interneurons (ChR2+, top row) or PV-Cre only mice (ChR2-, bottom row). Thin lines are individual animal data. **(C)** Change in the mean and standard deviation of landing positions, averaged across mice. Small circles are individual animal data. **(D)** Change in the mean head angle for up-down (pitch) and side-to-side (yaw) head position. **(E)** Same as (**D**) but change in standard deviation of pitch and yaw.

In summary, these experiments show that manipulating V1 activity significantly alters the behavior of mice on the distance estimation task, and the correspondence between the anatomical locus of suppression and the ocular condition supports a role for V1 in both binocular and monocular cue mediated distance estimation task performance.

2.5 DISCUSSION

We have established a visual distance estimation task in mice that engages an ethological, freely moving behavior. Previous research using similar versions of this task suggests that gerbils and rats utilize multiple cues to determine the distance to objects in the environment, including retinal image size, binocular vision, and motion parallax (Carey et al., 1990; Ellard et al., 1984; Goodale et al., 1990; Legg & Lambert, 1990). Importantly, this task is distinct from 'gap-crossing' tasks (Hutson & Masterton, 1986) where animals can use the whisker somatosensory system to determine the distance across a short gap. Furthermore, in contrast to other recently developed tasks that are designed to probe binocular depth perception and stereopsis (Boone et al., 2021), in this task mice are able to use monocular cues for depth, including those that are generated by self-movement. It can therefore be flexibly used to investigate a variety of distance estimation tactics by manipulating experimental conditions.

2.5.1 Cues for distance estimation

Binocular vision (and therefore stereopsis) was not required for accurate distance estimation in this task, consistent with previous studies on gerbils (Ellard et al., 1984). This provides the first demonstration that mice are able to use depth cues that are available besides stereopsis. Vertical head movements sufficient to generate motion parallax cues are increased in frequency under monocular conditions, suggesting that mice may use motion parallax under monocular vision in this task. Previous work found increasing frequency of vertical head movements as a function of gap distance in gerbils and rats (Ellard et al., 1984; Legg & Lambert, 1990). We did not see such a relationship, which could be due to species-specific differences or differences in task design. Interestingly, there was an inverse relationship between changes in movement frequency and head angle – animals with the smallest change in frequency of head movements showed the largest change in downward head angle. This could reflect the use of position parallax cues (comparing two perspectives, initial and final, rather than motion cues) and was also present in animals with binocular vision when binocular zone V1 activity was suppressed. These results do not rule out the use of binocular disparity under normal conditions –

in fact, changes in decision period head position and movement between binocular and monocular conditions, in addition to our optogenetic suppression experiments in binocular zone V1, provide evidence that animals use binocular cues as well under normal conditions.

The increase in decision period duration and total distance moved by mice in the monocular relative to binocular condition is consistent with an active sensing strategy that requires temporal integration. The inherent separation of the decision period from the jump execution in this task helps separate potential active sensing behavior from pure jumping behavior – in theory, the animal need not make any movements except those required to jump (Stamper et al., 2019). Closed-loop control of the sensory environment is a powerful tool for investigating active sensing (Biswas et al., 2018) in this task, altering the landing platform based on head movements would provide a causal test to determine whether mice use motion parallax cues, as was performed in locusts (Wallace, 1959). Furthermore, investigating the relationship between active sensing and memory-guided visual behavior could provide new insights into natural behavior and its neural basis. For example, experiments with a modified version of this task suggest that parietal cortex is necessary for context-dependent use of retinal image size cues (Ellard & Sharma, 1996). Future research could investigate how these contextual associations are formed through active sensing and bound into memory in the brain.

2.5.2 Eye movements during distance estimation

Using miniature cameras to track the eyes, we found that eye movements compensate for head movements to stabilize gaze leading up to the jump. This would allow the mouse to both maintain gaze toward the platform and reduce motion blur throughout large-amplitude head movements (Land, 1999). This is consistent with previous work showing that mouse eye movements stabilize gaze during both operant behavior (Meyer et al., 2020) and a natural behavior, prey capture (Michaiel et al., 2020). It will be interesting to determine whether jumping mice control their gaze to localize the platform on a specific subregion of the retina; that is, whether there exists a retinal specialization for determining distance. In the case of prey capture, the image of the cricket is stabilized in the retinal region with the highest concentration of alpha-ganglion cells (Holmgren et al., 2021). We also found that mice did not move their eyes to increase binocular overlap during the period immediately preceding the jump, similar to a previous finding demonstrating that rats do not align the gaze of the two eyes before crossing a short gap (Wallace et al., 2013). Smooth eye movements provide extra-retinal signals for

computing depth from motion parallax in primates (Kim et al., 2017; Nadler et al., 2008; Nadler et al., 2009), and future studies may address whether the compensatory movements we observed play a similar role in mice. Finally, these experiments show that mice are capable of performing this task with a tether and significant hardware weight on the head, which is a critical requirement for introducing additional techniques such as electrophysiology into this paradigm.

2.5.3 Neural circuits underlying distance estimation

We provide evidence for V1 specifically being important for distance estimation behavior in mice. Given the large volume of V1 (~5 mm^3) relative to the spread of laser-induced suppression in PV-ChR2 mice (~1 mm^3 , Li et al., 2019), animals likely maintained some V1 function across all conditions, which would explain why animals in all conditions could still perform the task to some degree, and did so with associated changes in how they orient their heads (i.e., using different parts of the visual field). Whether animals would perform this task with total V1 suppression is unclear given the option to abort trials. However, the fact that the anatomical locus of suppression (binocular vs. monocular zone V1) determined whether there was an effect on behavior between ocular conditions (binocular vs. monocular vision) supports the hypothesis that mice use V1 to estimate distance in this task. In fact, binocular V1 suppression led to the same changes in decision period head angle in mice with binocular vision that monocular occlusion induced, suggesting a common shift in strategy with the loss of stereopsis.

These results are consistent with previous work showing broad lesions of occipital cortex disrupt performance without affecting head movements, whereas lesions to superior colliculus and preoptic area had no effect on either (Ellard et al., 1986). This task could therefore be a useful tool for studying the specific computations performed in V1 that mediate accurate distance estimation, and both the visual and nonvisual input signals required to perform these computations. Additionally, the neural circuits that convert visual information into a jump command are also not well understood. Most work has examined jumping in nocifensive and defensive contexts rather than navigation (Barik et al., 2018; Wang et al., 2015), although a recent behavioral study demonstrated that squirrels learn to integrate multiple factors, including gap distance and branch flexibility, in executing a jump (Hunt et al., 2021).

2.5.4 Utility of studying natural distance estimation behavior

Natural behavior is often a continuous control process, which is fundamentally closed-loop, unlike stimulus-response paradigms that dominate behavior literature (Cisek, 1999). This task accordingly permits investigation of both how movement through the environment generates sensory cues useful for judging distance, and how the visual information is directly converted into a motor output. Furthermore, perception of spatial layout is an embodied process, and thus body- and action-scaling cues that are not available under conditions of restraint could provide distance information under the freely moving conditions of this task (Fajen, 2021). Critically, natural behaviors may be the most appropriate tool for studying the neural basis of sensory processing since theoretical considerations suggest that neural circuits may perform suboptimal inference under non-natural conditions (Beck et al., 2012). Finally, beyond studies of visual distance estimation, this task could provide a framework for integrated studies of motivation, motor control, and decision-making within an ethological context.

2.6 MATERIALS AND METHODS

2.6.1 Animals

Male and female mice between postnatal day 40 (P40) and P365 were bred in-house in a C56BL/6J background. For optogenetic experiments, transgenic mice were used to target the expression of channelrhodopsin-2 to parvalbumin-expressing neurons (PV-Cre [B6;129P2-Pvalbtm1(cre)Arbr/J, Jax #008069] crossed to Ai32 [B6;129S-Gt(ROSA)26Sortm32(CAG-COP4*H134R/EYFP)Hze/J, Jax #012569]; Hippenmeyer et al., 2005; Madisen et al., 2012). Mice were housed in a reverse 12 hr light–dark cycle room. Mice were placed under a water restriction schedule at the start of training, only receiving fluids during training/task periods. All procedures were performed in accordance with the University of Oregon Institute for Animal Care and Use Committee and Animal Care Services standard operating procedures.

2.6.2 Behavioral apparatus and jumping task

Two cohorts of mice were used in this study, with a subset of experiments (eye cameras and monocular V1 suppression) using a smaller arena, narrower landing platforms, and fewer gap distances. All other experimental components of the task were identical. The jumping arena was roughly 45 cm high, 70 cm wide, and 100 cm across for most experiments, and was 30 cm high and 60 cm across for eye camera/monocular V1 suppression. Mice self-initiated trials by

mounting a take-off platform (15 cm height, 10 cm width, 10 cm depth, with 4 × 5 cm overhang in front). While blocking the mouse's view of the arena with a barrier, the experimenter then placed one of three platforms (15, 20, or 25 cm width, 30 cm depth, 19 cm height for most experiments; 10, 20, or 30 cm width, 5 cm depth, 19 cm height for eye camera/monocular V1 suppression) at a random distance (10, 12, 14, 16, 18, 20, or 22 cm for most experiments; 8, 12, 16, 20, or 24 cm for eye camera/monocular V1 suppression) from the edge of the take-off platform. Platforms were custom built from ¼″ acrylic, and tops were coated in white rubberized coating (Plasti-Dip) or fine-grained white sandpaper to prevent animals from slipping. For eye camera/monocular V1 experiments, the platforms were white and a black strip was placed across the top leading edge of the landing platform, matched proportionally in height to platform width to maintain height/width ratio. Arena and platforms were constructed by the University of Oregon Technical Science Administration. A static white noise background composed of grayscale squares (~1° each of visual angle from take-off platform) was mounted at the back of the arena. Six LED puck lights were evenly spaced around the top of the arena for even illumination. Cameras (FLIR BlackFly S USB3) were mounted above and to the side of the arena, and the entire behavioral session was recorded (1440 × 1080 pixels at 99.97 fps, or 720 × 540 pixels at 60 fps) with camera timestamps using a custom Bonsai workflow (Lopes et al., 2015). A custom Python script was used to generate randomized platform/distance combinations for the experimenter and to log trial outcomes and approximate jump times. The moment the barrier was lifted and the mouse was able to see the landing platform constituted the trial start, and the time elapsed until the mouse jumped was the 'decision period.' There were three possible trial outcomes: (1) the mouse jumped and successfully reached the landing platform and received a reward (success), (2) the mouse jumped and missed the landing platform and received no reward (failure), or (3) the mouse dismounted the take-off platform and received a light airpuff and a time-out (abort).

All experiments included data from eight mice (same mice for binocular, monocular, and binocular V1 suppression; same mice for eye cameras, monocular V1 suppression, and supplementary ARHMM figure) and typically lasted 30–45 min. The mean ± standard error for the number of sessions and trials per mouse was as follows: binocular sessions 11.4 ± 0.6, trials 42.1 ± 2.3; monocular sessions 5.6 ± 0.2, trials 39.0 ± 3.4; eye cameras sessions 2.8 ± 0.3, trials 13.9 ± 0.7; binocular vision/binocular zone V1 suppression sessions 6.8 ± 0.2, trials 34.3 ± 2.6;

monocular vision/binocular zone V1 suppression sessions 4.8 ± 0.2, trials 33.5 ± 1.8; monocular vision/monocular zone V1 suppression sessions 3.6 ± 0.2, trials 27.2 ± 1.7. The number of trials was partly limited by the manual nature of the task, which required an experimenter to manually place a platform at a specific distance. However, this process typically took less than 3 s, whereas the time spent consuming the reward and freely investigating the arena between trials was significantly longer. Future versions of the task with automated platform placement and reward delivery could increase the number of trials per session, and increase the length of individual sessions. Acquisition files are available online at https://github.com/nielllab/nlab-behavior/tree/master/jumping/bonsai(copy archived at swh:1:rev:44a40fdbd63ed7740a73b8d085333c8d1b22c592; path=/jumping/bonsai/; *niellab, 2022*).

2.6.3 Behavioral training

Mice were habituated to the arena for 3 days with their cage mates, during which time they were individually handled and introduced to a clicker that indicated water reward (~25–50 µl), where each click is immediately followed by a reward. Mice were then individually clicker trained to mount a short take-off platform (10 cm height; click and reward upon mounting the platform), receiving water (administered by hand using a 1 ml syringe), and a small piece of tortilla chip (Juanita's). After 3–5 successful mounts, a landing platform (19 cm height) was placed against the take-off platform, and mice were clicker-rewarded for climbing up onto the landing platform. After three successful trials, the landing platform was moved slightly farther back, increasing the gap distance until jumping is required to reach the landing platform. At this point, the clicker was typically no longer required. Once the mouse could jump to the maximum distance, the taller take-off platform used in the task was introduced, and landing platforms were again introduced at short distances and slowly moved farther away. Training was complete when mice could jump to all three landing platforms at the farthest distance, and typically took 1–2 weeks with all mice successfully learning the task.

2.6.4 Surgical procedures

For all procedures, anesthesia was induced at 3% isoflurane and maintained at 1.5–2% in O2 at a 1 l/min flow rate. Ophthalmic ointment was applied to both eyes, and body temperature was maintained using a closed-loop heating pad at 37°C. In order to minimize stress when

plugging in optical tethers or miniature cameras, a small steel headplate was mounted on the skull using dental acrylic (Unifast LC) to allow for brief head-fixation before the experiment.

2.6.5 Monocular suture

The area immediately surrounding the eye to be sutured was wiped with 70% ethanol before ophthalmic ointment was applied. Two to three mattress sutures were placed using 6-0 silk suture, opposing the full extent of the lid. The forepaw and hindpaw nails ipsilateral to the sutured eye were trimmed to help minimize postprocedural self-inflicted trauma.

2.6.6 Optic fiber implant

A minimal portion of scalp was resected bilaterally over visual cortex, and a small trepanation was made over each primary visual cortex (+1.0, ±2.5 mm for monocular zone or +1.0, ±3.0 mm for binocular zone, relative to lambda suture). Bilateral optic fibers (ceramic ferrules, thorlab fiber 0.5 mm length from end of ferrule) were stereotactically lowered into the burr hole and secured in place with dental acrylic. Vetbond was then applied to secure the skin in place around the implant. Fiber transmission rates were measured prior to implant and accounted for during experiments.

2.6.7 Miniature head-mounted cameras

To obtain high-resolution video of the eyes during behavior, a miniature camera (iSecurity), magnifying lens (12 mm focus, 6 mm diameter), and an infrared LED were mounted on a custom designed 3D-printed plastic camera arm (Michaiel et al., 2020). Two miniature connectors (Mill Max 853-93-100-10-001000, cut to 2 × 4 pin) were glued to the headplate, and an equivalent connector on the camera arm was plugged in prior to the experiment. Camera power and data were passed through thin tethering wire (Cooner #CZ1174CLEAR) and acquired in Bonsai with system timestamps. The total hardware weight was approximately 3 grams. Eye videos were deinterlaced to achieve 60 fps (matching top/side cameras) prior to analysis

2.6.8 Data analysis

Full task videos were first split into individual trials using custom Python software; the trial start, jump, and landing frame numbers were determined and individual trial videos were saved. These trial videos were then labeled using markerless pose estimation with DLC. A set of sample frames were manually labeled and used to train two networks (top/side cameras and eye cameras) that were then used to track features in all video data. The distance jumped was calculated using the position of the left ear in the top-camera frame where the front paw touched

the platform and the animal decelerated (success trials) or when the front paw passed below the edge of the landing platform (failure trials). DLC points from the side-camera data during the decision period were passed through a median filter (n = 3) and a convolutional smoothing filter (box, n = 5). To extract individual movements, we identified all zero crossings in the velocity trace calculated from eye position, then extracted the 500 ms period around those time points in the eye position data. Movements that overlapped by more than 250 ms with a previous movement were excluded from further analysis. The x and y values were concatenated, and all movements across all conditions were fed into principal component analysis, after which *k*-means clustering was performed on the reduced data. We tried varying the number of clusters across a range of values (4–20) and found that $k = 10$ resulted in clusters that appeared to contain a single type of movement while minimizing repeated clusters of the same movement type. Pitch and yaw were both calculated from the angle between the eye and ear in the side and top video data, respectively.

For hidden Markov model (ARHMM) analysis, values of the points tracking the nose, eye, and ear were used as inputs for training after centering across experiments by subtracting off values of the point that tracked the edge of the take-off platform. Model training was performed using the SSM package in Python (Linderman et al., 2019). Model selection was based on the elbow in twofold cross-validation log-likelihood curves across model iterations while balancing model interpretability with model fit (final model: K = 6, lag = 1, kappa = 1e04, data temporally downsampled 2×, and one state discarded due to extremely low prevalence). ARHMM states were determined based on a posterior probability threshold of 0.8. Time points below the threshold were excluded from analysis. For lexical transition matrices, trials were first separated into binocular and monocular conditions. During the decision period of each trial, the transitions between unique ARHMM states were counted. The number of state transitions was then normalized by the total number of unique transitions per condition to calculate the relative frequency of transitions. For all summary analyses, data were first averaged within animal (across days) and then across animals within a group (e.g., monocular, binocular). Statistical significance was determined using ANOVA and the Student's t-test with Bonferroni corrections for multiple comparisons. Unless otherwise noted, data are presented as mean ± standard error of the mean. Analysis code is available online at https://github.com/nielllab/nlab-behavior/tree/master/jumping/python%20files (copy archived at

swh:1:rev:44a40fdbd63ed7740a73b8d085333c8d1b22c592; path=/jumping/python%20files/; *niellab, 2022*). Data are available at https://doi.org/10.5061/dryad.r7sqv9sg2.

2.6.9 Decoding analysis

We decoded the experimental condition (binocular vs. monocular, laser on vs. off) per animal from single-trial maximum a priori (MAP) motif sequences inferred using the ARHMM. Specifically, we trained binary decoders with a linear decision boundary (linear discriminant analysis) to decode the above categorical variables from the single-trial empirical state transition probability matrices derived from the MAP sequence of each trial, thus providing not only state usage information, but transitions between states as information the classifier could use. For each animal, correct trials were pooled across distances to provide enough trials per class for decoding. Data were split into training and test datasets in a stratified 10-fold cross-validation manner, ensuring equal proportions of trials of different types (distance, platform width, visual condition, laser) in both datasets. To calculate the statistical significance of decoding accuracies, we performed an iterative shuffle procedure on each fold of the cross-validation, shuffling training labels and testing on unshuffled test labels 100 times to create a shuffle distribution for each fold of the cross-validation. From these distributions, we calculated the z-score of decoding accuracy for each class in each cross-validation fold. These z-scores were then averaged across the folds of cross-validation and used to calculate the overall p-value of the decoding accuracy obtained on the original data. The decoding weights of the binary classifiers were examined as well to identify the significant transitions that contributed to decoding between visual conditions. The same shuffle procedure was used to assess significant elements of the classifier.

2.6.10 Statistics

All summary data in text and plots, unless noted otherwise, are mean ± standard error. All statistical tests were performed with SciPy, and in cases where data were not normally distributed, nonparametric tests were used.

2.7 ACKNOWLEDGEMENTS

We thank members of the Niell lab for helpful discussions, and Dr. Michael Goard, Dr. David Leopold, Dr. Matt Smear, and Dr. Michael Stryker for feedback on the manuscript. This work benefited from access to the University of Oregon high-performance computer, Talapas, and the University of Oregon Technical Science Administration. This work was supported by

grants from the National Institutes of Health: 5F32EY027696 (Parker), 1R21EY032708 (Niell, Parker), 1R34NS111669 (Niell). Some figure panels were made using https://biorender.com/.

2.8 ADDITIONAL INFORMATION

2.8.1 Ethics

This study was performed in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. All procedures were performed in accordance with the University of Oregon Institute for Animal Care and Use Committee and Animal Care Services standard operating procedures, under approved protocol AUP-21-21.

CHAPTER III
A DYNAMIC SEQUENCE OF VISUAL PROCESSING INITIATED BY GAZE SHIFTS

*This chapter contains previously published co-authored material.

3.1 JOURNAL STYLE INFORMATION

Originally published as PRL Parker*, DM Martins*, ESP Leonard*, NM Casey, **SL Sharp**, ETT Abe, MC Smear, JL Yates, JF Mitchell, and CM Niell. (2023) Reproduced from *Nature Neuroscienc*e, 26, 2192-2202.

*These authors contributed equally.

3.2 AUTHOR CONTRIBUTIONS

PRL Parker and CM Niell conceived the project. PRL Parker and ESP Leonard led mouse experiments. DM Martins and CM Niell led data analysis. NM Casey and SL Sharp contributed to mouse experiments. ETT Abes contributed to data analysis. MC Smear generated the audio track from mouse neural activity. JL Yates and JF Mitchell performed marmoset experiments. JL Yates, JF Mitchell, and DM Martins performed marmoset data analysis. All authors contributed to writing and editing the manuscript.

3.3 INTRODUCTION

Visual perception is an active process in which animals move through their environment and sample the visual scene through movements of the body, head and eyes. Self-motion and sampling behavior have a profound impact on the dynamic pattern of retinal visual input, which can play an integral role in processing and interpreting the visual scene (Boi et al., 2017; Ahissar & Arieli, 2001; Schroeder et al., 2010; Gibson, 1979). Correspondingly, a variety of movement-related signals are observed throughout the brain (Parker et al., 2020). In the visual cortex, eye movement signals have been studied in nonhuman primates and cats (Burchfiel & Duffy, 1974; Leopold & Logothetis, 1998; Nishimoto et al., 2017; Noda & Adey, 1974; Sommer & Wurtz, 2008) and recently in mice (Miura & Scanziani, 2022). Cortex-wide signals of locomotion and body/facial movements are present in head-fixed mice (Niell & Stryker, 2010; Stringer et al., 2019; Musall et al., 2019) and head rotation signals are seen in rodent visual cortex (Vélex-Fort et al., 2018; Meyer et al., 2018; Bouvier et al., 2020; Guitchounts et al.,

2020). Importantly, these signals can be attributed to the movement itself, the consequence of movement on the visual input or a combination thereof.

However, the computational role of head/eye movement signals in natural vision remains enigmatic, particularly since they are often studied under constrained conditions of head or gaze fixation, with sparse, artificial visual stimuli. Likewise, it is not clear whether there are common principles of active visual processing across rodents and primates, given differences in eye movement repertoires and other visual specializations. Indeed, it has recently been shown that the gain modulation of mouse V1 neurons by locomotion (Niell & Stryker, 2010) is present in primates, but differs in magnitude and sign (Liska et al., 2023). We therefore sought to characterize the neural representation of visual sampling movements in mice and marmosets, under natural conditions of free movement and viewing, respectively.

By measuring both head and eye movements together with neural activity in freely moving mice, we found that most V1 neurons do not respond to compensatory head/eye movements but respond to gaze-shifting movements with a variety of response dynamics, resulting in a temporal sequence of population activity following a gaze shift. Under head-fixation, when presented with sequentially flashed full-field stimuli, responses in many neurons closely matched the response to gaze shifts, suggesting that the pattern of gaze-shift transients is largely, though not entirely, explained by the response to rapid onset of new visual input. The time course of responses corresponded predictably with preference for increasing spatial frequency (SF), consistent with the coarse-to-fine theory of visual processing (Hegdé, 2008). Notably, we observed a similar sequence of processing following saccades in head-fixed marmosets that were allowed to freely gaze while viewing natural scenes. These results provide insight into neural signals previously observed around head and eye movements and reveal a computational principle for coarse-to-fine visual processing during active vision in complex natural scenes that is shared across rodents and primates.

3.4 RESULTS

3.4.1 V1 neurons respond specifically to gaze-shifting movements

First, we sought to characterize the responses of individual neurons in mouse V1 to large-amplitude head and eye movements. To monitor neural responses during free movement, we used a recording system (Figure 1a) consisting of a head-mounted camera to record the position of the right eye, a forward-facing wide-angle (120°) camera to capture the visual scene,

an inertial measurement unit (IMU) to record head orientation and acceleration, and a chronically implanted 64- or 128-channel linear silicon probe in V1 for single-unit electrophysiology (Figure 1b) (Michaiel et al., 2020, Parker et al., 2022). To assess the impact of the recording system on behavior, we calculated the mean head pitch and roll, which revealed minimal effects on posture (Extended Data Figure 1a,b).

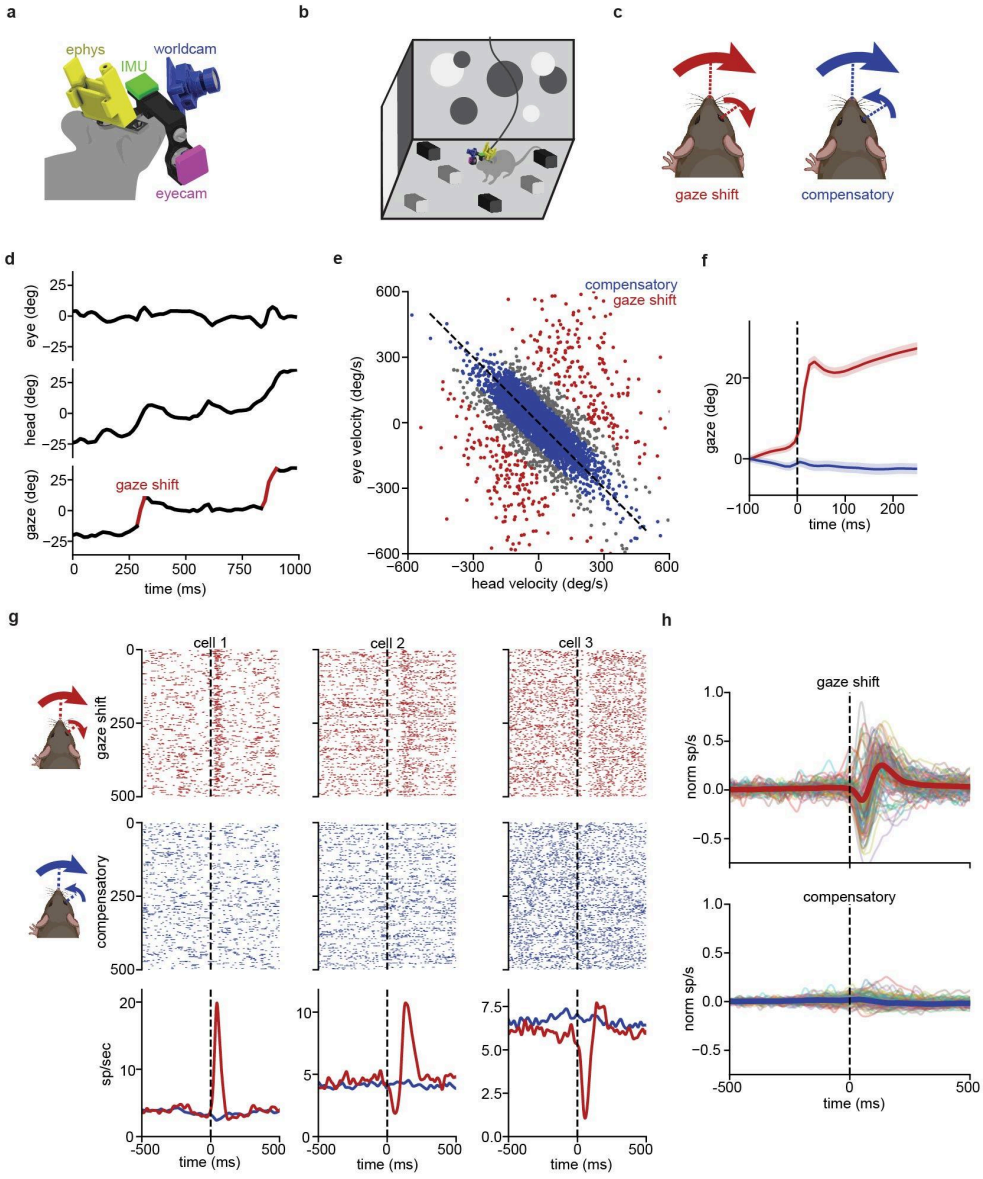


Figure 1. V1 neurons preferentially respond to gaze-shifting eye/head movements. a, Schematic of head-mounted recording system. **b**, Experimental paradigm, in which a mouse freely moves through a visually complex environment. **c**, Schematic of gaze-shifting and compensatory eye/head movements, in which the eyes move either with or against the movement of the head. Note that the arrows are not to scale, and that eye versus head movement amplitudes are not always equal in either condition. **d**, Mice change their horizontal eye position (top) to compensate for changing horizontal head position (middle), punctuated by rapid gaze-shifting eye movements.

Figure 1 continues on next page

Figure 1 continued
e, Scatter plot of eye and head velocities subsampled (25×) from an example 64-min recording, showing segregation of compensatory versus gaze-shifting eye movements. **f**, Mean ± s.e.m. position of the animal's gaze around gaze-shifting (red) and compensatory (blue) movements in the leftward head direction across all recordings. **g**, Response of three example cells to 500 randomly sampled gaze-shifting (top) or compensatory (middle) eye/head movements as a spike raster. Bottom, gaze-shifting and compensatory PETHs. **h**, Normalized PETHs of gaze-shifting (top) and compensatory (bottom) eye/head movements for 100 example units with a baseline firing rate >2 Hz, with median of all cells (n = 716) overlaid.

Mice coordinate their head and eye movements in a 'saccade-and-fixate' pattern of selecting and stabilizing the input to the retina that is observed across the animal kingdom23. When the mouse moves its head, the accompanying eye movements can be generally categorized into two functional types: compensatory or gaze-shifting (Michaiel et al., 2020; Meyer et al., 2020). Gaze refers to the direction the eyes are pointing relative to the visual scene—the sum of head direction and eye position within the head—which determines the retinal input. Gaze-shifting eye movements accompany head movements in the same direction to change the focus of the animal's gaze, and are typically abrupt, saccadic movements (Figure 1c, left). Conversely, compensatory eye movements counteract head movements of the opposite direction to keep gaze stabilized during self-motion (through vestibulo–ocular and other reflexes) and are typically smooth in trajectory (Figure 1c, right). Together, these two paired eye/head movements lead to a pattern of abrupt shifts of gaze (saccade) interspersed by periods of stabilized gaze (fixate) during continuous head movement. The vast majority of gaze-shifting movements in freely moving rodents are horizontal; vertical eye movements are generally compensatory (Miura & Scanziani, 2022; Michaiel et al., 2020, Meyer et al., 2020; King et al., 2023).

As observed previously, large-amplitude horizontal head movements in our recordings were usually accompanied by a horizontal eye movement (Michaiel et al., 2020; Meyer et al., 2020). The coordinated impact of gaze-shifting and compensatory movements is observed through gaze direction (sum of eye and head), showing a saccade-and-fixate pattern with periods of gaze stabilization (fixations) interspersed by abrupt shifts (saccades; Figure 1d). This head movement categorization can be visualized in a plot of eye versus head velocity (Figure 1e), where compensatory movements fall along the negative diagonal, since head and eyes move nearly equal but oppositely, while gaze shifts form a distinct off-diagonal pattern. Based on this, we defined gaze-shifting movements as timepoints with gaze velocity >240 deg s^{-1}, and compensatory as timepoints with gaze velocity <120 deg s^{-1}, where gaze velocity is defined as the sum of the eye and head velocities (Methods). In both cases, we required a head velocity

>60 deg s^{-1} to be considered a movement timepoint. During locomotion faster than 2 cm s^{-1}, the frequency of gaze-shifting movements doubled relative to slow and stationary periods (locomotion: 107 ± 6 per min; slow and stationary periods: 50 ± 3 per min; Extended Data Figure 1c, left). A similar increase was observed for compensatory movements (locomotion: 361 ± 11 per min; slow and stationary periods: 229 ± 15 per min; Extended Data Figure 1c, right). Thus, the difference in frequency of compensatory and gaze-shifting is not attributable to locomotion. Aligning gaze position to the time of gaze-shifting or compensatory movements reveals the characteristic step with gaze shifts and little change in gaze position for compensatory movements (Figure 1f).

We extracted single-unit neural activity in freely moving mice around the time of each gaze-shifting or compensatory eye/head movement (Figure 1g). This revealed striking responses following gaze-shifting movements, including positive, biphasic and negative responses. Notably, the same modulation is not generally found around compensatory movements. Across a population of 716 neurons in 9 mice, responses to gaze shifts were abundant (74.2%; 531 of 716) but compensatory responses were not (6.7%; 48 of 716; Figure 1h). Threshold values to separate compensatory and gaze-shifting movements were chosen to isolate these two movement types based on previous findings (Michaiel et al, 2020) (Extended Data Figure 1d); however, we also explored intermediate gaze shifts (eye/head movements that shift the gaze >120 deg s^{-1} and <240 deg s^{-1}; Extended Data Figure 1e–g). Neuronal responses were largest to gaze shifts, with relatively low-amplitude responses to these intermediate movements (Extended Data Figure 1h,i). Thus, V1 neurons predominantly respond to gaze-shifting eye/head movements, which produce abrupt changes in the retinal input, and do not generally respond to compensatory movements, which stabilize the retinal input. These findings support the notion that some component of previously observed responses in V1 to head movements may be attributed to an accompanying saccadic eye movement and the resulting change in visual input, rather than the head movement itself (Guitchounts et al., 2020).

3.4.2 Diversity and temporal sequence of gaze-shift responses

Neural responses to gaze shifts were diverse, both in sign of modulation and temporal dynamics. To characterize this diversity, we performed *k*-means clustering of units based on principal component analysis (PCA) of the normalized peri-event time histograms (PETHs) for gaze shifts (Extended Data Figure 2a), with the onset of the event defined as the onset of the

gaze shift, which will introduce a new visual stimulus onto the retina. For most units, the response for one direction of movement was stronger than the opposite direction. We determined their preferred direction as the direction of movement yielding the largest peak response and used this PETH for clustering. This clustering provides a means to describe the diversity within the population, but it is important to note that this is subdividing a continuous population. We selected a value of $k = 5$ clusters as it captured the diversity of responses while still providing distinct gaze-shift motifs in each cluster.

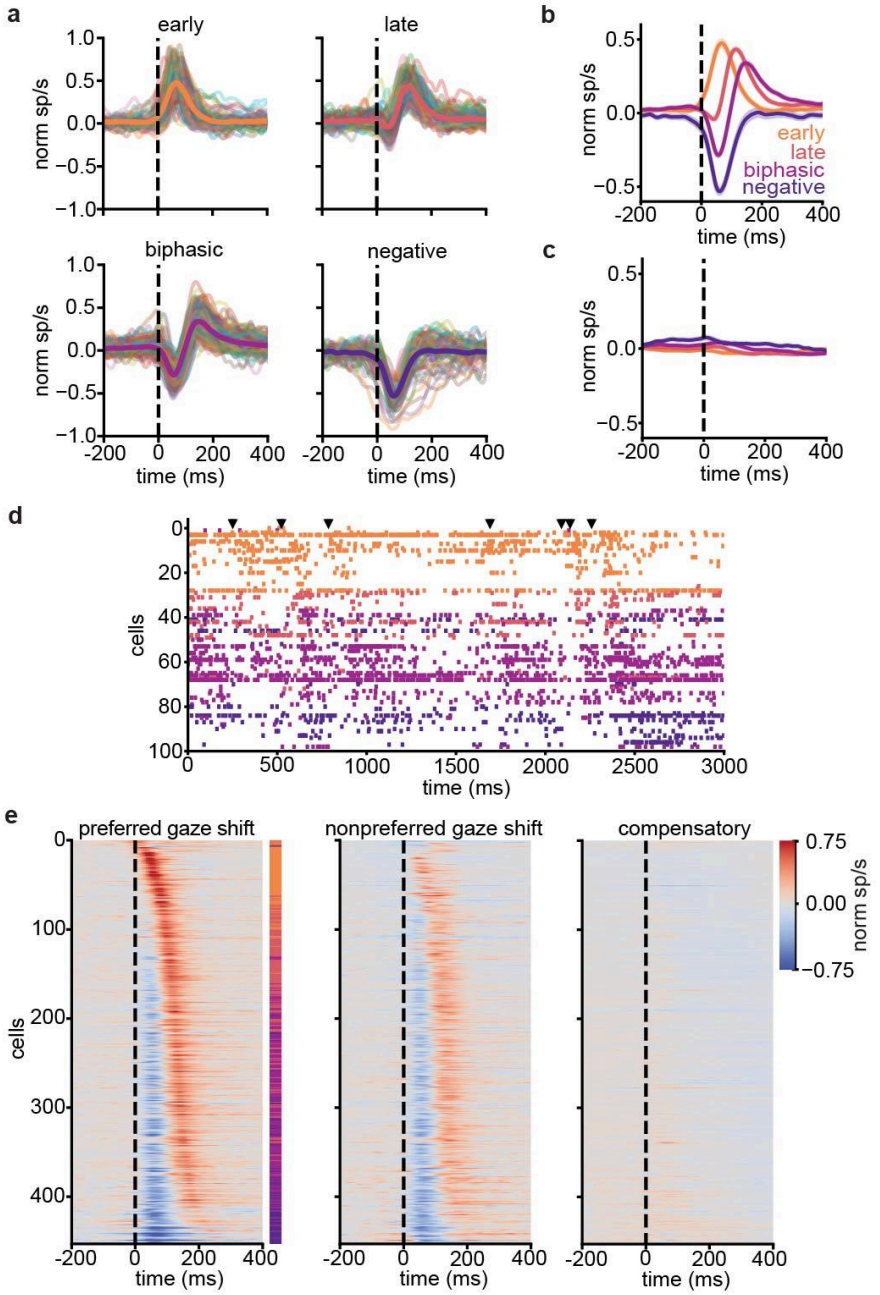


Figure 2. Diversity and temporal sequence of gaze-shift responses. a, Normalized PETHs for preferred direction of gaze shifts in responsive units, grouped by *k*-means clustering, with each cluster's mean overlaid. *n* = 9 mice, 716 units (early = 82, late = 135, biphasic = 170, negative = 66, unresponsive = 263). **b**, Mean normalized PETH in response to gaze-shifting eye/head movements, clustered by response type. **c**, Same as **b** for compensatory movements with cells clustered on their gaze-shift response. **d**, Example spike rasters from 3-s period of freely moving activity, showing all units responsive to gaze shifts (*n* = 98 of 128), color-coded by response type cluster. Gaze-shift onsets are indicated by black arrows above. **e**, Normalized PETHs for gaze-shifting movements in the preferred (left) or nonpreferred (middle) directions, and for compensatory movements (right). All responsive cells are shown, sorted along the *y* axis by response latency. The vertical colorbar to the side of the left panel shows each cell's assigned cluster using the colors in **a**.

This resulted in four responsive clusters (Figure 2a,b), which showed early positive (11.5% of population; Extended Data Figure 2b), late positive (18.9%), biphasic (23.7%) and negative (9.2%) responses to gaze shifts, as well as a largely unresponsive fifth cluster (36.7%). Mean firing rates and the relative fraction of putative excitatory/inhibitory cells based on spike waveform were similar for each cluster (Extended Data Figure 2d–f). Some units showed direction selectivity for leftward or rightward gaze shifts, especially in the early cluster (Extended Data Figure 2g). Intriguingly, while most clusters were uniformly distributed in cortical depth, neurons in the early positive population were more likely to be located in the deep layers (Extended Data Figure 2h). These clusters also corresponded in a predictable manner to tuning for angular velocity of horizontal head rotation (Extended Data Figure 2i). Early positive neurons were positively tuned for angular velocity, and direction-selective, while biphasic and negative clusters showed negative tuning, consistent with their response immediately after a gaze shift. On average, none of the clusters had a consistent response to compensatory movements (Figure 2c), though most neurons that responded to compensatory movements also responded to gaze shifts. However, biphasic/negative clusters had a greater number of compensatory responsive units than early/late clusters (Extended Data Figure 2j–l).

Notably, the mean traces for each cluster (Figure 2b) and the projection in principal component space (Extended Data Figure 2a) suggested a temporal sequence, from early positive through negative. Sorting units based on the latency of their peak positive response to gaze shifts revealed a continuous sequence of activation across the population following gaze shifts that was evident in single trial spike rasters (Figure 2d). Applying this sorting to the mean PETHs (Figure

2e) revealed a similar sequence, with positive peaks gradually becoming more delayed and an early negative component appearing (Figure 2e, left). We confirmed that this sequence was not an artifact of sorting through cross-validation, which showed that both temporal order and peak response latencies (Extended Data Figure 3a,b) were maintained when sorted on one half of the dataset (training set) and applied to the other half (test set; Pearson correlation coefficient, $r = 0.870$, $P = 2.51 \times 10{-}140$). The temporal sequence pattern was present, though weaker, in the nonpreferred direction (Figure 2e, middle), and was absent for compensatory movements (Figure 2e, right). Notably, the four responsive groups showed significantly different latencies to peak response (Extended Data Figure 2c; $P < 0.05$ for all comparisons with no effect of experimental session; $P = 0.220$, linear mixed effects model) and generally maintained their grouping in the continuous sequence of temporal responses, with the early positive group showing the earliest peak positive response, followed by the late positive group, the biphasic group and the negative group (Figure 2e, colorbar in left panel). Thus, the diverse responses to gaze shifts represent a temporal sequence across the population.

3.4.3 Temporal dynamics of gaze-shift response require visual input

The selectivity of cells to gaze-shifting head and eye movements, but not compensatory movements, led us to hypothesize that visual input may predominantly drive the observed responses in many neurons, as opposed to primarily motor efference or vestibular signals. To test the role of visual input in generating gaze-shift responses, we recorded neural activity during free movement both in the light and in complete darkness.

We clustered cells using the gaze-shift PETHs measured under normal light conditions (Figure 3a) and compared their responses in the light and dark (Figure 3b). The distinct cluster responses present in the light were absent in the dark, with most clusters (late positive, biphasic and negative) showing an average suppression of activity with an onset before the start of the gaze shift (Figure 3c; $n = 265$ units, 7 mice), consistent with previous reports of saccadic suppression (Ibbotson & Krekelberg, 2011). As in the light, neurons also showed little response to compensatory movements in the dark (Figure 3d). Sorting units based on their response latency in the light revealed that the precise temporal sequence of responses found in the light was also not maintained in the dark (Figure 3e). Together with the absence of modulation around compensatory movements, this suggests that the specific dynamic responses following a gaze shift rely on the corresponding change in visual input.

The mean response of the early positive group appeared flat in the dark, which resulted from a mix of units that showed suppression and a small fraction of units with strongly gaze direction-tuned responses that persisted in the dark (3.3% cells, $n = 9$ of 269; Extended Data Figure 4). The latter may represent a small population that receives a direct corollary discharge signal for saccades, as found in previous work (Miura & Scanziani, 2022). However, for the vast majority of neurons, the diversity of responses we observed around gaze shifts in the light was replaced by a suppression in the dark that lacked a clear temporal sequence.

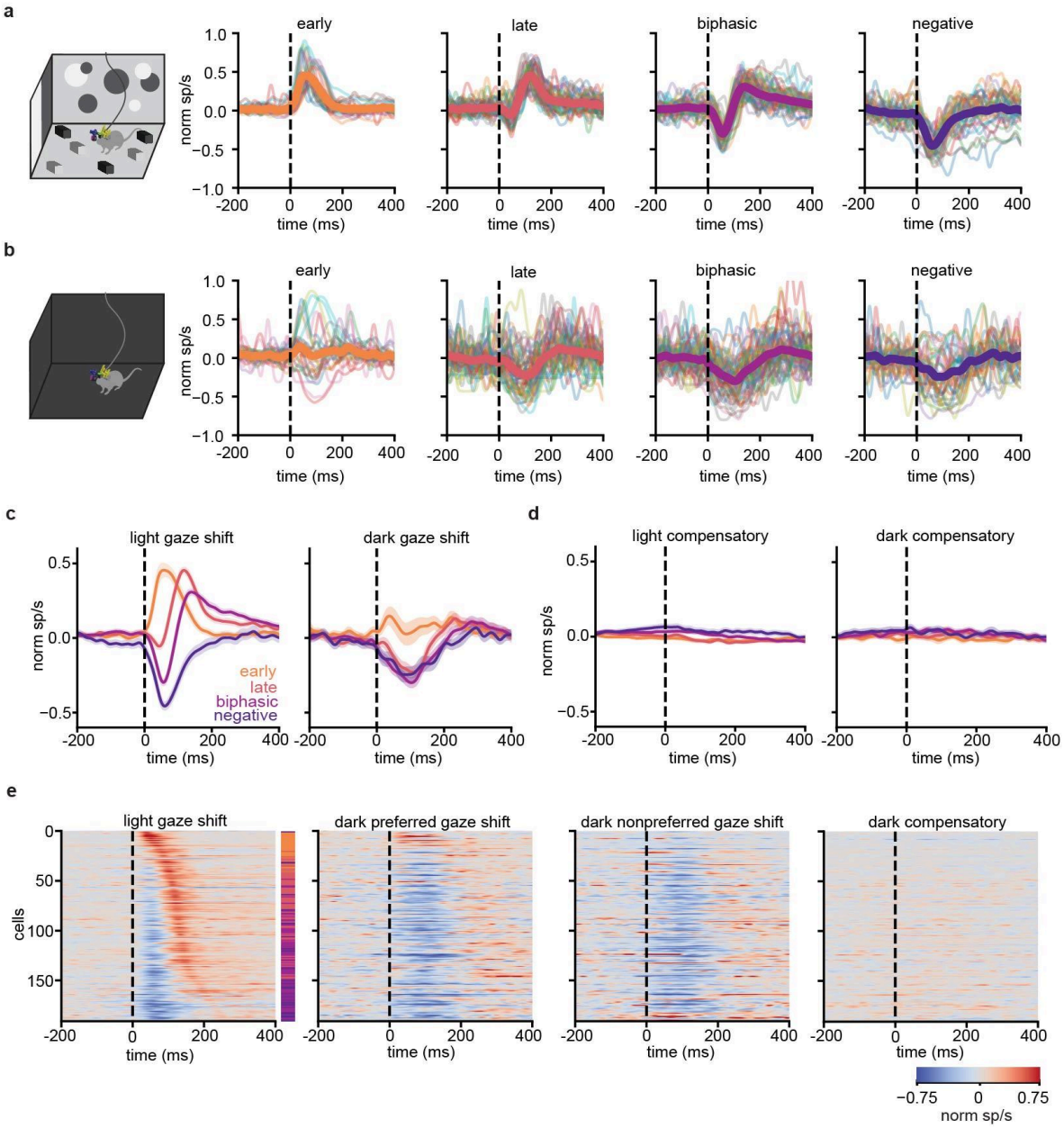


Figure 3. Temporal dynamics of gaze-shift responses depend on visual input. a, Normalized PETHs in light condition for the preferred direction of gaze shifts, clustered using *k*-means weights used for cells in Figure 2, $n = 7$ mice, 269 units (early = 27, late = 56, biphasic = 66, negative = 42, unresponsive = 78). **b**, Same as **a** for dark condition. Cells are grouped by their cluster from **a**. **c**, Mean gaze-shift responses of each cluster in light (left) and dark (right) conditions with standard error. **d**, Same as **c** for compensatory movements. **e**, Temporal sequence of responses to gaze shifts and compensatory movements in the light and dark. All temporal sequences are sorted by the cell's gaze-shift response latency for light condition (left).

3.4.4 Flashed stimulus responses resemble gaze-shift responses

We sought to further test whether the observed responses to gaze shifts were visually driven by presenting stimuli that mimic the rapid onset of a new visual image that results from a gaze shift in a complex visual environment. Head-fixed mice were shown continuously flashed stimuli, including reversing checkerboard and sparse noise, before recording the same units during free exploration of the arena, and we compared the stimulus onset responses with gaze-shift responses.

Time-aligned rasters and PETHs for gaze shifts and reversing checkerboard stimulus are shown for example units in Figure 4a. Notably, the visually evoked response to continuously flashed full-field stimuli resembled the neuron's response to gaze shifts, including positive, biphasic and negative responses. This similarity was evident at the population level as well, as the mean stimulus onset PETHs in response to the flashing checkerboard (Figure 4b, left), clustered based on each unit's gaze-shift response type, closely matched responses to gaze shifts (Figure 4c), as did mean responses to the shorter sparse noise stimulus (Figure 4b, right; Figure 4d). To quantify the similarity between gaze-shift responses and those to flashed stimuli, we computed the correlation coefficient between PETHs for gaze shifts and the flashed stimuli for each unit (0–250 ms following onset). This revealed that gaze-shifting and flashed stimulus PETHs were highly correlated for most units (Figure 4E,F; checkerboard median = 0.64, $n = 345$; sparse noise median = 0.68, $n = 263$).

Furthermore, the temporal sequence of checkerboard and sparse noise responses, when sorted based on each unit's peak latency from freely moving gaze shifts, closely matched the temporal sequence from gaze shifts (Figure 4g). Correspondingly, the latency of each unit's responses to head-fixed sequential stimuli was highly correlated with its latency for gaze-shift responses (Figure 4h,i). These results cannot be explained by the mouse performing gaze shifts with stimulus onset, as mice made only infrequent eye movements during head-fixation (sparse noise: 11.0 ± 10.1 saccades per min; checkerboard: 13.6 ± 15.8 saccades per min). Interestingly,

responses to head-fixed stimuli were slightly faster than responses to gaze shifts (checkerboard versus gaze shift: slope = 0.55, intercept = 0.041; sparse noise versus gaze shift: slope = 0.44, intercept = 0.045). Although it is not apparent why this might be, one possibility is that the appearance of a flashed stimulus on the screen is abrupt, whereas saccades extended over tens of milliseconds, so the timepoint when the input is stabilized following a gaze shift may be substantially delayed relative to the gaze-shift onset. Regardless, the close correlation between responses to gaze shifts during free movement and responses to sequentially flashed stimuli strongly suggests that the temporal sequence of gaze-shift responses primarily results from large, rapid changes to visual input.

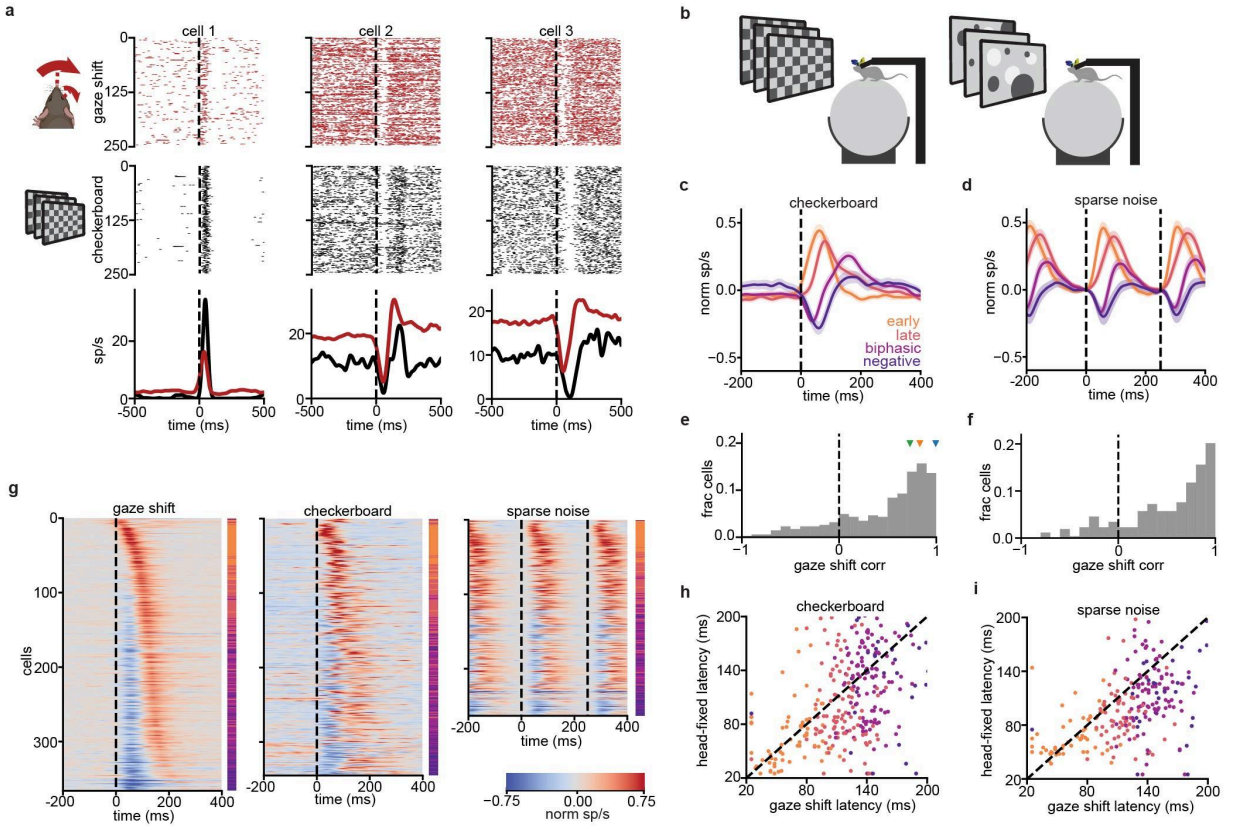


Figure 4. Head-fixed flashed stimulus responses resemble freely moving gaze-shift responses. a, Three example cells with positive (left), biphasic (middle) and negative (right) responses to freely moving gaze shifts. Spike rasters at the time of 250 randomly sampled gaze shifts during free movement (top) and 250 full-field reversals of a black and white checkerboard during a head-fixed recording of the same cell (middle). Bottom, gaze-shift (red) and checkerboard (black) PETHs for each cell. **b**, Schematic of head-fixed flashed checkerboard and sparse noise stimuli. **c**, Mean ± s.e.m. normalized response of cells to head-fixed flashed checkerboard, clustered on their responses to freely moving gaze shifts. $n = 9$ mice, 716 cells (checkerboard responsive = 472 of 716, early = 67, late = 100, biphasic = 123, negative = 55, gaze-shift unresponsive = 127). **d**, Same as **c** for sparse noise stimulus (sparse noise responsive = 333 of 716, early = 51, late = 69, biphasic = 99, negative = 44, gaze-shift unresponsive = 70). **e**, Histogram of Pearson correlation coefficient between gaze-shift PETH and flashed checkerboard, plotted as fraction of cells (frac) versus gaze shift correlation (corr), (median = 0.64, $n = 345$; arrows

Figure 4 continues on the next page

Figure 4 continued
indicate correlation, for example, cells 1 (blue, 0.99), 2 (orange, 0.83) and 3 (green, 0.73)) in **a**. **f**, Same as e for flashed sparse noise stimulus (median = 0.68, $n = 263$). **g**, Temporal sequence of gaze-shift responses (left, cells responsive to gaze shifts and either of the flashed head-fixed stimuli, $n = 366$), checkerboard (middle, cells responsive to both gaze shift and checkerboard, $n = 345$) and sparse noise (right, cells responsive to both gaze shift and sparse noise, $n = 263$) sorted on the latency of gaze-shift responses in freely moving conditions. **h**, Latency of positive peak in gaze-shifting responses versus head-fixed stimulus responses for checkerboard (Pearson correlation coefficient, $r = 0.487$, $P = 5.89 \times 10{-}22$, slope = 0.55, intercept = 0.041). **i**, Same as **h** for sparse noise stimulus (Pearson correlation coefficient, $r = 0.526$, $P = 4.19 \times 10{-}20$, slope = 0.44, intercept = 0.045).

3.4.5 Coarse-to-fine processing around gaze shifts

We next sought to determine whether there was a relationship between the visual tuning properties of neurons and their temporal responses to gaze shifts. We presented drifting sinusoidal gratings at three spatial (0.02, 0.08, 0.32 cycles per deg, or cpd) and two temporal (2, 8 cycles per s) frequencies to head-fixed mice before recording of the same units during free movement (Figure 5a,b). When organized by gaze-shift response type, the mean PETHs for gaze-shift clusters to drifting gratings reveal a large-amplitude onset transient in the positive clusters, and a smaller, slower rise time onset transient with a larger sustained response in the biphasic and negative clusters (Extended Data Figure 5a,b). Note that the temporal dynamics of these responses differ from those to gaze shifts and flashed stimuli due to the continuously moving drifting gratings.

To compare the relative SF preference across the response sequence, we performed four separate analyses. First, we calculated mean tuning curves for each gaze-shift cluster, which showed a clear progression shifting from lower to higher SFs across clusters (Figure 5c). The early positive group was biased toward lower and mid SFs, which gradually shifted to a peak at the highest SFs through the biphasic and negative groups (multivariate analysis of variance (ANOVA), cluster × SF tuning, $F = 9.20$, $P = 8.71 \times 10{-}10$). Calculating the preferred SF for units meeting a ≥2:1 selectivity index also revealed corresponding differences in SF preference across clusters (Extended Data Figure 5c), though this progression was somewhat weaker as this is a categorical metric and excludes a substantial number of neurons. We also computed an SF preference for each neuron based on the weighted average of its response to the three drifting gratings SFs (Figure 5d), which provides a metric to capture the central tendency of the individual neuron SF tuning curves. This metric also showed an increase from lower to higher SFs across the four groups. This relationship was not due to differences in sampling across recording sessions (linear mixed effects model, $P = 0.247$ grouped by subject). Finally, when

sorting units based on gaze-shift response latency, the median SF preference increases as a function of latency (Figure 5e). Together, these different analyses demonstrate that the temporal sequence of responses following gaze shifts corresponds to a progression from lower-SF-preferring neurons to higher-SF-preferring neurons. These results are consistent with the coarse-to-fine theory of visual processing, which posits that the spatial components of new visual input are processed sequentially from lower to higher SFs (Hegdé, 2008).

3.4.6 Distinct temporal frequency tuning for early/later responses

The relationship between gaze-shift response and weighted temporal frequency (TF) preference was qualitatively opposite to that of SF. The early positive group preferred the highest TF across the population, while the three later groups preferred relatively low TFs (Extended Data Figure 5d,e; multivariate ANOVA, TF × cluster $F = 21.45$, $P = 3.45 \times 10{-}13$). Correspondingly, the median TF preference decreased as a function of increasing response latency (Extended Data Figure 5f). This relationship was not due to differences in sampling across recording sessions (linear mixed effects model, $P = 0.659$ grouped by subject). Calculating the preferred TF of units meeting a ≥2:1 selectivity index, as an alternative to the weighted TF preference metric, revealed a shift in the bias of TF tuning from early to later responses (Extended Data Figure 5g). The observed relationship between TF preference and response latencies provides a potential explanation for the sequence we observe following gaze shifts: cells that prefer higher TF values respond quickly to gaze shifts, corresponding to the initial and abrupt change in visual input as a result of a shifting gaze. Cells that prefer lower TF values, conversely, respond later after the gaze shifts during gaze fixation and maintenance, when the stimulus is not rapidly changing.

A simple linear–nonlinear model provided support for this intuition (Supplementary Text and Extended Data Figure 6) by demonstrating that temporal kernels corresponding to low and high TF can qualitatively reproduce the different response profiles observed. This model also provides an explanation for how these response profiles differ from responses to standard visual stimuli that contain a gray interstimulus interval (ISI), since new stimuli arrive continuously without an opportunity for neurons to return to baseline.

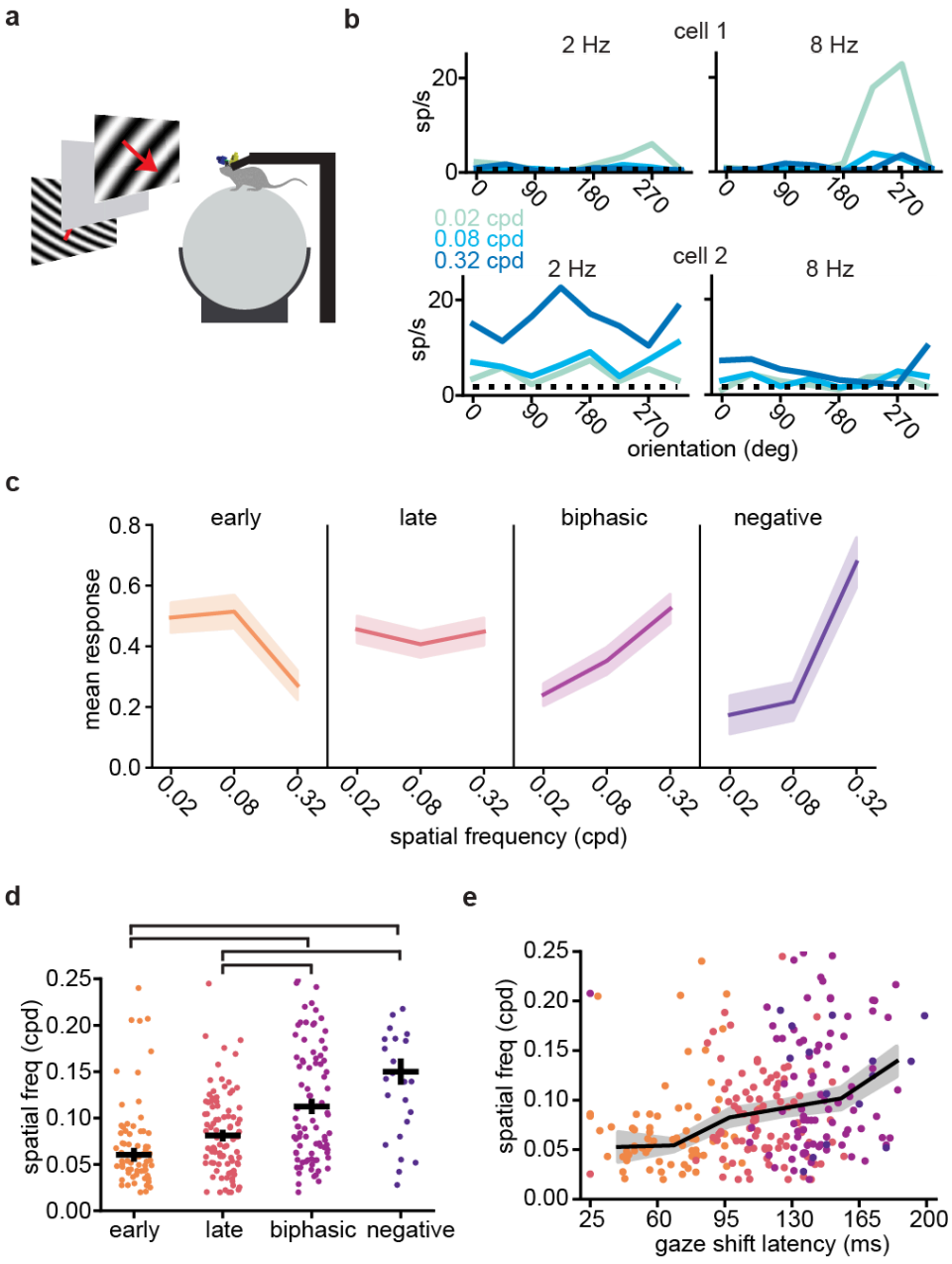


Figure 5. SF tuning demonstrates coarse-to-fine processing around gaze shifts. **a**, Schematic of head-fixed drifting gratings recordings. **b**, Orientation tuning curves for two example cells from head-fixed drifting gratings stimulus. Responses are shown for eight orientations of gratings (0 deg is the horizontal rightwards direction), three SFs, and low (left) and high (right) TFs. Spontaneous firing rates are shown as a horizontal dotted black line. **c**, Mean ± s.e.m. SF tuning curve for gratings-responsive cells in each cluster (multivariate two-way ANOVA, SF × cluster, $F = 9.20$, $P = 8.71 \times 10{-}10$; $n = 9$ mice, grating responsive cells = 384 of 716, early = 71, late = 96, biphasic = 98, negative = 29, unresponsive = 90). **d**, SF preference for gratings-responsive cells in each gaze-shift response cluster, calculated as a weighted mean of responses. Median and standard error are shown for each cluster. Bars above indicate statistical significance at $P < 0.05$ (linear mixed effects model, $n = 9$ mice, $n = 384$ cells; early versus late $P = 0.145$, early versus biphasic $P = 3.04 \times 10{-}7$, early versus negative $P = 5.49 \times 10{-}06$, late versus biphasic $P = 4.54 \times 10{-}5$, late versus negative $P = 3.96 \times 10{-}4$, biphasic versus negative $P = 0.620$). Of 384 cells, 20 had a weighted mean SF response greater than 0.25 cps and are not visible in the scatter plot but were included in the

Figure 5 continues on next page

Figure 5 continued
calculation of median. **e**, Weighted SF preference versus gaze-shift response latency, for all cells responsive to gratings. Running median ± s.e.m. for all cells is overlaid. The color of each point indicates the cluster from gaze-shift responses (Pearson correlation coefficient, $r = 0.286$, $P = 1.73 \times 10{-}6$).

3.4.7 Coarse-to-fine temporal sequence in freely gazing marmosets

Given the finding of a coarse-to-fine visual processing sequence following gaze shifts in freely moving mice, we wondered whether a similar temporal sequence occurs in V1 of nonhuman primates during active visual sampling. Mouse vision differs in several ways from nonhuman primates, but particularly striking is the lack of a fovea and high-acuity vision (Seabrook et al., 2017; Huberman & Niell, 2011). We therefore recorded the activity of single units in foveal V1 of head-fixed marmosets to determine the neural response to saccades across complex visual scenes.

Marmosets were first head-fixed and allowed to freely view natural image stimuli (Figure 6a). Aligning spike times to the onset of saccades revealed gaze-shift responses in marmosets that were similar to those in freely moving mice. Spike rasters and PETHs from example units show positive, biphasic and negative responses following saccades (Figure 6b). We clustered responsive units based on their normalized gaze-shift PETHs (k-means, $k = 4$ as the cluster of nonresponsive units was removed based on set criteria; Extended Data Figure 7a). This resulted in qualitatively similar groups to those identified in mice (Figure 6c,d and Extended Data Figure 7b,c). We also found a similar temporal sequence of responses across the marmoset V1 population after sorting by latency of peak positive response (Figure 6e), with the clusters maintaining their groupings across the temporal sequence in the same order as the mouse V1 units. We confirmed the sequence and latencies (Extended Data Figure 3c,d) by cross-validation.

To determine whether this temporal sequence represents a similar coarse-to-fine mechanism, we examined SF tuning based on rapidly flashed sinusoidal gratings, presented while the monkey performed saccades to forage for embedded Gabor patches (Ringach et al., 1997) (Figure 7a and Methods). This dataset was originally designed to study neural signals associated with visual foraging, but for the present study the same data provide an accurate estimate of SF preference for the same neurons as in the recordings during natural image viewing. Specifically, the sinusoidal gratings represent a Hartley subspace that allows rapid mapping of receptive fields (Ringach et al., 1997) as well as orientation and SF tuning, though not TF tuning.

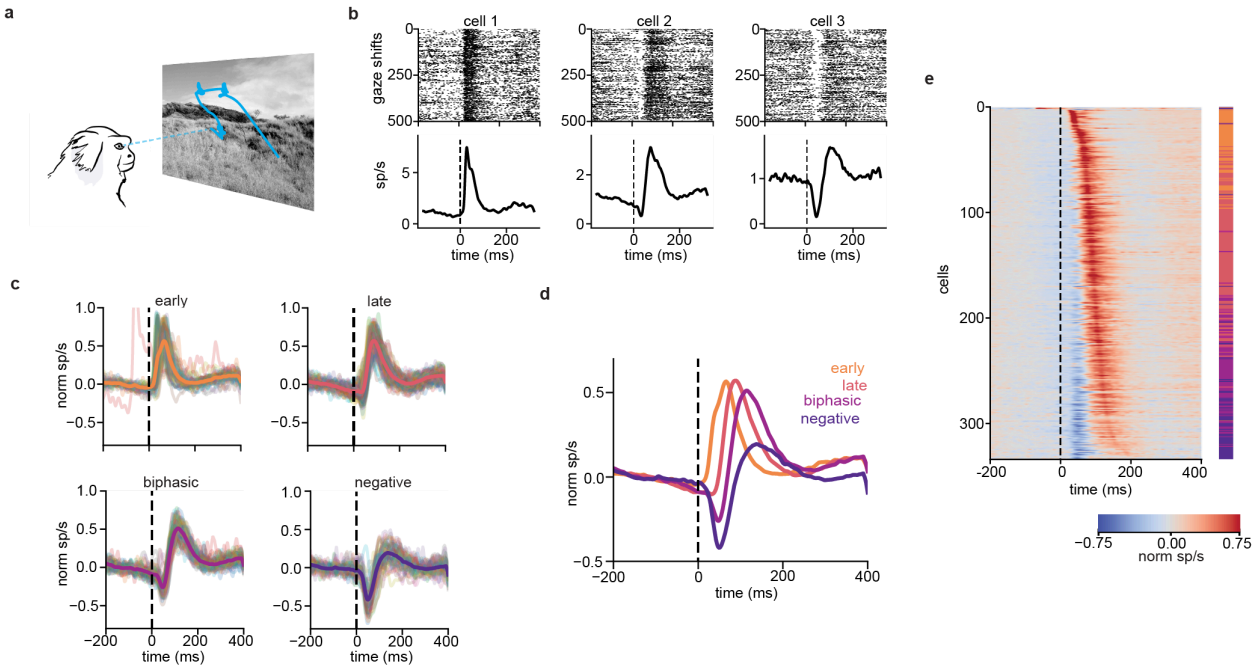


Figure 6. A similar temporal sequence of V1 saccade responses in freely gazing marmosets. a, Schematic of marmoset natural image free-viewing. **b**, Three example cells' spike rasters (top) and PETHs (bottom) for 500 randomly selected saccades. **c**, Normalized PETHs for saccades, grouped by *k*-means clustering, with each cluster's mean overlaid. $n = 2$ marmosets, 334 units (early = 64, late = 136, biphasic = 84, negative = 50). **d**, Mean normalized saccade response of cells in each cluster. **e**, Normalized saccade PETHs sorted along the y axis by positive peak latency. The vertical colorbar indicates the response cluster for each cell using the colormap from **c**.

Similar to the SF tuning of the gaze-shift response clusters in mice, marmoset neurons with earlier, positive saccade responses had tuning curves biased to lower SF preferences, centered around 2 cpd, and biphasic/negative were biased to higher SFs, centered around 4 cpd (Figure 7b; multivariate ANOVA, SF × cluster $F = 12.43$, $P = 1.82 \times 10{-}18$). We observed a similar progression across the groups based on the weighted mean SF preferences (Figure 7c). This relationship was not due to differences in sampling across recording sessions (linear mixed effects model, $P = 0.534$ grouped by subject). Likewise, we observed an increase in median weighted SF preference as a function of saccade response latency (Figure 7d). Finally, calculating the peak SF for each unit showed a shift from 2 to 4 cpd across the four clusters (Extended Data Figure 7d). We thus find that marmoset foveal V1 shows a temporal sequence of gaze-shift responses similar to that of freely moving mice, consistent with a coarse-to-fine pattern of visual processing shared across mouse and primate V1.

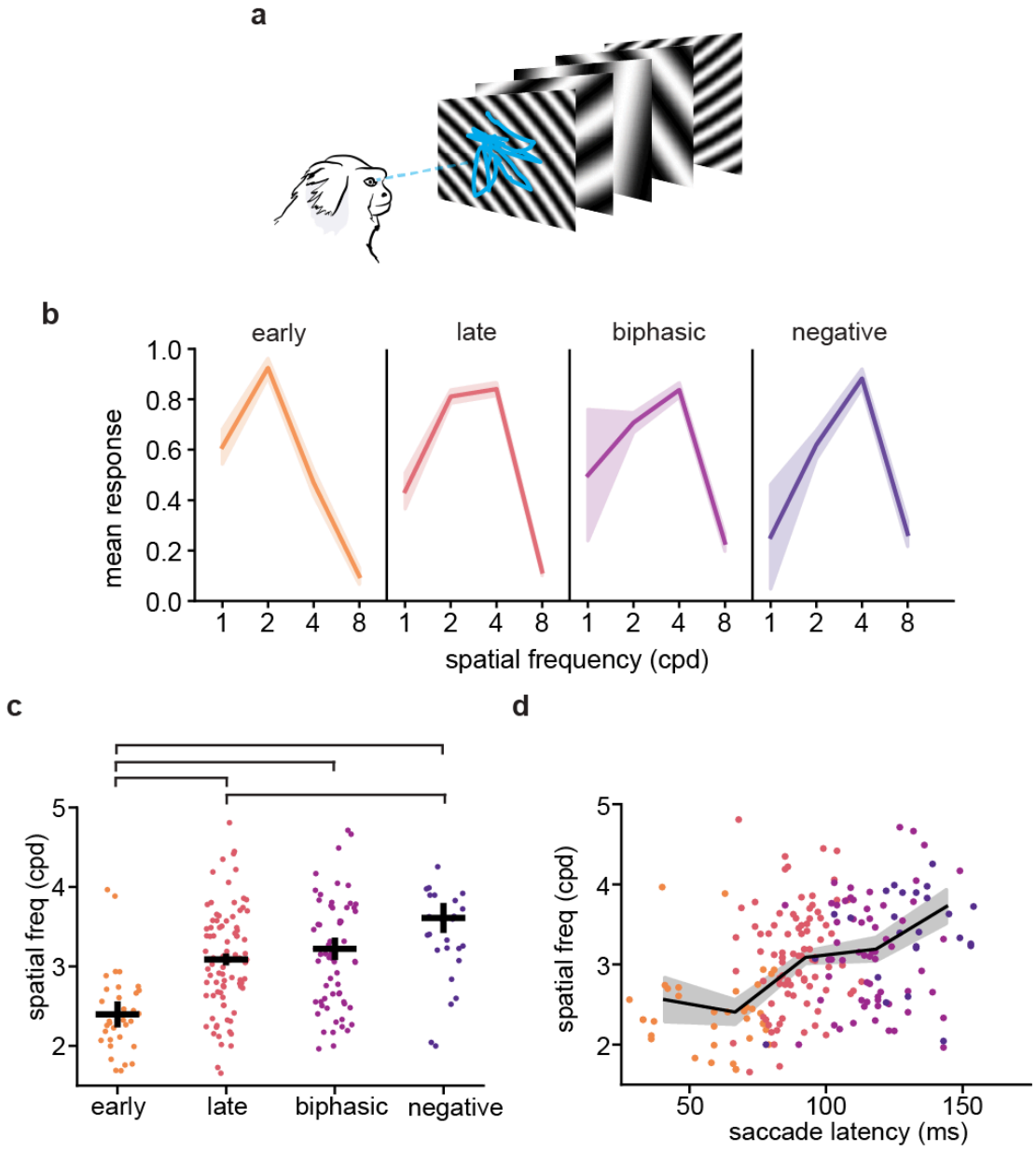


Figure 7. A coarse-to-fine temporal sequence in freely gazing marmosets. a, Schematic of grating stimulus. Sinusoidal grating of varying SF and orientation were presented rapidly (60 Hz) while monkeys performed saccades to forage for an embedded Gabor patch. **b**, Mean ± s.e.m. SF tuning curve for gratings-responsive cells in each cluster (multivariate two-way ANOVA, SF × cluster $F = 12.43$, $P = 1.82 \times 10^{-18}$; $n = 2$ marmosets, $n = 238$ cells, early = 41, late = 88, biphasic = 70, negative = 39). **c**, Weighted SF preference for each saccade response cluster, calculated as a weighted mean of responses for four presented SFs. Median and standard error are shown for each cluster. Only gratings-responsive cells are shown. Bars above indicate statistical significance at $P<0.05$ (linear mixed effects model, $n = 2$ marmosets, $n = 238$ cells; early versus late $P = 5.77 \times 10^{-3}$, early versus biphasic $P = 1.29 \times 10^{-4}$, early versus negative $P = 9.71 \times 10^{-6}$, late versus biphasic $P = 0.126$, late versus negative $P = 0.013$, biphasic versus negative $P = 0.253$). **d**, Weighted SF preference versus saccade response latency below 25ms ($n = 1$) or above 160 ms ($n = 12$) are not shown for plotting purposes but were included in statistical tests. $n = 334$ units (early = 40, late = 88, biphasic = 69, negative = 29, $r = 0.313$, $P = 8.54 \times 10^{-7}$).

3.5 DISCUSSION

We recorded the activity of V1 neurons as mice moved freely in an arena to determine how visual processing is influenced by eye and head movements at the cortical level. A wide scope of previous work has shown that activity of cortical neurons is correlated with movement, with V1 neurons showing responses to locomotion, head movements and eye movements. We found a diverse array of responses to gaze-shifting, but not compensatory, head and eye movements in V1 that together formed a temporal sequence of activity. When the animal moved in an arena lacking visual input (complete darkness), the continuous temporal sequence of responses was not maintained, and most neurons showed weak suppression of activity around the time of a gaze shift. Further, we found V1 neurons mimic their response to gaze shifts when shown full-field sequentially flashed stimuli under head-fixation despite no corresponding head or eye movements. These findings suggest that a substantial component of the responses to gaze-shifting head and eye movements results from changes to visual input as opposed to motor efference or nonvisual sensory reafferent signals. We did observe evidence of nonvisual signals from responses in darkness—a small set of 'early positive' responses (3.3% of the population) and a common slow suppression signal. These signals could be combined, either additively or nonlinearly, with visual input in the light. A previous study reported a higher incidence of V1 neurons with apparent motor efference responses resulting from saccades (~35% direction-selective saccade responses (Miura & Scanziani, 2022)). The discrepancy in the prevalence of these signals could be due to differences in analysis, such as threshold and normalization methods, or differences in stimulus conditions (recordings performed with the mouse in front of a gray screen versus in darkness here). A recent study reported a 9.5% incidence of saccade-responsive neurons in V1 (King et al., 2023).

Additionally, we found that neurons responding earlier to a gaze shift preferred higher TFs and lower SFs, and neurons responding later following a gaze shift preferred lower TFs and higher SFs. These findings are consistent with the coarse-to-fine model of visual processing, in which the visual system responds to coarse aspects of the visual scene before finer details. The temporal sequence that we observe is locked to the 'saccade-and-fixate' pattern that occurs during natural vision, during which rapid saccadic eye movements are interspersed with slow fixational movements in a pattern that formats the visual input arriving on the retina1. Our findings therefore reflect the fact that natural visual input is neither continuous nor organized

into temporally distinct trials as in highly controlled experimental paradigms. This impacts both the statistics of the visual input and the corresponding neural dynamics, underscoring the need to study visual processing under naturalistic conditions.

These findings are remarkably consistent across mice and marmosets, both in terms of the specific temporal response profiles that we observe, and the correspondence with SF preference. This is notable since although many basic aspects of cortical visual processing are shared across mammals, one might expect that aspects related to active vision could differ. Primates have a pronounced foveal specialization accompanied by an expanded repertoire of eye movements, including voluntary targeted saccades independent of head movements. On the other hand, mice have less pronounced retinal specializations (Bleckert et al., 2014; van Beest et al., 2021) and eye movements are largely coupled to head movement and image stabilization, with little evidence of independently targeted saccades (Michaiel et al., 2020; Meyer et al., 2020; Zahler et al., 2021; Holmgren et al., 2021). Nonetheless, when analyzed in a shared framework of gaze shifts versus fixations, a similar computational principle emerges—a temporal sequence of response dynamics corresponding to coarse-to-fine processing.

3.5.1 Temporal dynamics and gaze-shift responses

Our results provide insight into the nature of neural signals measured around head and eye movements in natural conditions. In one of the few studies of free-viewing of natural scenes in primates, a mix of positive and negative responses was observed following saccades (Gallant & Van Essen, 1998). Likewise, neural responses around microsaccades, when the fovea moved within the stimulus, revealed positive, negative and biphasic responses (Leopold & Logothetis, 1998). Our results suggest that these diverse responses may correspond to the different temporal tuning of V1 neurons, which determines their response when the retinal input shifts abruptly following saccades or microsaccades. Likewise, a recent study observed modulation of V1 activity around head-orienting movements in rats (Guitchounts et al., 2020), with a diversity of positive and negative tuning in light and dark, with no systematic relationship. By measuring eye movements along with head movements, we found that most neurons primarily respond to the instants within an ongoing head movement that are associated with a saccadic movement of the eyes, which result in a gaze shift. Furthermore, our results explain the logic behind the diversity of head movement responses, based on the temporal sequence relative to gaze shifts and the SF/TF preferences measured with grating stimuli—from early positive neurons with lower

SF/higher TF preference to biphasic and then negative neurons with higher SF/lower TF preference (Figures 5 and 7 and Extended Data Figure 2c). This time course also predicts responses in the dark, as all but the earliest positive neurons shift to suppression in the dark (Figure 3).

It is striking that we see relatively little response to head movements that are not accompanied by a gaze shift, especially since a subset of V1 neurons show vestibular coding during head rotation in head-fixed subjects (Vélez-Fort et al., 2018; Bouvier et al., 2020). However, in comparing head-fixed versus freely moving conditions, it is important to note that studies of coding in the vestibular nuclei have shown that the vestibular reafference resulting from active, volitional rotation (as in free movement) is canceled out, while in passive rotation (as in head-fixed conditions) it is not (Angelaki & Cullen, 2008; Duffy & Burchfiel, 1975). Together, these data suggest that V1 may similarly differentiate between active and passive head movements, potentially by inheriting these signals from the vestibular nuclei.

Surprisingly, a large fraction of neurons had a negative response to gaze shifts in the light and to flashed stimuli. Negative responses in visual neurons around the time of a saccade have long been hypothesized to be suppression of visual perception while performing an eye movement across the visual scene. We find that such responses are common and represent part of a continuum from early positive responses through biphasic and finally negative responses. Notably, the neurons with the largest negative responses, and correspondingly latest positive responses, also preferred the lowest TFs. This suggests that the negative responses, at least in part, are a result of the neurons not responding to the high TF transient associated with the saccade, leading to a drop in activity, then slowly responding to the sustained visual input during fixation. This hypothesis is supported by a simple linear–nonlinear model (Supplementary Text), which recapitulates the diverse temporal dynamics, including negative responses, based purely on different temporal kernels for a neuron's response that correspond to higher and lower TF, or transient and sustained responses.

Notably, these temporal dynamics arise as a result of the abrupt arrival of new visual input, during both gaze shifts and continuously flashed stimuli, standing in stark contrast with the standard visual neuroscience approach of presenting a gray 'blank' during an ISI. ISIs allow neurons to return to their baseline firing rate between visual stimuli to isolate responses with their corresponding stimulus; responses to stimuli that lack an ISI can therefore depend on the

cell's response to the previous input. Under natural conditions, visual stimulation is a continuous input without ISIs. Our finding that responses to rapidly shifting full-field stimuli recapitulate gaze-shift responses supports the notion that neural dynamics depend on not only what stimuli are presented, but how they are presented. Accordingly, it is important to understand the visual stimulus as it arrives at the retina under natural conditions (Gibson, 1979).

For most neurons the typical temporal response sequence around gaze shifts was replaced by suppression in the dark. This result is consistent with physiological saccadic suppression via corollary discharge, which has been reported in other species, including nonhuman primates and cats (Duffy & Burchfiel, 1975; Adey & Noda, 1973; Toyama et al., 1984). It is possible that this suppressive signal is present in the light but is masked by the strong visually driven responses that are absent in the dark. It may also contribute to the negative component observed in gaze shifts in the light, amplifying the visually driven component. Alternatively, previous work in mice has also shown that responses to head rotation in V1 are luminance-dependent in a cell-type-specific manner (Bouvier et al., 2020), and thus the sign of this signal may switch between the light and dark conditions. We also observed a small population of neurons (3.3%) that showed positive responses to gaze-shifting movements in both the light and dark, which may represent a direct motor corollary discharge signal. Similar results have been reported in head-fixed mice in the absence of visual input (Miura & Scanziani, 2022), though we observed a much smaller fraction of units with such responses

3.5.2 Coarse-to-fine processing during active vision

There is substantial evidence from human psychophysics that supports coarse-to-fine processing, with lower SF components being recognized before higher SF components (Navon, 1977; Oliva & Schyns, 1997; Petras et al., 2019). At the neural level, single-unit recordings in head-fixed nonhuman primates and rodents have shown a progression of tuning from lower to higher SF following stimulus onset (Bredfeldt & Ringach, 2022; Skyberg et al., 2022; Mazer et al., 2002; Purushothaman et al., 2014; Vreysen et al., 2012). Here, we demonstrate that this coarse-to-fine visual mechanism is engaged during active vision in both mice and marmosets, resulting in a temporal sequence of responses to visual input with gaze shifts.

There are multiple proposed reasons for why the visual system may utilize a coarse-to-fine order of processing(Hegdé, 2008). According to Marr's primal sketch theory, initial visual processing may be devoted to setting a basic layout of the visual scene, followed by

later activity to establish greater dimensionality in the percept (Marr, 2010). Alternately, the order of responses may represent a matching to the typical input following a saccade and subsequent fixation, with lower SFs predominant immediately after a saccade due to the 1/f frequency distribution of natural scenes, followed by a whitening transformation of the distribution due to small movements during the fixations (Rucci et al., 2007). Additionally, recent work in rodents suggests that coarse-to-fine processing may support efficient coding of natural scenes (Skyberg et al., 2022), and modeling shows that a coarse-to-fine framework improves binocular disparity computation in stereovision (Lopes et al., 2015).

The underlying mechanisms that generate coarse-to-fine processing in V1 are unknown. One possibility is the pooling in V1 of magnocellular and parvocellular type inputs from the lateral geniculate nucleus, which respond in both a temporally and spatially specific manner (Allen & Freeman, 2006). In primates, the magnocellular pathway conveys high TF/low SF information (‘fast’ and ‘coarse’), while the parvocellular pathway shows low TF/high SF tuning (‘slow’ and ‘fine’) (Allen & Freeman, 2006). Although there isn’t a clear delineation of magnocellular and parvocellular pathways in mice, a similar relationship between transient/sustained response types and corresponding SF tuning has been observed in mouse lateral geniculate nucleus (Piscopo et al., 2013), and previous work supports a relationship between TF and SF tuning in mouse V1 (Gao et al., 2010). Alternatively, a rapid refinement of SF tuning could result from recurrent processing in cortex, for example, from active suppression of low SF responses (Bredfeldt & Ringach, 2002). Genetic tools available in mice may allow experiments to test these potential circuit mechanisms.

Together, these results present a dynamic view of processing during active vision, where each gaze shift initiates a temporal sequence supporting coarse-to-fine processing. This processing mode is shared across rodents and primates and is dramatically different from the regime typically probed in head- and gaze-fixed vision. This emphasizes the importance of accounting for the actual dynamics of visual input that occur during behavior under natural conditions, to understand the computations performed by the visual system.

3.6 METHODS

3.6.1 Mouse methods

3.6.1.1 Animals

All procedures were conducted in accordance with the guidelines of the National Institutes of Health and were approved by the University of Oregon Institutional Animal Care and Use Committee. The 3–8-month-old adult mice (Mus musculus, C57BL/6J, Jackson Laboratories and bred in-house) were kept on a 12-h light/dark cycle. In total, eight male and six female mice were used for this study. Mice were housed with sibling cagemates until electrophysiology implant when they were singly housed. Humidity was between 40% and 60% and temperature was 21 ± 1 °C. Data collection and analysis were not performed blind to the conditions of the experiments. No animals or data points were excluded from analysis. Data collection was not randomized as there was only a single experimental group.

3.6.1.2 Surgery and habituation

Mice were initially implanted with a titanium headplate over V1 to allow for head-fixation and attachment of head-mounted experimental hardware. After 3 days of recovery, widefield imaging (Wekselblatt et al., 2016)was performed to help target the electrophysiology implant to the approximate center of the left monocular V1. A miniature connector (Mill-Max 853-93-100-10-001000) was secured to the headplate to allow repeated, reversible attachment of a camera arm, eye/world cameras and IMU (Michaiel et al., 2020; Parker et al., 2022). To simulate the weight of the real electrophysiology drive for habituation, a ‘dummy’ electrophysiology drive was glued to the headplate. Animals were handled by the experimenter for several days before surgical procedures, and subsequently habituated (~45 min total) to the spherical treadmill and freely moving arena with hardware tethering attached for several days before experiments.

The electrophysiology implant was performed once animals moved comfortably in the arena. A craniotomy was performed over V1, and a linear silicon probe (64 or 128 channels, Diagnostic Biochips P64-3 or P128-6) mounted in a custom three-dimensionally printed drive (Yuta Senzai, UCSF) was lowered into the brain using a stereotax to an approximate tip depth of 750 µm from the pial surface. The surface of the craniotomy was coated in artificial dura (Dow DOWSIL 3-4680) and the drive was secured to the headplate using light-curable dental acrylic (Unifast LC). A second craniotomy was performed above the left frontal cortex, and a reference

wire was inserted into the brain. The opening was coated with a small amount of sterile ophthalmic ointment before the wire was glued in place with cyanoacrylate. Animals recovered overnight and experiments began the following day.

3.6.1.3 Hardware and recording

The camera arm was oriented approximately 90 deg to the right of the nose and included an eye-facing camera (iSecurity101 1000TVL NTSC, 30 frames per s (fps) interlaced), an infrared light-emitting diode to illuminate the eye (Chanzon, 3-mm diameter, 940-nm wavelength), a wide-angle camera oriented toward the mouse's point of view (BETAFPV C01, 30 fps interlaced) and an inertial measurement unit acquiring three-axis gyroscope and accelerometer signals (Rosco Technologies; acquired at 30 kHz, downsampled to 300 Hz and interpolated to camera data). Fine-gauge wire (Cooner, 36 AWG, no. CZ1174CLR) connected the IMU to its acquisition box, and each of the cameras to a USB video capture device (Pinnacle Dazzle or StarTech USB3HDCAP). A top-down camera (FLIR Blackfly USB3, 60 fps) recorded the mouse in the arena.

The electrophysiology headstage (built into the silicon probe package) was connected to an Open Ephys acquisition system via an ultra-thin cable (Intan no. C3216). Electrophysiology data were acquired at 30 kHz and bandpass filtered between 0.01 Hz and 7.5 kHz. We first used the Open Ephys GUI (https://github.com/open-ephys/plugin-GUI) to assess the quality of the electrophysiology data, then recordings were performed in Bonsai (Lopes et al., 2015) using custom workflows (https://github.com/nielllab/FreelyMovingEphys). System timestamps were collected for all hardware devices and later used to align data streams through interpolation.

During experiments, animals were first head-fixed on a spherical treadmill to permit measurement of visual tuning properties using traditional methods, then were transferred to an arena where they could freely explore. Recording duration was approximately 40 min head-fixed, and 1 h freely moving. For head-fixed experiments, a 70-cm monitor (BenQ GW2780) was placed approximately 27.5 cm from the mouse's right eye, and visual stimuli were presented using Psychtoolbox-3 (Kleiner et al., 2007)). Head-fixed stimuli were also recorded using the head-mounted world camera. First, we presented 15 min of a band-limited Gaussian noise stimulus57 (SF spectrum 0.05 cpd to 0.12 cpd, flat TF spectrum with a low-pass cutoff at 4 Hz) to confirm V1 targeting based on spike-triggered average receptive fields. We presented a contrast-reversing square-wave checkerboard stimulus for 2 min with an SF of 0.04 cpd and TF

of 0.5 Hz. Flashed sparse noise consisted of full- and minimum-luminance circular spots on a gray background played for 5 min. Spots were 2, 4, 8, 16 and 32 deg in diameter and presented so that each size made up an equal fraction of the area on the screen, totaling 15% on average. Each stimulus frame was presented for 250 ms and immediately followed the previous frame with no ISI (Piscopo et al., 2013). Drifting sinusoidal gratings were presented at eight evenly spaced directions of motion for three SFs (0.02, 0.08, 0.32 cpd) and two TFs (2, 8 cps) for 15 min, with a 1-s stimulus duration and 0.5-s gray ISI, and stimulus conditions randomly interleaved (total of 12 presentations per unique stimulus).

For experiments comparing stimuli flashed with and without an ISI, responses were recorded for two stimulus sets: a sparse noise stimulus presented for 500 ms followed by a gray screen ISI lasting 500 ms, shown for 20 min (~1,200 stimulus frames). Next, the same stimulus flashed every 500 ms with no ISI interleaved between frames, shown for 10 min total (~1,200 stimulus frames).

The arena was approximately 48-cm long by 37-cm wide by 30-cm high. A 61-cm monitor (BenQ GW2480) covered one wall of the arena, while the other three walls were clear acrylic covering custom wallpaper including black and white high- and low-SF gratings and white noise. During recordings, the arena's stimulus monitor played a sparse moving noise stimulus similar to the flashed sparse noise stimulus used during head-fixed recordings52. For moving sparse noise, full- and minimum-luminance spots were 4, 8 and 16 deg in diameter. Each spot was assigned to move in one of eight evenly spaced directions and at one of five speeds (10, 20, 40, 80, 160 deg s^{-1}). Spots appeared on the appropriate edge of the screen and moved across until they disappeared on the far edge. The floor was a gray silicone mat (Gartful) and was densely covered with black and white Lego bricks to provide three-dimensional visual contrast. Small pieces of tortilla chips (Juanita's) were lightly scattered around the arena to encourage foraging during the recording; however, animals were not water or food restricted.

To eliminate all possible light within the arena during dark recordings, the entire behavioral enclosure was sealed in light-blocking material (Thorlabs BK5), all potential light sources within the enclosure were removed and all external light sources were turned off. Animals were first recorded in the dark (~20 min), then the arena lights and wall stimulus monitor were turned on (~20 min). As a result of the dark conditions, the pupil dilated beyond the margins of the eyelids, which made eye tracking infeasible. Thus, before the experiment, one

drop of 2% pilocarpine HCl ophthalmic solution was applied to the animal's right eye to constrict the pupil to a size similar to that seen in the light. Once the pupil was restricted enough for tracking in the dark (~3 min) the animal was moved into the dark arena for recording until the effects of the pilocarpine wore off (~20 min), at which time the light recording began.

3.6.1.4 Data preprocessing

The raw electrophysiology data from head-fixed and freely moving recordings acquired in the same session were concatenated into a single file for spike sorting, to allow us to confidently track single units across the entire experiment. Common-mode noise was removed by subtracting the median across all channels at each timepoint. Spike sorting was performed using Kilosort 2.5 (https://github.com/MouseLand/Kilosort), and isolated single units were then selected using Phy (Phy 2.0 beta 5; https://github.com/cortex-lab/phy) based on parameters including contamination (<10%), firing rate (mean > 0.5 Hz across entire recording), waveform shape and autocorrelogram. After data were spike sorted in Kilosort and curated in Phy, the resulting spike data were then split back out into individual recordings for analysis.

To group units into narrow- and broad-spiking putative cell types, we first normalized the mean spike waveform of each unit using the equation below, where W is the waveform and Wb is the mean of the first 200 µs. We performed k-means clustering on the normalized waveforms ($k = 2$). The resulting clusters closely resembled the segregation of narrow- and broad-spiking units in previous studies based on explicit waveform criteria (Niell & Stryker, 2008).

$$W_{norm} = (W - W_b) / (-\min(W - W_b))$$

Laminar depth was calculated from the multi-unit local field potential (LFP) of head-fixed neural data. The mean-centered LFP of each channel for a given shank of the linear silicon probe was bandpass filtered (1 to 300 Hz) and the power for the channel was calculated. The position of the maximum LFP power along the ordered sites of each shank was considered the center of layer 5 of cortex (Senzai et al., 2019).

$$\text{Power}_{\text{channel}} = \text{sqrt}(\text{mean}(\text{LFP}^2_{\text{ch}}))$$

We extracted pupil position from the eye camera data (Michaiel et al., 2020). Briefly, eye camera data were first deinterlaced to achieve 60-fps video, then eight points around the pupil

were tracked with DeepLabCut (Mathis et al., 2018). We then fit an ellipse to these eight points and computed pupil position in terms of angular rotation. The head-mounted world camera was deinterlaced to achieve 60-fps video and distortions from the camera lens were corrected using OpenCV. The mouse's position in the arena was tracked with DeepLabCut using top-down camera frames. We defined running as periods when the animal's neck point had a velocity >2 cm s^{-1} and a stationary state as periods when the neck point had a velocity <2 cm s^{-1}.

Horizontal head rotation velocity was extracted from the IMU, converted to deg s^{-1} and interpolated to eyecam frame timestamps. We defined leftward and rightward directions as the direction of head movement from the animal's perspective. For example, a movement of the right eye in the nasal direction would be a leftward movement. To select eye/head movement onsets, we first determined if the head velocity was >60 deg s^{-1} in the leftward or rightward direction. If a sufficiently large head movement was made, we then separated it into either gaze-shifting or compensatory movements using gaze velocity, where gaze is defined as the sum of horizontal eye and head velocities (Michaiel et al., 2020). If there was a high gaze velocity (>240 deg s^{-1}) concurrent with the head movement, the movement was considered gaze-shifting. Those that resulted in low gaze velocity (<120 deg s^{-1}) were considered compensatory. Eye/head movements that resulted in intermediate gaze velocities (>120 deg s^{-1} and <240 deg s^{-1}) were excluded from our analysis to avoid contamination between the two categories of movements (Extended Data Figure 1e–i). For eye/head movements spanning multiple eye camera frames, only the first timepoint, representing the onset of the movement, was used to calculate the onset of the PETH. Compensatory movements which occurred 250 ms before or after a gaze-shifting movement were excluded to avoid contamination by gaze shifts. For head-fixed recordings, only high eye velocity (>240 deg s^{-1}) was used to identify saccades, since the head was immobilized.

To maintain internal consistency with eye cameras and obviate the need for a separate video stimulus synchronization signal, we determined stimulus onsets in head-fixed recordings directly from the head-mounted cameras. For the reversing checkerboard stimulus that updated every 500 ms, we identified frame transitions from the head-mounted camera video based on *k*-means clustering ($k = 2$) of the video into the two separate contrasts, and selected transitions between the clusters. For the flashed sparse noise stimulus that updated every 250 ms, we determined the timestamps of stimulus onset based on the root mean square pixel-wise change in the image, which showed clear peaks at stimulus transition. For drifting sinusoidal gratings, we

determined direction of motion by computing optic flow from the worldcam video, SF based on the mean gradient magnitude from paired Sobel operators and TF from the mean Fourier transform of each pixel over time.

3.6.1.5 Analysis of neural responses

All analyses were performed in Python (v.3.8, python.org). Neural responses were calculated as a PETH from electrophysiology spike times using kernel density estimation with a gaussian kernel with bandwidth (standard deviation) of 10 ms, sampled at 1-ms intervals. The PETH included neural activity from −1,000 ms to 1,000 ms around the event onset for all stimuli/events (unless noted otherwise). A modulation index was calculated for PETHs as the peak of the firing rate (R) from the event onset at 0 ms until 250 ms after the event, minus the baseline before the event (R_b), which was calculated as the mean of R from −1,000 ms to 200 ms around the event onset (unless noted otherwise).

$$\text{mod} = \max(\text{abs}(R - R_b))$$

For the drifting gratings stimulus, the PETH was calculated similarly to other stimuli, but from −1,500 ms before until 1,500 ms after stimulus onset to capture the full stimulus and ISI interval.

The modulation index for gaze-shift responses in the leftward and rightward directions were compared to determine a preferred direction. A direction selectivity index (DSI) was calculated as the difference between the preferred and nonpreferred maximum modulations divided by their sum.

$$\text{DSI} = (\text{mod}_{\text{pref}} - \text{mod}_{\text{nonpref}})/(\text{mod}_{\text{pref}} + \text{mod}_{\text{nonpref}})$$

To normalize PETHs, we subtracted the baseline Rb from the PETH (R) to give an evoked firing rate, and then divided by the maximum of R during a response window (R_{rw}) within a response range of −250 ms before to 250 ms after the event. The denominator includes responses up to 250 ms before the event to ensure that spurious increases in firing rate that are not time-locked to an event were not amplified by the normalization (this was extremely rare and only occurred in unresponsive cells).

$R_{norm} = (R - R_b)/(\max(R_{rw}))$

When normalizing the PETH of compensatory eye/head movements, we used that neuron's R_{rw} from gaze shifts in the preferred left/right direction so that all eye/head responses were normalized relative to the neuron's best response. Because the head-fixed sparse noise stimulus presentation was of short duration (250 ms), we used the spike rate at the time of the stimulus onset, 0 ms, as the baseline firing rate (R_b) for normalization to avoid including responses to previously presented flashed stimuli.

Neurons were considered responsive to a stimulus or eye/head movement onset if they changed their firing rate by at least 10% and 1 spike per second (sp s^{-1}) in the 250 ms following the onset of the event. For the gratings stimulus, suppressed-by-contrast cells were not considered responsive, and removed by requiring that gratings PETHs have a peak firing rate >0.5 normalized sp s^{-1} during the 1 s of stimulus presentation.

The latency of a neuron's peak response was calculated as the timepoint of the maximum firing rate of its normalized PETH in the period between 25 ms to 250 ms after the onset of the event. For temporal sequence plots, normalized PETHs were sorted by the latency of their peak for the preferred direction of gaze-shifting eye/head movements. Peak latency was not calculated for units unresponsive to gaze shifts. Peak latency sorting was cross-validated by randomly assigning gaze-shift events into a train set and a test set, calculating a PETH for each half of the data, and sorting the test set by the peak times of the training set (Extended Data Figure 3).

SF and TF preferences were based on the evoked firing rate for each stimulus condition, computed as the mean rate from 25 to 1,000 ms following stimulus onset, minus the mean baseline rate in the 500 ms before stimulus onset. The mean SF and TF tuning curves for all cells within gaze-shift clusters were calculated using each cell's mean evoked firing rate for each SF/TF normalized by the maximum response to its best SF/TF.

Each cell's weighted TF response, W_{TF}, was determined using the mean evoked responses, R, for each of the two presented TFs and at the cell's optimal SF and preferred orientation.

$W_{TF} = (2 \times R_{low} + 8 \times R_{high})/(R_{low} + R_{high})$

Likewise, weighted SF response was calculated at the cell's weighted TF response and preferred orientation using the mean evoked responses for each of the three presented SFs.

$$W_{SF} = (0.02 \times R_{low} + 0.08 \times R_{mid} + 0.32 \times R_{high})/(R_{low} + R_{high})$$

As an alternate metric for TF and SF bias in each gaze-shift cluster, we calculated the fraction of cells with a >2:1 preference for one of the presented SFs over the other two SFs (Extended Data Figure 5c) and for one TF over the other (Extended Data Figure 5g).

3.6.1.6 Response clusters

To cluster units with similar gaze-shift responses, we first performed PCA on the normalized gaze-shift responses in each unit's preferred direction in the period from −50 to 300 ms relative to the gaze-shift onset. Units that did not meet the above spike rate criteria had their input for PCA replaced with an array of zeros to exclude their nonresponsive PETHs from clustering, while seeding the formation of an unresponsive group of units for clustering. We used the first four principal components, which cumulatively explained 95% of the variance, to perform *k*-means clustering ($k = 5$), resulting in four clusters of responsive units and one cluster of unresponsive units. To maintain consistency, the same PC weights and *k*-means model were applied to cluster the smaller number of units that were acquired subsequently in light/dark experiments.

To identify units that respond positively in dark conditions, we selected units that were positively modulated in the dark by the onset of a gaze shift by >1 sp s^{-1} and 10% of their normalized spike rate in the 100 ms following the gaze-shift onset.

For presentation of sparse noise with and without an ISI (Extended Data Figure 6a,b), neurons were grouped into response types by applying the PCA and *k*-means models fit with separate gaze-shift PETH data to the PETHs of the continuously presented flashed sparse noise. Cells were grouped for the ISI condition using the results of their clustering of the continuous stimulus. To avoid the inclusion of suppressed-by-contrast cells, we removed ISI-condition responses that failed to increase their firing rate by 0.1 normalized sp s^{-1} in the first 100 ms after stimulus onset.

3.6.1.7 Linear-nonlinear model

To model the temporal dynamics of responses to a time-varying stimulus with abrupt transitions, such as gaze shifts or flashed stimuli, we implemented a rate-based linear–nonlinear

model (Dayan & Abbott, 2005). The stimulus was described by a set of discrete levels *s(t)*, and a neuron's temporal kernel as *k(τ)*, each on a discrete timebase with d*t* = 25 ms. The neuronal response at each time *t* was computed as the convolution of the temporal kernel with the stimulus, followed by an exponential nonlinearity.

$$r(t) = exp(\sum_{r=1}^{n} s(t - \tau)k(\tau))$$

For each 500-ms interval the stimulus was a randomly selected integer in the range 1–8, either with or without a subsequent 500-ms ISI with a value of 0. For the temporal kernel, we use (a) the positive phase of a sine function, for a unimodal positive kernel; (b) the full cycle of a sine function, for a biphasic kernel; or (c) the full cycle of a sine function with the negative phase multiplied by 0.25, for an intermediate between a and b. Each kernel had a total duration of 250 ms.

3.6.2 Marmoset methods

3.6.2.1 Animals and surgical procedures

All surgical and experimental procedures were approved by the Institutional Animal Care and Use Committee at the University of Rochester in accordance with the US National Institutes of Health guidelines. Two adult male common marmosets (Calithrix jacchus) aged 2 and 7 yr were used in this study. Housing temperature was 27 °C and humidity 50%. Data collection and analysis were not performed blind to the conditions of the experiments. No animals or data points were excluded from analysis. Data collection was not randomized as there was only a single experimental group. Marmosets underwent an initial surgery to implant a headpost to stabilize their head during behavioral sessions (Mitchell et al., 2014; Nummela et al., 2017). A second surgery was performed to implant a recording chamber with a 3 × 3-mm^2 craniotomy over foveal and parafoveal V1. The exposed dura was sealed under a thin layer (<1 mm) of a silicone elastomer (Spitler & Gothard, 2008).

3.6.2.2 Hardware and recordings

Electrophysiological recordings were performed using 2 × 32-channel silicon electrode arrays (http://www.neuronexus.com). Probes included two sharpened tip shanks of 50-µm width spaced 200 µm apart, each containing 32 channels separated by 35 µm. In one animal we used a

semichronic microdrive (EDDS Microdrive system, https://microprobes.com) to place electrodes in cortex for 1–2 weeks over which we made 3–6 recordings. In the second animal we used a custom microdrive (https://marmolab.bcs.rochester.edu/resources/) to place and remove electrodes daily. Arrays were lowered slowly through silastic into cortex using a thumb screw.

Data were amplified and digitized at 30 kHz with Intan headstages using the Open Ephys GUI (https://github.com/open-ephys/plugin-GUI). The wideband signal was high-pass filtered by the headstage at 0.1 Hz, preprocessed by common-average referencing across all channels and then high-pass filtered at 300 Hz. The resulting traces were spike sorted using Kilosort2. Outputs from the spike sorting algorithms were manually labeled using the Phy GUI (https://github.com/kwikteam/phy). Any units that were either physiologically implausible based on the lack of a waveform with a trough followed by a peak or with an interspike interval distribution with more than 1% of the spikes under 1 ms were excluded from analyses.

3.6.2.3 Eye-tracking and saccade detection

Gaze position was monitored using infrared eye-tracking methods (Yates et al., 2023). Briefly, the first and fourth Purkinje images (P1 and P4) were visualized using a collimated infrared light source and tracked at 593 fps to estimate the two-dimensional eye angle. The eye tracker was manually calibrated to adjust the offset and gain (horizontal and vertical) by showing marmoset monkeys small windowed face images at different screen positions to obtain their fixation (Mitchell et al., 2014; Nummela et al., 2017). Saccadic eye movements were identified automatically using a combination of velocity and acceleration thresholds (Cloherty et al., 2020).

3.6.2.4 Head-fixed visual stimuli and free-viewing tasks

Visual stimuli were presented on a Propixx Projector (Vpixx) with a linear gamma table using Psychtoolbox-3 (Kleiner et al., 2007) and Matlab (MathWorks). Stimulus and physiology clocks were aligned using a Datapixx (Vpixx) (Eastman & Huk, 2012). MarmoV5 stimulus code is available (https://github.com/jcbyts/MarmoV5).

Marmosets performed free-viewing and visual-foraging tasks. For natural image stimuli, the marmosets were allowed to free-view a screen containing a full-field (±15 visual degrees) grayscale natural image for 10 s and then received a single drop of juice (5–10 µl of marshmallow water). On the foraging trials the marmosets viewed a random sequence of flashed full-field gratings (60- or 120-Hz frame update of varying orientation and SF gratings) as they searched for a small Gabor target (Yates et al., 2023). A drop of juice reward was given for

fixating within 2 deg of the small target for more than 100 ms after which time it was repositioned to another random location within 5 visual degrees of the screen center. During foraging the full-field flashed gratings were presented at 25% contrast and were drawn from a polar grid of 8 equally spaced orientations and a grid of 4 log-spaced range of SFs from either 1 to 8 cpd or 2 to 16 cpd (8 × 4 stimulus space). On each video frame either a single grating was presented or with 50% chance a blank gray background was shown.

3.6.2.5 Estimation of orientation and SF tuning

To measure orientation and SF preferences of individual neurons we analyzed the response to flashed full-field sine-wave gratings described above for the foraging task. A linear estimate of the receptive field can be recovered from this stimulus using subspace reverse correlation by correlating the firing rate during each video frame with the stimulus history (Ringach et al., 1997). We first computed a temporal kernel of the response to gratings as compared with blanks. The average temporal response at latencies from −50 to 200 ms (−3 to 12 video frames at 60 Hz) was computed by time-locking on video frames including gratings and averaging. It was compared against the baseline firing rate, which was estimated by instead time-locking to blank video frames. If a peak was identified in the temporal response to the grating stimulus between 20 and 80 ms that was at least 8 s.d. above the baseline, then the neuron was included in analyses for tuning. We marked a temporal window that included the frame of the peak response along with the adjacent frames next to the peak frame if they were significant by at least five standard deviations above the baseline.

To estimate orientation and SF tuning, the mean response was computed from the identified temporal window for each grating stimulus. The joint tuning for orientation and SF was defined by the mean responses in the 8 × 4 grid of stimuli sampled from which we averaged across orientation or SF to get the marginal tuning curves. To determine if neurons exhibited a selective response, we computed a tuning index by taking the standard deviation of the values in the tuning curve divided by the mean. Neurons with weak tuning, having an index less than 0.2 in both dimensions, were not included for subsequent analyses that aimed to correlate tuning preference with saccade modulation. The weighted preference was computed using a weighted mean from the tuning curve instead of taking the peak at a discrete value, similar to the analysis for mouse data described above. To emphasize the peak in the weighted mean, since a wider range of SFs was presented than for mouse data, we used a weighting function computed by

subtracting the mean of the tuning curve with the resulting negative values floored at zero and then squaring the remaining nonzero values.

3.6.2.6 Neural responses to eye movements

Neural responses around eye movements were computed from −200 ms before to 400 ms after the saccadic onset. The PETH was computed using kernel density estimation with a Gaussian kernel with half-bandwidth of 10 ms, sampled at 1-ms intervals. To normalize PETHs (R), we computed the baseline firing rate (R_b) from −100 to 0 ms before saccade onset and normalized the PETH as

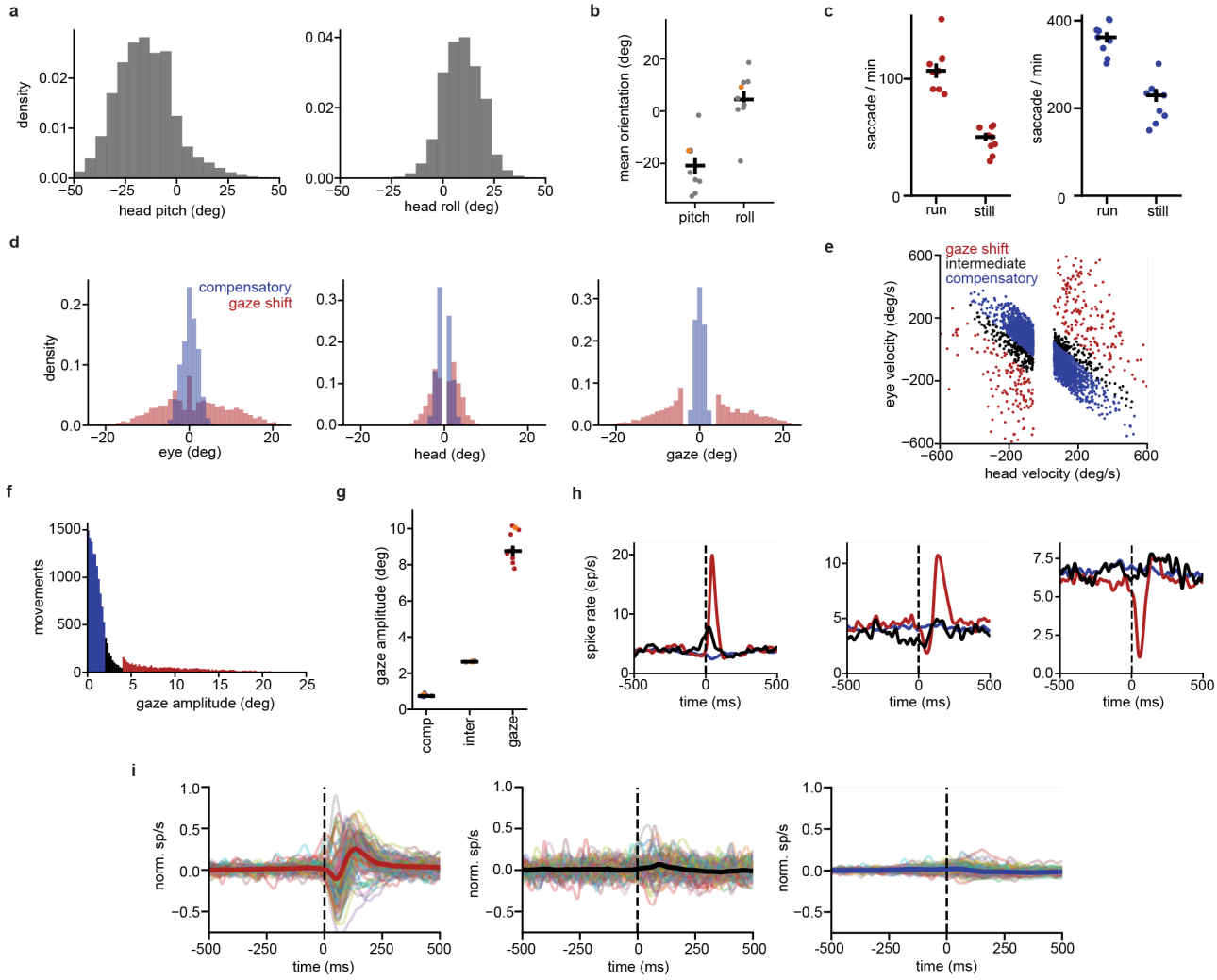
$$R_{norm} = (R - R_0)/\max(R).$$

The maximum of R was computed from 0 to 400 ms. The PCA was computed on the normalized PETH from −200 to 400 ms. The six PCs explained 95% of the variance and were k-means clustered ($k = 4$) to group units into saccade response types.

3.6.3 Statistical analysis

Statistical tests across cell types were performed using ANOVA with correction for multiple comparisons or a linear mixed effects model. Correlations between latencies and response properties were tested using Spearman's correlation coefficient. Bar plots show mean and s.e.m. unless otherwise noted. No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those reported in previous publications (Liska et al., 2023; Senzai & Scanziani, 2022). Data distribution was assumed to be normal, but this was not formally tested.

3.7 EXTENDED DATA

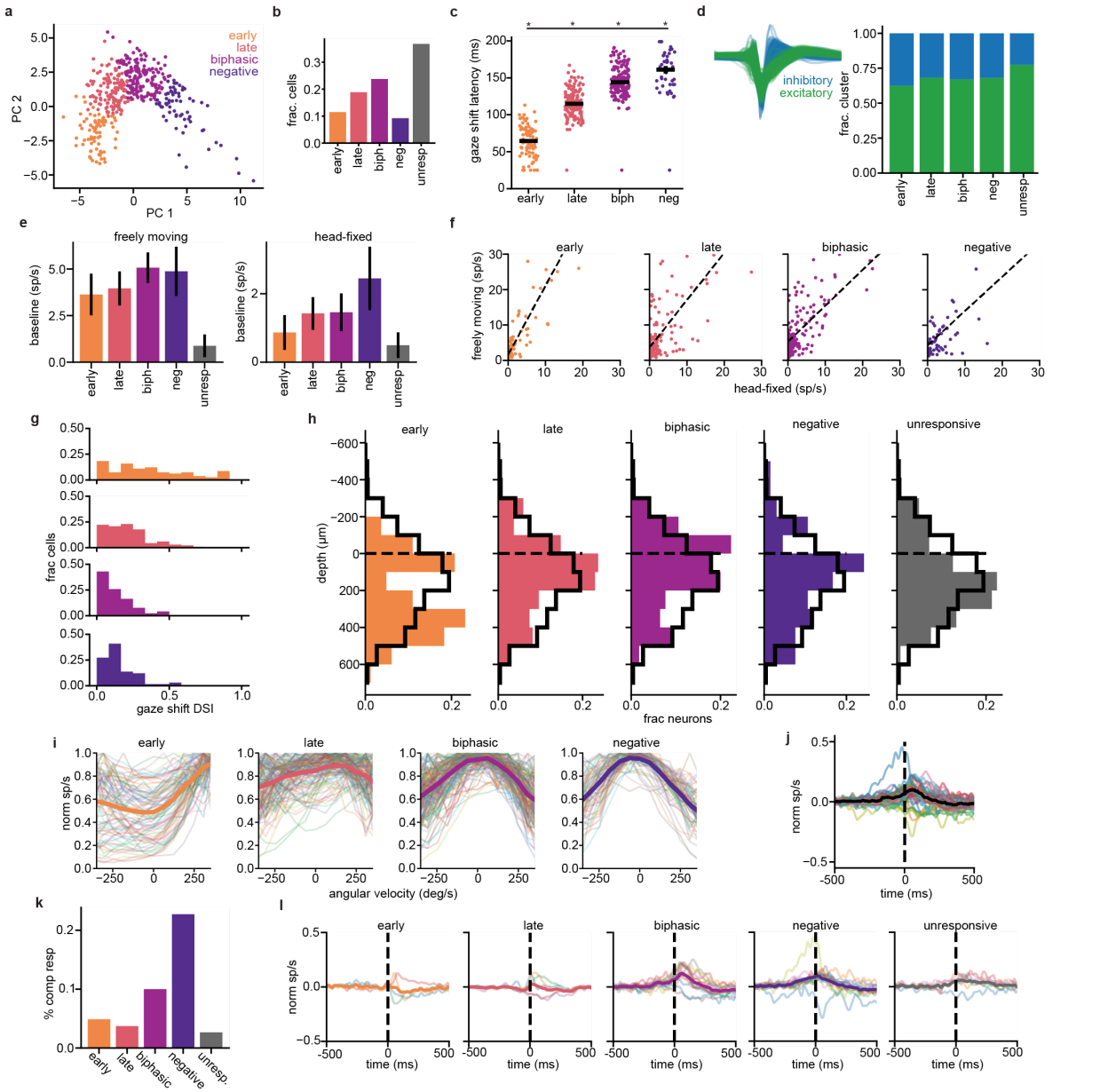


Extended Data Figure 1. Characterization of free movement. **a**, Mean head pitch and roll during free motion for one example recording. Pitch mean = −15.1 ± 0.02 deg; Roll mean = 9.2 ± 0.01 deg. **b**, Mean head pitch and roll, indicating the center point during free movement. Each point is a mouse ($n = 9$ mice). Black bars indicate mean and standard error (pitch: −20.8 ± 3.1 deg; roll: 4.5 ± 3.3 deg). Recording from a shown in orange. **c**, Rate of gaze-shifting (left; run median = 106 ± 6 saccades/min; still median = 50 ± 3 saccades/min) and compensatory (right; run median = 362 ± 11 saccades/min; still median = 229 ± 15 saccades/min) movements during periods of locomotion greater than 2 cm/s measured from the top camera ('run') compared to periods of slower locomotion, fine motion, and/or stationary periods less than 2 cm/s ('still'). Each point is a mouse ($n = 9$ mice). Black bars indicate median and standard error. **d**, Amplitude of position change for eye (left), head (middle) and gaze (right; defined as eye + head) during gaze-shifting and compensatory eye/head movements at the onset of the movement for the example recording used in **a**. **e**, Scatter plot of eye and head velocities subsampled (25x) from the example recording used in **a**, showing compensatory, gaze-shifting, and intermediate movements, the latter of which are

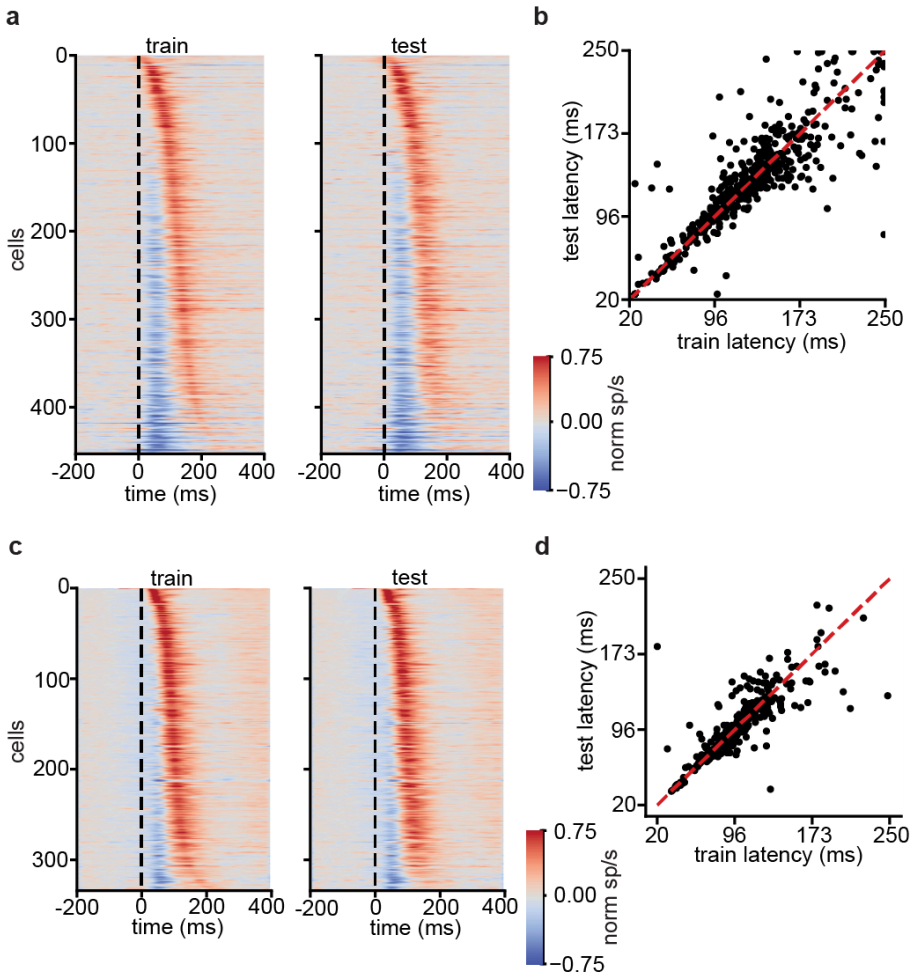
Extended Data Figure 1 continues on next page

Extended Data Figure 1 continued

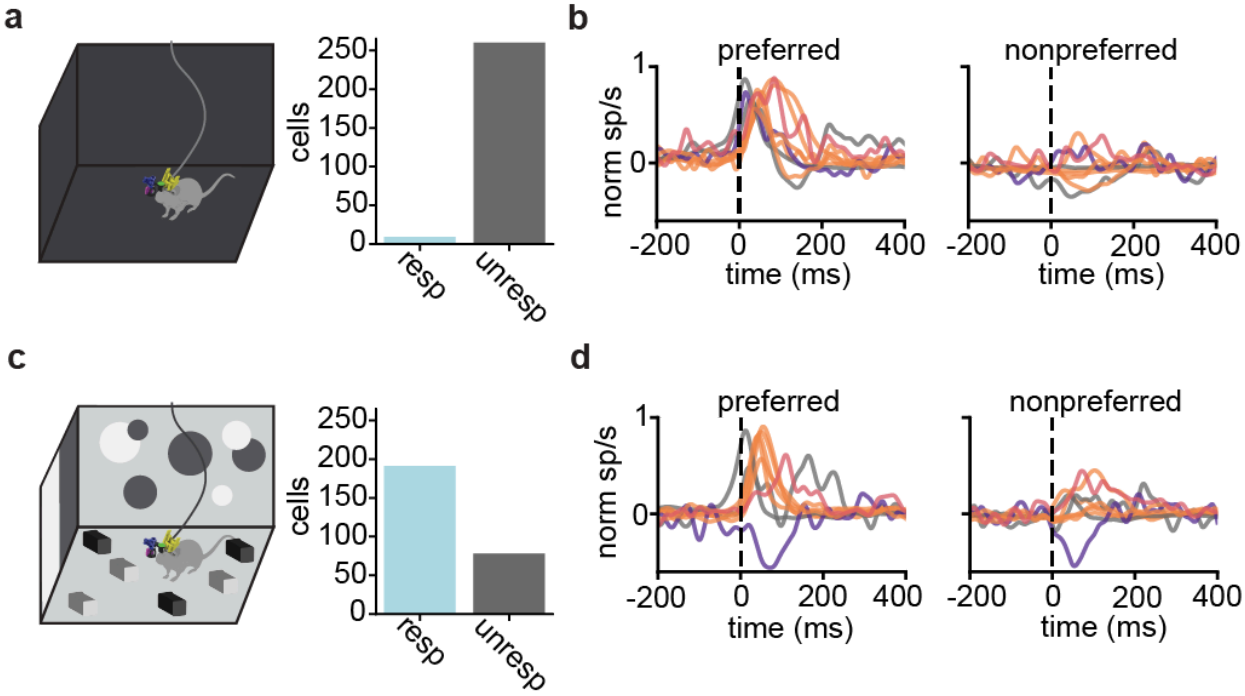
excluded from the analysis in the main text. **f**, Amplitude of gaze changes at onset of movement for the example recording used in **a**. **g**, Median ± SEM amplitude of gaze change for all recordings. Each recording is a point ($n = 9$ mice). Recording in *f* is shown in orange. Compensatory: 0.73 ± 0.02 deg; intermediate: 2.63 ± 0.01 deg; gaze-shifting: 8.76 ± 0.29 deg. **h**, PETHs for example cells from Fig. 1g including PETH for responses to intermediate saccades in black. **i**, Normalized PETHs of gaze-shifting (left), intermediate (middle), and compensatory (right) eye/head movements for 100 example units with a baseline firing rate >2 Hz, with median of all cells ($n = 716$) overlaid.



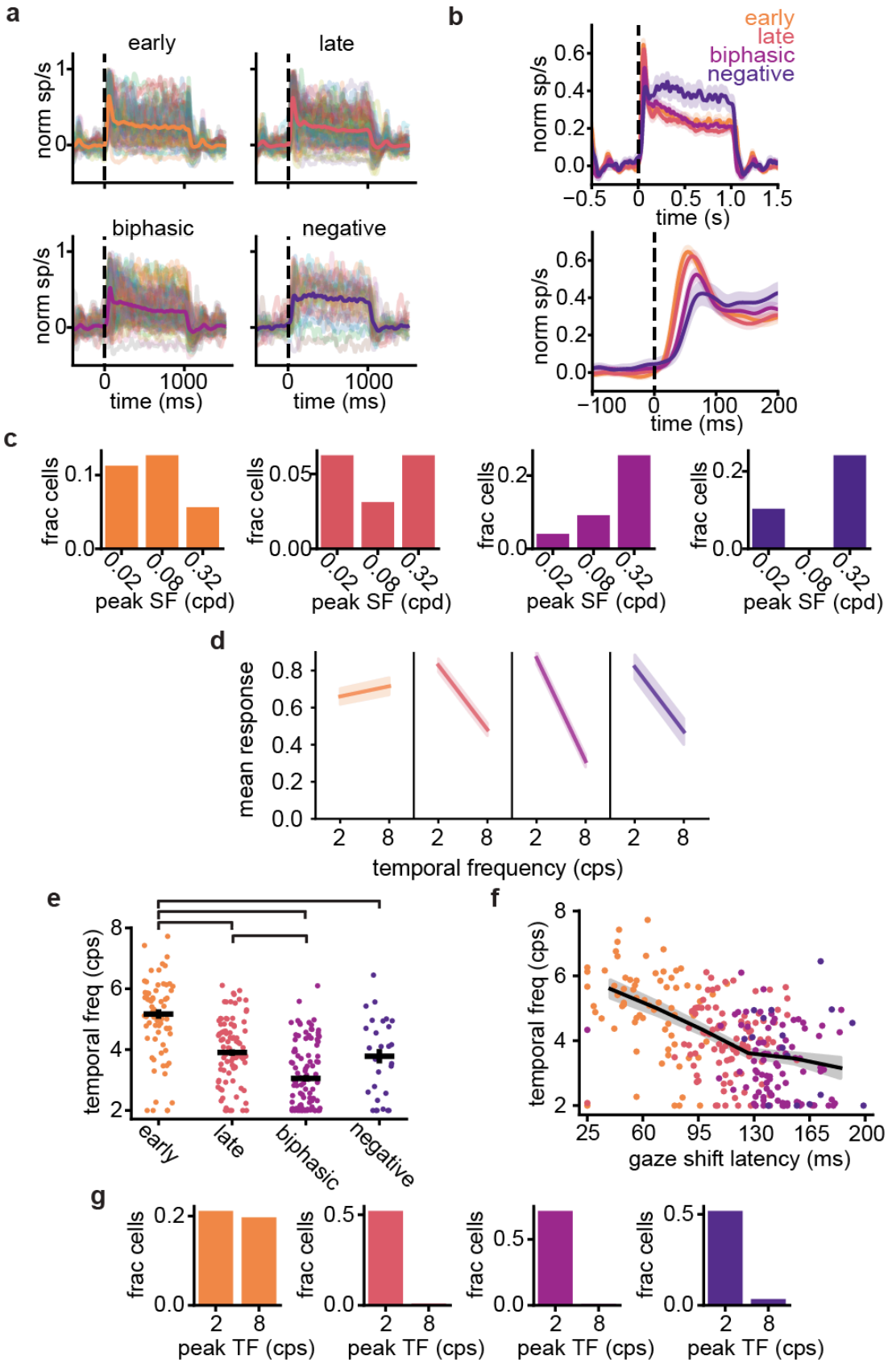
Extended Data Figure 2. Additional characterization of gaze shift response types. **a**, PCA of gaze shift PETHs. Only the two PCs with the highest explained variance are shown. Cells in the scatter plot are colored by the cluster they were assigned by *k*-means clustering of PCs. **b**, Fraction of units in each gaze shift response cluster. **c**, Latency of peak responses were significantly different for all comparisons between clusters ($p < 0.05$ with no effect of experimental session $p = 0.220$, linear mixed effects model; early vs. late $p = 7.10e\text{-}42$, early vs. biphasic $p = 3.86e\text{-}116$, early vs. negative $p = 3.83e\text{-}162$, late vs. biphasic $p = 2.32e\text{-}25$, late vs. negative $p = 2.30e\text{-}65$, biphasic vs. negative $p = 6.67e\text{-}22$). **d**, Fraction of putative cell types in each gaze shift response cluster. Excitatory and inhibitory groups were identified by *k*-means clustering on spike waveforms (waveforms shown above). **e**, Median ± SEM baseline firing rate of units during freely moving (left) and head-fixed (right) recordings ($n = 9$ mice, $n = 716$ cells). Freely moving baseline was calculated as the pre-saccadic period before gaze shifts. Head-fixed baseline was calculated as the firing rate during presentation of gray screen during head-fixation. **f**, Scatter plot of head-fixed and freely moving baseline firing rates. Each point is a cell. Linear regression shown as dashed black line. (early: $r = 0.87$, $p = 1.01e\text{-}26$, m = 1.94; late: $r = 0.70$, $p = 7.06e21$, m = 1.33; biphasic: $r = 0.70$, $p = 3.01e\text{-}26$, m = 1.01; negative: $r = 0.69$, $p = 1.45e\text{-}10$, m = 0.97). **g**, Gaze shift left/right direction selectivity index by cluster. **h**, Laminar depth of all cells determined using the local field potential from multi-unit activity power along each shank of the probe. Black outline shows the distribution of depths for all cells. Dashed line (0 µm) is the estimated depth of cortical layer 5, to which depths were aligned. **i**, Normalized horizontal angular velocity tuning for all cells, separated by response clusters. Positive values for angular velocity represent each unit's preferred horizontal direction of gaze shift. **j**, PETH for compensatory eye/head movements for cells responsive to compensatory movements ($n = 48/716$). Only the preferred direction is shown. Responsiveness defined as 10% modulation and modulation by at least 1 sp/s. **k**, Percent of each gaze shift response cluster that is responsive to compensatory movements (total=48/716, early=4/82, late=5/135, biphasic=17/170, negative=15/66, unresponsive=7/263) **l**, Same as **j** grouped by gaze shift response cluster.



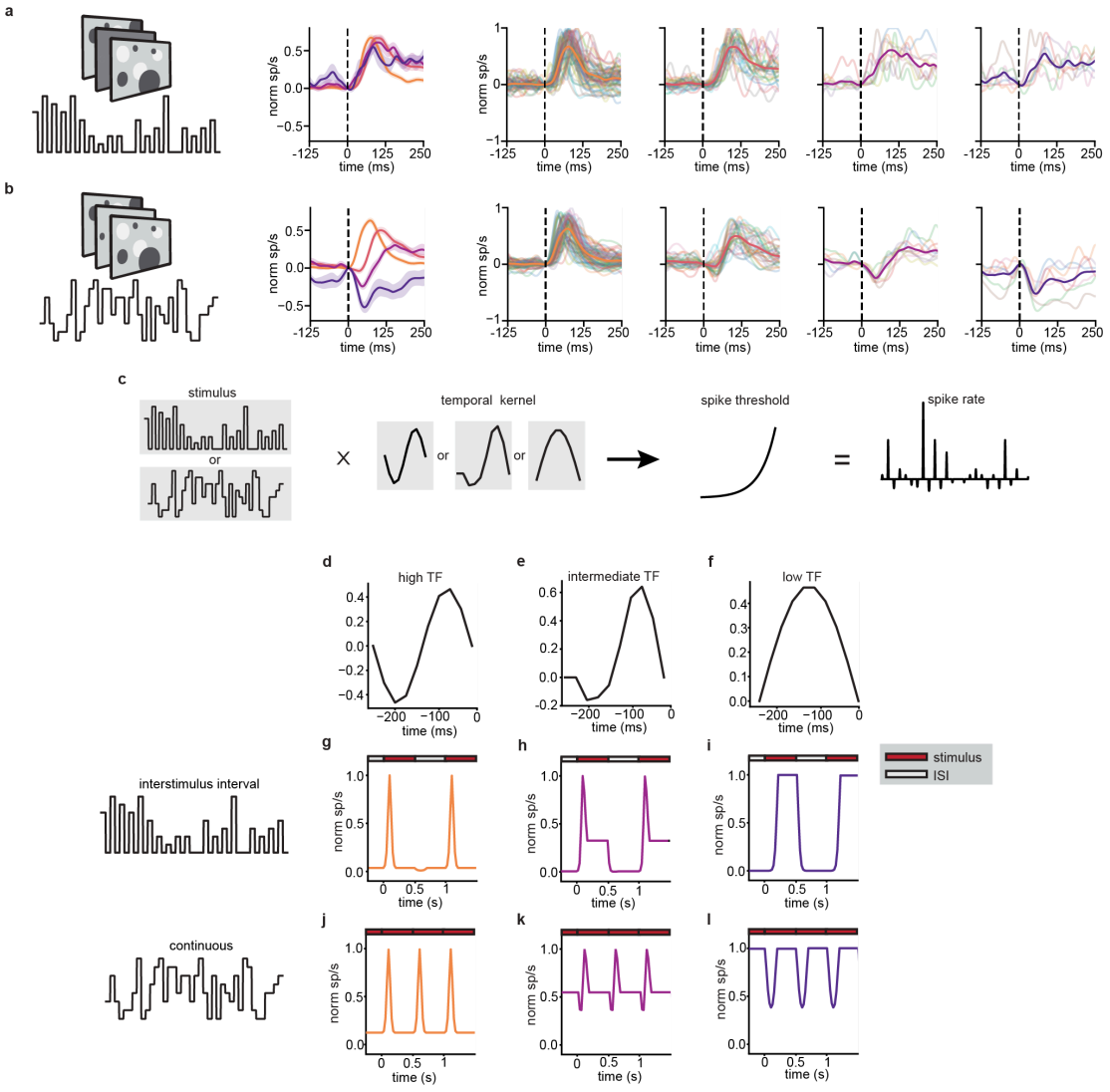
Extended Data Figure 3. Cross validation of response latencies. a, Cross-validation for mouse gaze shift PETHs of all responsive cells. Gaze shift times were randomly divided into two sets used to calculate PETHs in the train (left) and test (right) sets. The test set was sorted by the latency of the positive peak in the train set. **b**, Latency of gaze shift response for train versus test sets ($r = 0.870$, $p = 2.51e\text{-}140$). **c**, Same as **a** for marmoset saccades. **d**, Same as *b* for marmoset saccades ($r = 0.875$, $p = 1.44e\text{-}106$).



Extended Data Figure 4. Additional characterization of responses in the dark. **a**, Fraction of cells responsive in the dark condition (responsive=9/269, unresponsive=260/269). **b**, Dark condition PETHs for cells that responded to gaze shifts in freely moving dark conditions. Units are colored by clustering from responses in light condition. $n = 9/269$ (early=5, late=1, biphasic=0, negative=1, unresponsive=2). **c**, Same as **a** for the light condition (responsive=191/269, unresponsive=78/269). **d**, Responses of units in **b** for the light condition.


a
early
late
biphasic
negative
norm sp/s
time (ms)
b
early
late
biphasic
negative
norm sp/s
time (s)
time (ms)
c
frac cells
peak SF (cpd)
d
mean response
temporal frequency (cps)
e
temporal freq (cps)
f
temporal freq (cps)
gaze shift latency (ms)
g
frac cells
peak TF (cps)

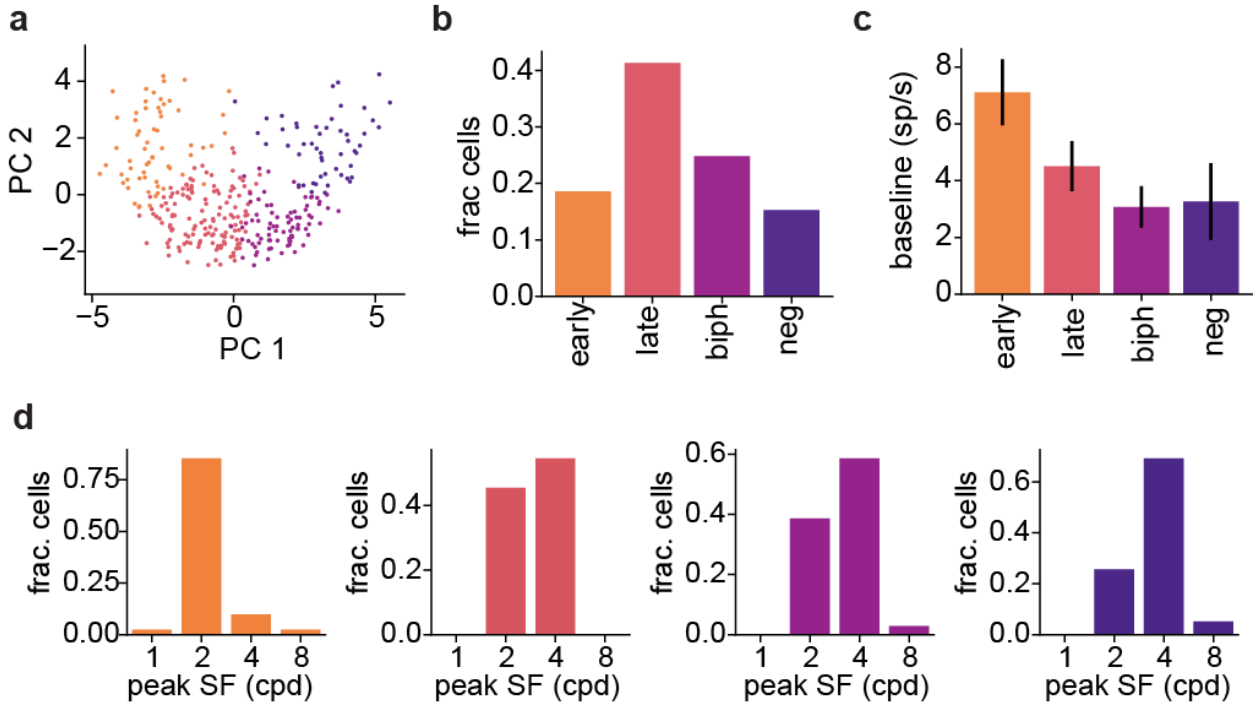
Extended Data Figure 5. Additional characterization of drifting gratings responses. a, Head-fixed drifting gratings PETHs for gaze shift response clusters with mean response overlayed. Stimulus is presented for 1 s with gray ISI between stimuli. $n = 9$ mice, $n = 384/716$ cells responsive to gratings (early=71, late=96, biphasic=98, negative=29, unresponsive=90). Cells below firing rate threshold are not shown. **b**, Mean ± SEM normalized gratings PETHs clustered by gaze shift response for full stimulus presentation (top) and highlighting stimulus onset (bottom). **c**, Fraction of cells in each cluster with a ≥ 2:1 preference for the presented spatial frequencies compared to the sum of responses for the two other spatial frequencies. **d**, Mean ± SEM temporal frequency tuning curve by cluster (Multivariate two-way ANOVA, TF x cluster $F = 21.45$, $p = 3.45\text{e-}13$). **e**, Temporal frequency preference for gratings-responsive cells in each gaze shift response cluster, calculated as a weighted mean of responses ($n = 9$ mice, 384 cells). Median and standard error are shown for each cluster. Bars above indicate statistical significance at $p < 0.05$ (linear mixed effects model, $n = 9$ mice, $n = 384$ cells; early vs. late $p = 3.64\text{e-}7$, early vs. biphasic $p = 2.24\text{e-}21$, early vs. negative $p = 4.42\text{e-}9$, late vs. biphasic $p = 2.32\text{e-}6$, late vs. negative $p = 5.69\text{e-}2$, biphasic vs. negative $p = 9.37\text{e-}2$). **f**, Weighted temporal frequency preference versus gaze shift response latency, for all cells responsive to gratings. Running median ± SEM for all cells is overlaid. The color of each point indicates the cluster from gaze shift responses. ($r = -0.468$, $p = 2.12\text{e-}16$). **g**, Same as **c** for temporal frequency.



Extended Data Figure 6. Temporal tuning of neurons can explain diverse responses to gaze shifts.
a, Responses to flashed sparse noise stimulus presented with an inter-stimulus interval (ISI; $n = 3$ mice; early=71,

Extended Data Figure 6 continues on next page

Extended Data Figure 6 continued
late=33, biphasic=9, negative=7). Left: mean ± SEM for each cluster; right: individual neuron responses overlaid with mean. **b,** Same as **a** for a continuously flashed sparse noise stimulus. **c**, Schematic of modeling approach. A scalar stimulus, presented continuously or with an inter-stimulus interval (ISI), is passed through a variable temporal kernel of either high, intermediate, or low temporal frequency (TF), and a non-linearity is used to generate a spiking output. **d**, High TF kernel. **e**, Intermediate TF kernel. **f**, Low TF kernel. **g**, Resulting response of model using high TF kernel to visual stimuli presented with an ISI. **h**, Same as *g* for intermediate TF kernel. **i**, Same as *g* for low TF kernel. **j**, Resulting response of model using high TF kernel to visual stimuli presented continuously. **k**. Same as **j** for intermediate TF kernel. **l**, Same as **j** for low TF kernel.



Extended Data Figure 7. Additional characterization of marmoset saccade response types. a, PCA of marmoset gaze shift PETHs for the 2 PCs with highest explained variance, colored by k-means clusters. **b**, Fraction of units in each saccade response cluster. **c**, Median ± SEM baseline firing rate of units in each cluster ($n = 2$ marmosets, $n = 238$ cells). **d**, Fraction of cells with maximal response to each presented spatial frequency.

3.8 SUPPLEMENTARY INFORMATION

Refers to Figure S6: Temporal tuning of neurons can explain diverse dynamics in response to gaze shifts. The diversity of visual responses profiles (positive, biphasic, negative) that we observed following gaze shifts (Figure 2) and sequentially flashed stimuli (Figure 4) are strikingly different from the positive responses typically observed in standard visual stimulus paradigms, such as the drifting gratings shown in Figure 5. We hypothesized that this may be due to the presence of a gray inter-stimulus interval (ISI) in the grating stimuli, which is standard in presentation of many visual stimuli but is absent during gaze shifts in visually complex environments, where a new visual input arrives with each saccade without a return to baseline. To test this, we presented flashed sparse noise either continuously or with a gray screen ISI and

grouped cells into continuous gratings response clusters using PCA and k-means models from gaze shift responses. While responses to continuously presented flashed stimuli resembled response profiles seen following freely moving gaze shifts, this similarity was not preserved when an ISI was interleaved between stimuli (Figure S6a,b).

Furthermore, the correspondence between temporal frequency preference and gaze-shift response, along with the similarity in response to flashed stimuli, suggests that these distinct responses could result from the neurons' inherent temporal dynamics applied to visual input that are changing abruptly in rapid succession. To investigate this, we developed a simple linear-nonlinear model of a neuron's response to a time-varying stimulus (Figure S6c). We model the stimulus as a scalar input representing the visual drive, which varies discretely across eight levels at 500 ms intervals. We implemented two forms for this time-varying input: one in which the input returns to zero on alternate intervals, consistent with standard visual stimulus presentation with a gray ISI, and the other in which the input varies randomly at each interval without returning to zero, consistent with continuous stimulus presentation as occurs in sequential gaze shifts. We then modeled the neuron's response as a linear temporal kernel applied to this stimulus, followed by an exponential nonlinearity to produce a spike rate. We used three temporal kernels: a biphasic kernel (Figure S6d) which mathematically acts as a differentiator and corresponds to transient, high TF responses; a unimodal positive kernel (Figure S6f) which acts as an integrator and corresponds to sustained, low TF responses; and an intermediate between these two (Figure S6e). These represent the distribution of temporal kernels typically observed in LGN and V1 (Piscopo et al., 2013; DeAngelis & Freeman, 1995; Benardete & Kaplan, 1999) with a corresponding relationship between the temporal kernel and temporal frequency tuning (Idrees et al., 2022).

Applying this model to the input with a gray ISI and calculating the mean PETH across all stimulus transitions recapitulated expected responses from standard ISI stimulus presentation (Figure S6a), with the high TF kernel evoking a rapid transient response relative to ISI baseline (Figure S6g), the intermediate kernel evoking a transient followed by sustained response (Figure S6h), and the low TF kernel evoking a slower but sustained response (Figure S6i). Strikingly, when the model was applied to the continuous input, it reproduced response types that we observed for gaze shifts and flashed stimuli with no ISI (Figure S6b), with the high TF kernel

evoking short latency positive responses (Figure S6j), the intermediate kernel evoking a biphasic response (Figure S6k), and the low TF kernel evoking negative responses (Figure S6l). The intuition behind this is that during continuous stimulus presentation, for the high TF kernel each new stimulus evokes a transient response and then returns to baseline, resulting in a positive peak. For the low TF kernel, rather than resting at a baseline firing rate before stimulus onset as would be the case with a gray ISI, the pre-stimulus rate is determined by the sustained response to the previous stimulus. Following the transition to a new stimulus, the firing rate drops due to the offset of the previous stimulus and the slow (low TF) response to the new stimulus, resulting in what appears as a negative response. In other words, the negative response need not result from active suppression, but simply the loss of response to the previous stimulus and lack of rapid onset response to the new stimulus. For the intermediate kernel, there is a combination of these two effects, leading to a brief drop from the previous stimulus offset followed by a mix of transient and sustained responses, giving a biphasic response.

Together, this modeling suggests that the distinctive response patterns following a gaze shift can be explained, at least at the qualitative level, by each neuron's temporal dynamics in response to individual stimuli, together with the fact that during gaze shifts, the response to each new stimulus rides on the response to the previous stimulus. Note that a recent study demonstrated that a related model can explain suppressive responses to sequential stimuli at the level of the retina (Idrees et al., 2022).

CHAPTER IV
IMPACT OF ACTIVE VISION ON MOUSE SUPERIOR COLLICULUS

4.1 AUTHOR CONTRIBUTIONS

The following is unpublished materials with Jhoseph Shin, Dylan M Martins, Keaton Jones, and Cristopher M Niell. SL Sharp and CM Niell conceived the project. J Shin led data analysis. SL Sharp led experiments and contributed to data analysis. DM Martins contributed to data analysis. K Jones contributed to histology.

4.2 ABSTRACT

Vision is an active process that involves exploration of the environment through head, eye, and body movements. Many studies have highlighted the importance of active sensing during visual exploration, however due to experimental constraints there is limited data on the impact of free movement on visual processing. It is of particular interest to investigate the impact of active vision on superior colliculus (SC) given its role in visuo-motor integration. SC is known for its multimodal organization, with superficial SC (sSC) being responsive to visual stimuli and deep SC (dSC) responding to multimodal sensory input and initiating motor output. Many have highlighted the importance of head movements in redirecting gaze during visual sampling but the role SC plays in this visuo-motor behavior has not been studied across depth in free moving rodents. To determine the impact of active vision we record neural activity across depth of SC during gaze shifts while simultaneously recording eye and head position during free movement. We find that sSC neurons respond to the influx of new visual information during saccadic movements, whereas dSC neurons respond to the head movements themselves, as demonstrated by their sustained response in darkness. Additionally, we find motor responses in dSC are most responsive to head movements rather than eye movements. Further, we compare sSC visual responses initiated by gaze shifts to known gaze shift responses in the primary visual cortex (V1) finding similarities in a few key response types, although unique collicular responses emerge. A striking similarity between SC and V1 gaze shift dynamics emerges when comparing gaze shift initiated temporal sequence responses, as activity in both areas forms a visually dependent temporal sequence that corresponds to coarse-to-fine visual processing. Our results showcase distinct visual processing differences between colliculus and cortex as well as highlighting the various roles SC plays during gaze-shifting movements.

4.3 INTRODUCTION

When analyzing the visual scene one often utilizes movement of their eyes, head, and body to sample the environment. This combination of movements is essential to active vision, which is necessary for most visually dependent tasks. The SC is a highly conserved midbrain structure known for its involvement in visuo-motor behaviors (Basso & May, 2017; Liu et al., 2022). SC has a laminar structure with superficial layers dedicated to visual processing (Liu et al., 2022; Gale & Murphy, 2014; Basso et al., 2021; Cazemier et al., 2024; Cang et al., 2018) and deeper layers responding to multimodal input and producing motor output (Wilson et al., 2018; Gandhi & Katnani, 2011); this organization makes SC primed to analyze aspects of the visual scene and initiate appropriate movements. A large breadth of research shows that SC is involved in the generation of head and eye movements across species. Specifically in mice, activity of SC has been studied as it relates to visuo-motor behaviors.

While much research has focused on either superficial SC (sSC) or deep SC (dSC), few studies have explored the neural responses of both regions simultaneously. A recent study explored how superior colliculus sifts visual information across depth finding strong differences in response preferences to visual stimuli between sSC and dSC (Lee et al., 2020). It is of particular importance to investigate the activity of SC across depth as many visuo-motor tasks rely on the coordination of sSC and dSC to identify visual information and initiate appropriate behavior. Prey capture is a prime example of an ethologically relevant visuo-motor behavior that relies on the integration of sSC responses to visual input to identify prey and dSC motor output to successfully catch it (Hoy et al., 2019; Zhao et al., 2023). In fact distinct cell types within SC have been identified to contribute differently to prey capture behavior (Hoy et al., 2019). Further, the cooperation of sSC and dSC can be seen when looking at another ethologically relevant behavior, escape (Campagner et al., 2023; Shang et al., 2018). It has been shown that sSC identifies looming stimuli (Campagner et al., 2023) while circuits in dSC initiate increased attentiveness or restlessness associated with escape behavior (Bittencourt et al., 2005). Altogether this work highlights the need to study sSC and dSC together in action in order to better understand SC's role in visual behavior.

The role of mouse SC during gaze-shifting head and eye movements has not been thoroughly investigated. Traditionally neural responses in SC have been studied using head-fixed or head-rotating experimental paradigms which allow for precise experimental control but give

limited insight into active visual processing. Head-fixed neural recordings in sSC showcase robust, cell type specific (Hoy et al., 2019; Gale & Murphy, 2014), responses to various visual stimuli. Of the few studies looking at movement modulation in mouse SC using head-fixed approaches, locomotor modulation seems modest (Savier et al., 2019) but interestingly head-fixed saccades are consistently accompanied by attempted head movements (Meyer et al., 2020), suggesting important involvement of head and eye movements in SC visual processing. In dSC, free moving experiments have found neurons tuned for different directions of head movements around the three axes of rotation (Wilson et al., 2018).

Further, our recent work in mouse primary visual cortex during free movement finds that neural responses to gaze-shifting head and eye movements have unique temporal dynamics (Parker et al., 2023). These dynamics are visually dependent and correspond with coarse-to-fine visual processing. Importantly these response dynamics are the result of the abrupt change in the visual input that occurs during saccadic movements, not the head or eye movements that occur during saccades. It is unknown if similar dynamics exist in SC during gaze-shifting movements considering the different roles of collicular and cortical regions in visual processing (Basso et al., 2021), and SC's known involvement in the generation of head and eye movements (Gandhi & Katnani, 2011).

Together these findings highlight the importance of studying vision in unrestrained conditions allowing for head and eye movements. While there is a breadth of research studying the visual responses in sSC and motor responses in dSC it is unknown how these regions respond together during active vision. We therefore sought to create an experimental paradigm to investigate neural responses across the depth of SC during free moving head and eye movements. This work highlights distinct visual processing differences between SC and V1 as well as unveiling the various roles SC plays during active vision.

4.4 RESULTS

4.4.1 Measuring head and eye movements together with neural activity in SC of freely moving mice

First, we wanted to characterize the responses of neurons in mouse SC across depth during gaze-shifting head and eye movements. To do this we used chronic electrophysiology and a head-mounted camera system we previously employed in Parker et al. (Parker et al., 2023). Briefly we used a recording set-up that allowed for measurement of right eye position with a

camera aimed at the eye, the animal's field of view with a forward-facing camera, head angle and acceleration with an inertial measurement unit (IMU), and a single-unit electrophysiology with a chronically implanted 128 channel silicon electrode positioned in SC (Figure 1A). Our experimental set-up allowed for sequential head-fixed and free moving recordings, which enabled us to measure responses to controlled visual stimuli in head-fixation, and neural activity during active sensing in free movement (Figure 1B). To determine electrode depth we presented contrast-reversing checkerboard stimuli to measure neurons local field potential (LFP) along with histology. These LFP responses were plotted and the point of maximum deflection within a recording determined electrode depth as described in Ito et al. (Ito et al., 2017) and further explained in the methods section (Figure 1 C-E).

Mice sample the visual scene in a 'saccade-and-fixate' pattern in which coordinated head and eye movements select and stabilize input to the retina (Land, 2019). During head movement the mouse's eyes do one of two things: move in the same direction as the head to shift its gaze, or move in the opposite direction of the head to stabilize gaze; we refer to these movements as gaze-shifting or compensatory, respectively (Michaiel et al., 2020; Meyer et al., 2020). During free movement these behaviors occur in a continuous pattern that represents gaze-shifting head and eye movements (saccade) and stabilizing head and eye movements (fixate). In this work we classify only horizontal gaze-shifting movements during free movement as it is known that vertical eye movements in mice are largely compensatory (Michaiel et al., 20202; Meyer et al., 2020; Miura & Scanziani, 2022; Giovannetti & Rancz, 2024). Additionally, we analyze the impact of head movement alone.

Consistent with previous findings during free movement (Parker et al., 2023; Michaiel et al., 2020; Meyer et al., 2020), we observe many large amplitude horizontal head movements that are accompanied by horizontal eye movements. We then classify these movements as either saccadic or compensatory by calculating the gaze direction (sum of head and eye movement) defined in detail in the methods. Additionally, we analyze the impact of head movement independent of eye movement via head-only IMU measurements. This head and eye movement categorization can be easily visualized by plotting eye vs head velocity, where compensatory movements align along the diagonal, representing equal but opposite head and eye movements, whereas gaze-shifting movements fall beyond the diagonal (Figure S1A).We define gaze-shifting and compensatory movements as previously outlined in Parker et. al., 2023, briefly, timepoints

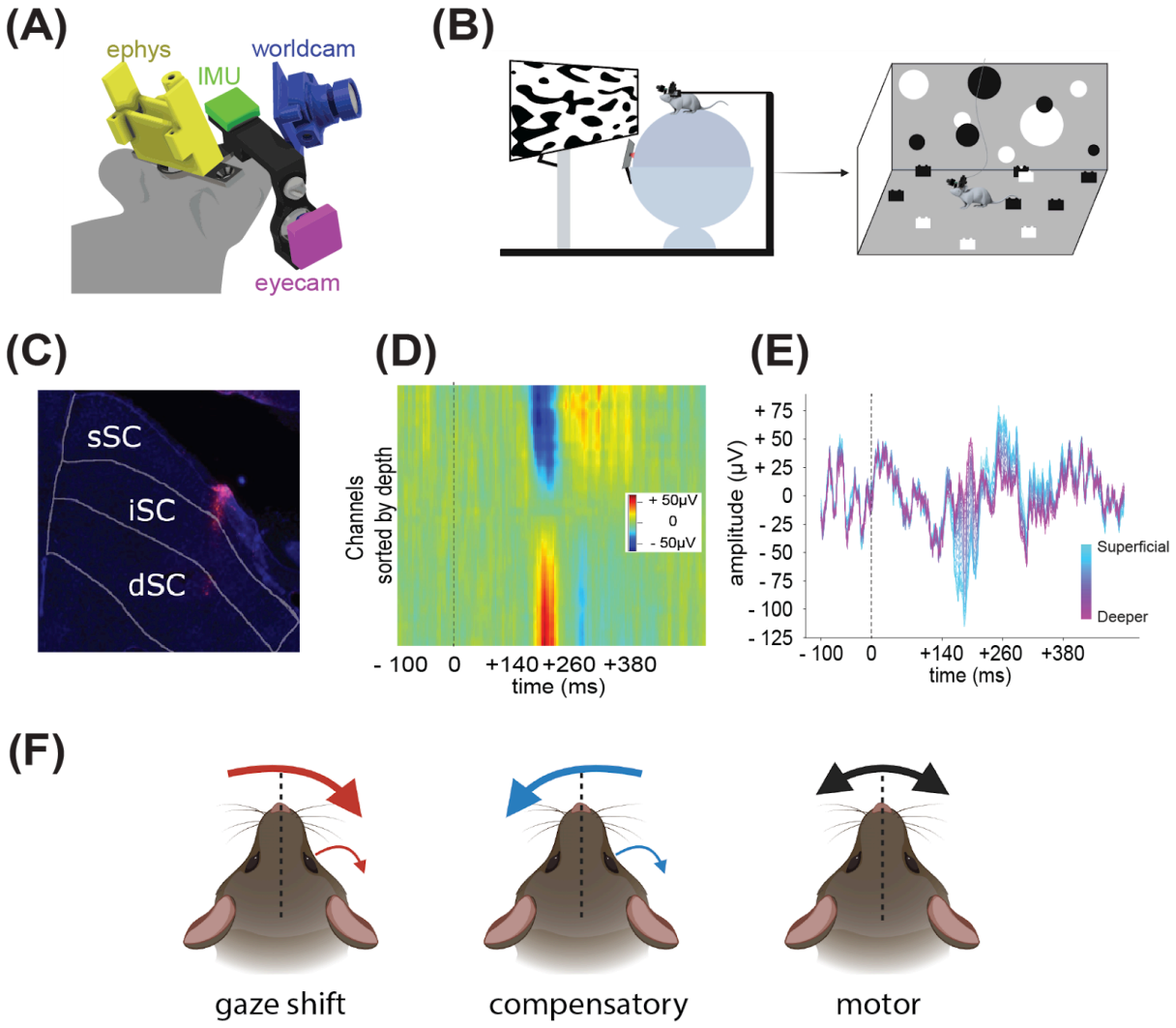


Figure 1. Experimental set-up. **(A)** Schematic of head-mounted recording system which includes 128 channel silicon electrode implanted in SC, eye camera, world camera, and inertial measurement unit (IMU). **(B)** Experimental paradigm where neural responses are recorded during head-fixed set-up immediately followed by freely moving recordings. **(C)** Histology of electrode tract (shown in pink) going through all layers of SC. **(D)** Local field potential (LFP) responses for each channel sorted by depth. **(E)** LFP spikes for example electrode shank showing amplitude responses for each channel after reversing checkerboard stimulus presentation (shown by dashed line) with deeper cells in purple and more superficial cells in turquoise. Depth was calculated based on maximum amplitude deflection described in methods and modeled after Ito et al., 2017. **(F)** Schematic of the head and eye movements we focus on in this work. Gaze-shifting (left) movements occur when the animal moves their head and eyes together in the same direction to shift their gaze. Compensatory head and eye movements occur when the animal moves eyes and head in opposite directions to stabilize gaze. Motor responses are characterized by head movements in either direction regardless of eye movement.

where gaze velocity (sum of head and eye velocity) is >240 deg s^{-1} are gaze-shifting movements and timepoints where gaze velocity is <120 deg s^{-1} are compensatory movements. Eye movements that had velocity values between 120 deg s^{-1} and 240 deg s^{-1} were excluded from analysis to avoid contamination of the two movements.

4.4.2 Temporal dynamics of gaze shift responses in SC

Using chronically implanted silicon electrodes we extracted single-unit neural activity across the depth of SC in freely moving mice around the time of compensatory and gaze-shifting head and eye movements. Figure 2A,B shows example responses of neurons in sSC and dSC to gaze-shift movements. The sSC neurons responded with a positive or negative change in firing rate immediately following the gaze shift, and their response was similar for both directions of gaze shift. On the other hand the dSC neurons began responding before the gaze shift with a more extended duration, with a strong positive response to one direction of gaze shift and a negative response to the other direction. In our population of 601 neurons from 11 mice, we find a large majority of cells (42.3%) respond to gaze-shifting movements,which produce abrupt changes in retinal input, whereas responses to compensatory movements, which stabilize retinal input, are more rare (9.9%). Neuronal responses in sSC predominantly respond to gaze-shifting head/eye movements and do not respond to compensatory movements (Figure S2A). While neurons in dSC appear to respond to the head/eye movement that occurs during gaze-shifts.

These response profiles can also be seen throughout trials with sSC neurons responding either positively or negatively to the onset of gaze shift and dSC responding to the head/eye movements that occur during the gaze shift, as shown in spike rasters across trials in Figure 2A,B and average peri-event time histograms (PETHs) in Figure 2C,D. This response type difference can also be seen in single trials when comparing rasters of neural responses at the population level during free movement; sSC neurons respond in a unique temporal dynamic at the onset of a gaze-shift (Figure 2C, bottom) while in contrast dSC neurons respond predominantly in a synchronous burst like fashion that initiates prior to the gaze-shift and persists after (Figure 2D, bottom).

The gaze shift initiated temporal dynamic is further distinguished between sSC and dSC when looking at population level temporal sequences sorted by depth from most superficial sSC to deepest dSC, as in Figure 2G. The most superficial cells respond quickly after the onset of the

gaze shift with either a positive or negative change in firing rate. Whereas cells in the dSC respond around 500 ms prior to gaze shift onset, and persist for 500 ms after. Neurons in sSC generally respond similarly to preferred and non-preferred gaze shift directions whereas neurons in dSC show clear direction preference with negative responses to non-preferred gaze shifts, and positive responses to preferred gaze shifts (Figure 2H). We further quantified these differences by taking the mean PETH responses to gaze shifts in sSC and dSC. In sSC we see early positive and negative responses around the onset of gaze shift (Figure 2I). Neurons in dSC show motor responses with activation for preferred gaze shift direction and suppression for non-preferred gaze shift direction (Figure 2H). Additionally, dSC neurons respond to compensatory head/eye movements (Figure S2A). However, we see smaller amplitude motor responses during compensatory movements as compared to gaze-shifting movements, this difference makes sense given that compensatory movements on average have lower head amplitude than gaze-shifting movements ($p < 0.0001$; Wilcoxon rank-sum test) (gaze-shift: 373.42 deg s^{-1}, compensatory: 285.11 deg s^{-1}).

4.4.3 dSC gaze-shift responses persist in the dark, but sSC responses do not

To confirm that this temporal dynamic is visually dependent, we recorded gaze shift responses from the same neurons during free movement in both light and dark conditions. In darkness, the temporal dynamic following a gaze-shift changes drastically for sSC cells whereas dSC cells response remains in the absence of visual input (Figure 3 A, B). Specifically, positive and negative cell sSC responses are largely eliminated in darkness while dSC responses are unaffected (Figure 3B). These differences are also evident in the mean PETH gaze shift responses in light and dark (Figure 3C,D). Thus during active vision sSC and dSC respond differently to the visual environment with sSC responding largely to the saccadic movement in a visually dependent manner, and dSC responding to both gaze-shifting and compensatory head/eye movements, even in the absence of visual input, consistent with encoding the motor component itself.

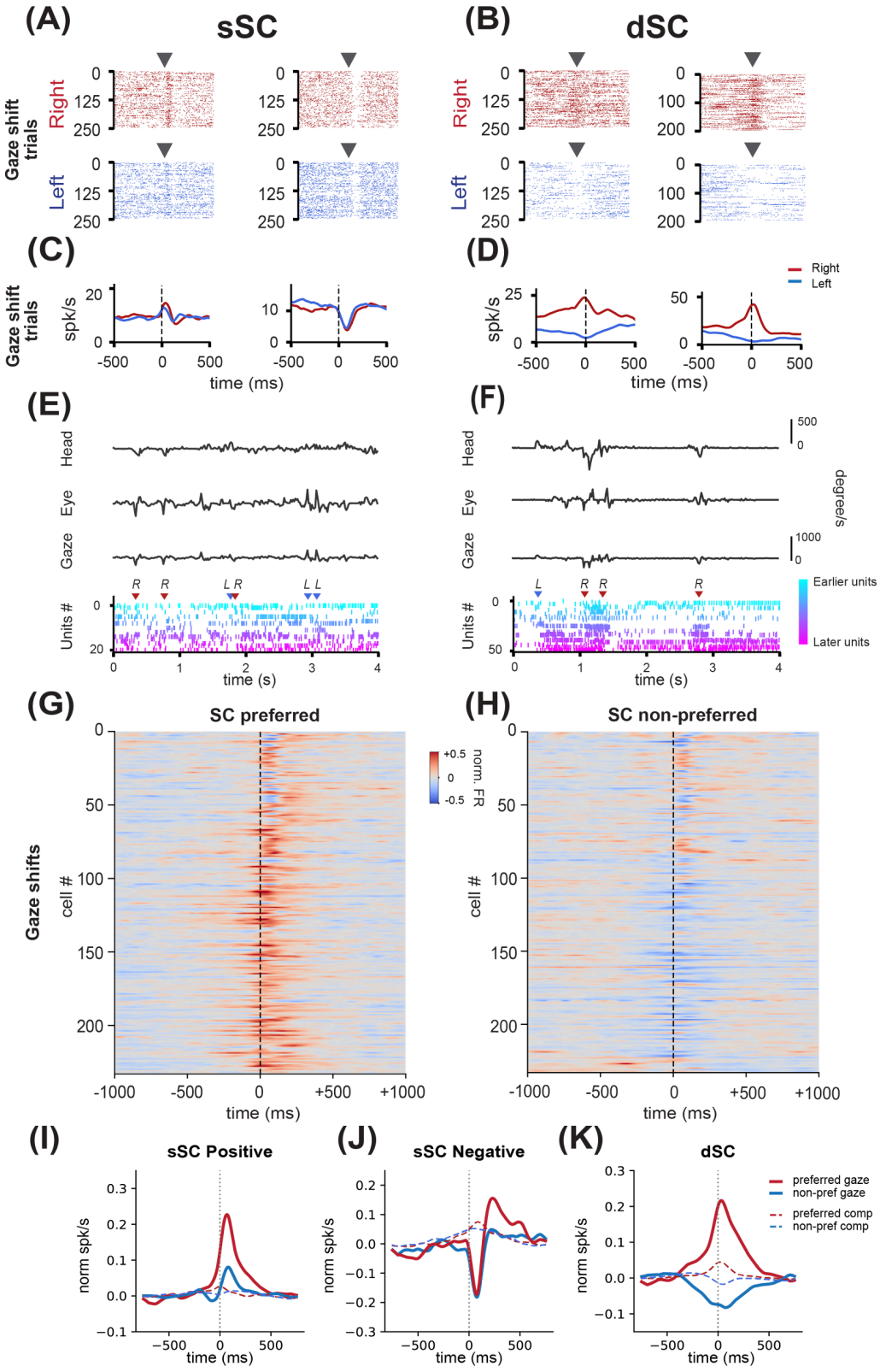

(A)
sSC
(B)
dSC
Gaze shift trials
Right
Left
(C)
(D)
spk/s
time (ms)
Right
Left
(E)
(F)
Head
Eye
Gaze
Units #
time (s)
degrees
Earlier units
Later units
(G)
SC preferred
(H)
SC non-preferred
Gaze shifts
cell #
norm FR
time (ms)
(I)
sSC Positive
(J)
sSC Negative
(K)
dSC
norm spk/s
preferred gaze
non-pref gaze
preferred comp
non-pref comp

Figure 2. Temporal dynamics of gaze shift responses in SC. **(A)** Trial spike rasters showing different two response types: positive (left) and negative (right) in sSC for leftward gaze shift movements (top) and rightward gaze shift movements (bottom). Arrowheads indicate onset of gaze shift. **(B)** Trial spike rasters showing two examples of motor responses in dSC, leftward gaze shift movements (top) and rightward gaze shift movements (bottom). Arrowheads indicate onset of gaze shift. **(C)** Corresponding example PETHs to above positive and negative response trial rasters, rightward gaze shift movement (red) and leftward gaze shift movement (blue). Dashed line indicates gaze shift onset. **(D)** Same as **C** but for dSC motor responsive example neurons shown above in **B**. **(E)** Example traces from free moving experiment showing head movement, eye movement , and gaze (sum of head and eye movements). Bottom raster shows responses from neurons in sSC during this experiment. Heat map indicates latency to gaze shift onset. L and R arrows indicate time of gaze shift, and arrow color indicates which direction gaze shift was, red for right, blue for left. **(F)** Same ad E but for example dSC experiment. **(G)** Temporal sequence of preferred gaze shift responses across depth, organized by depth, with cell 0 being the most superficial. **H)** Same as **G** but for non-preferred gaze shifts in SC. **(I)** Mean PETHs response for sSC positive responsive cells. Solid red line shows preferred gaze shift response. Solid blue line shows non-preferred gaze shift direction response. Dashed red line shows preferred compensatory movement response and dashed blue line shows response for non-preferred compensatory. (n = 11 animals, **(J)** Same as **I** but for sSC negative responsive cells. **(K)** Same as **I**, **J** but for dSC motor responsive cells. (n = 11 mice, 267 cells)

4.4.4 sSC neurons are tuned to contrast while dSC neurons are tuned to head movement

We next sought to further investigate the relative responsiveness of SC neurons to visual input versus head movement itself. During head-fixation we presented band-limited white noise contrast-modulated stimuli, that progresses from grey screen up to full contrast and back down to low contrast over a period of 10 seconds (Niell & Stryker, 2008). This stimulus allowed us to measure the contrast response function of each neuron to temporally varying contrast levels. We see a clear demarcation in visual responsiveness between sSC and dSC. Figure 4A shows contrast response functions measured from white noise, for the same example sSC and dSC units as in Figure 2A,B. The positive and negative sSC cells have strong contrast response, whereas no distinct contrast response preference is seen in the dSC neurons (Figure 4A). We further quantified contrast responses of the population across SC depth by calculating a contrast modulation index (CMI) for each gaze-shift responsive cell based on its contrast response function Briefly, for each cell we calculate its CMI based on the response to maximum contrast R_{max} and to zero contrast R_0 as $((R_{CMax} - R_{C0})/(R_{CMax} + R_{C0}))$. Again a clear distinction in the degree of contrast modulation emerges across SC depth. We see that cells that are within the sSC

have varying degrees of contrast modulation ranging from 1 to -1 whereas dSC do not appear to be modulated by contrast (Figure 4B). These results from head-fixation suggest that gaze-shift responsive neurons in sSC are responsive to standard visual stimuli whereas dSC cells are not. Together with the fact that sSC responses are eliminated in the dark, while dSC responses persist, this suggests that sSC gaze shift responses are driven by the visual input while dSC neurons are driven by the head movement themselves.

The white noise stimulus also allowed us to identify a neuron's receptive field via their spike-triggered average response. Example neurons in sSC show robust receptive fields whereas the dSC neuron lacks a defined receptive field (Figure 4C). These RF trends persist at a population level (Figure 4D). These results from head-fixation suggest that gaze-shift responsive neurons in sSC are responsive to standard visual stimuli whereas dSC cells are not. These results suggest that sSC gaze shift responses are driven by the visual input while dSC neurons respond to head movements themselves.

We next examined the degree to which dSC neurons are driven by head movement by calculating a head velocity tuning curve for each cell in dSC. Figure 4E shows head velocity tuning curves for two individual dSC neurons. Both units show strong head direction preference, one example unit has a maximum firing rate at -250 deg s^{-1} whereas the other fires maximally at +250 deg s^{-1}. We further quantified head movement modulation by calculating a head-movement modulation index (HMI) for each cell across the depth of SC using head movement metrics from the IMU during free movement. HMI is calculated by taking an average of each cell's firing rate at +360 deg and -360 deg $((R_{+360°} - R_{-360°})/(R_{+360°} + R_{-360°}))$ (Figure 4F). When looking at head movement modulation across the population the majority of cells in sSC do not seem to be modulated by head movement, with a small proportion of sSC cells appearing weakly modulated by head movement. Whereas the large majority of dSC cells are modulated by head movement (Figure 4F). As a final comparison between contrast modulation and head-movement modulation we calculated average modulation of CMI and HMI across depth in SC. Clear distinctions between visual and head movement modulation across depth are seen, as sSC cells are strongly modulated by contrast but not head-movement whereas dSC cells are not modulated by contrast but heavily modulated by head-movement (Figure 4G).

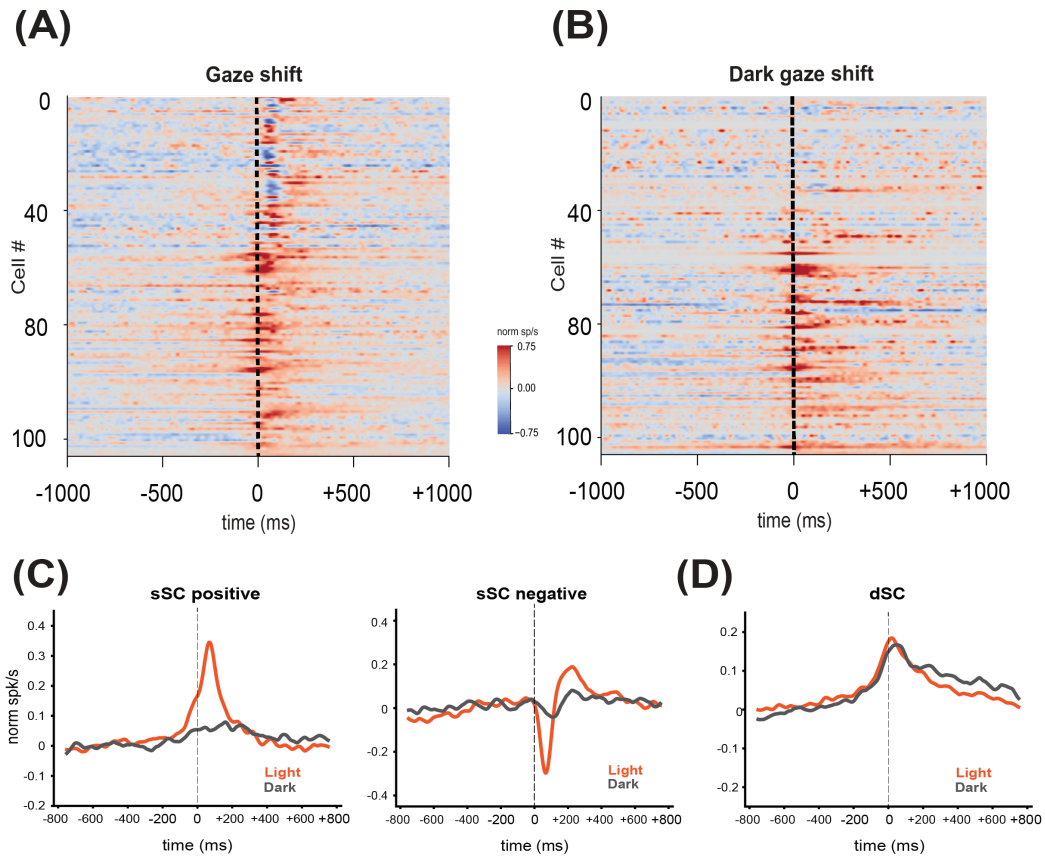


Figure 3. sSC gaze shift responses are visually driven, dSC are not. **(A)** Temporal sequence of gaze shift responses during light recording plotted by depth (n = 11 animals, 105 cells). **(B)** Same as A but during complete darkness experiments. **(C)** Mean PETH of sSC early positive cells (left) and negative (right) in light and dark conditions. **(D)** Same as **C** but for cells in dSC.

4.4.5 sSC neurons have unique gaze shift initiated temporal dynamics when compared to V1

We next wanted to characterize the temporal sequence of sSC to gaze shifts. In order to directly compare sSC to V1, we used the same clustering as previously used in Parker et al. (Parker et al., 2023) to segregate V1 neurons. Briefly, to characterize the diversity of neural responses, we performed *k*-means clustering of units based on principal component analysis (PCA) of the normalized peri-event time histograms (PETHs) for gaze shifts at the onset of the gaze shift, when a new visual stimulus hits the retina. We also calculated the preferred direction of movement for each unit by determining which direction yielded the largest peak response and used preferred responses for PETH clustering. Clustering allows us to categorize the population

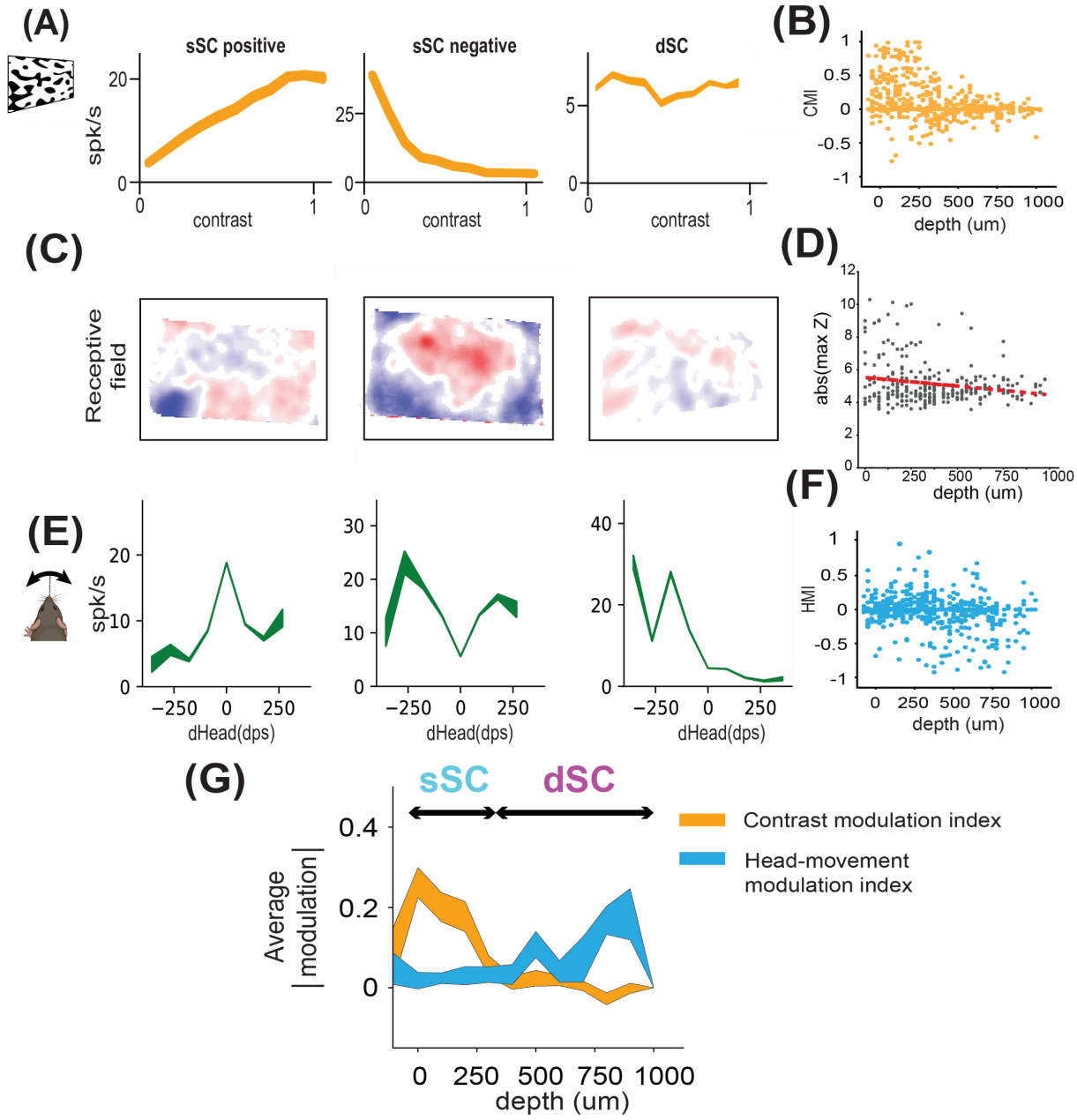


Figure 4 sSC is tuned to contrast, dSC is tuned for head-movement (A) Contrast response functions (spk/sec) for each response type, positive sSC, negative sSC, motor dSC, calculated from response to presentation of contrast variable white noise stimulus. **(B)** Contrast modulation index (CMI) is calculated using cells' response to contrast-modulated white noise stimulus ((R_{CMax} - R_{C0})/(R_{CMax} - R_{C0})). **(C)** Receptive field (RF) maps for each of the example units in A. RF was calculated from spike triggered average responses to white noise stimulus. **(D)** Z-score scatter calculated from white noise spike triggered average (β= -0.20, SE = 0.008, p< 0.001).

Figure 4 continues on next page

Figure 4 continued

(E) Head velocity tuning curves from example dSC neurons. **(F)** Head-movement modulation index (HMI) was calculated for each cell (($R_{+360°}$ - $R_{-360°}$)/($R_{+360°}$ + $R_{-360°}$)). **(G)** Average modulation (absolute value) of SC neurons across depth to CMI and HMI). Presumptive sSC and dSC areas noted with text above responses. (n = 11 mice, 267 cells)

---

dynamics of sSC however, it should be noted that we are subdividing a continuous population. A value of $k = 5$ clusters was implemented to allow for direct comparison to visual cortex (V1) gaze shift dynamics in Parker et al., 2023. This resulted in four responsive clusters, early positive (21.3%), late positive (8.3%), biphasic (10.1%), and negative (5.9%) responses to gaze shifts, as well as a large unresponsive fifth cluster (54.4%) (Figure 5A). It should be pointed out that while all four response types appear in sSC this is likely due to using the same *k*-means clustering metrics as were used for V1 experiments (Parker et al., 2023). For a direct comparison between gaze shift responses in both regions we kept the same metrics for sSC analysis but if sSC cells' responses were to be clustered independently it is possible that different response types would appear. When directly comparing the distribution of response types within sSC the population is significantly different than in V1 (Chi-squared test $X^2_{(1)}= 43.96$, $p<0.001$; Figure 5B). These results suggest that sSC and V1 respond to gaze shifts with different temporal dynamics.

In addition to clustering response types in sSC we also sorted units based on the latency of their peak positive response to gaze shifts to analyze the temporal sequence of the population. This revealed a unique sSC temporal dynamic to gaze shifts (Figure S3A). This temporal sequence response in sSC can also be visualized on single trial spike rasters (Figure 2C,D). We confirmed the emerging temporal sequence was not an artifact of sorting through cross validation, where we sorted one half of the dataset (training set) and applied it to the other half of the dataset (test set; Pearson correlation coefficient, $r = 0.61$, $p < 0.0001$), which showed that temporal order and peak response latency were maintained (Figure S3B). Direct comparison of response types between the regions finds notable differences in prominent response type, with sSC majority response type being early positive where V1 majority response type is biphasic. These results suggest sSC and V1 encode visual information during gaze-shifts uniquely.

4.4.6 Coarse-to-fine processing around gaze shifts in sSC

We sought to further investigate the visual tuning properties of sSC gaze shift responsive cells. Specifically, we were interested to see if cells in sSC had a coarse-to-fine relationship

between their visual tuning responses and temporal responses to gaze shifting head/eye movements, similar to that observed in V1(Parker et al., 2023). During head-fixation we presented drifting sinusoidal gratings at three spatial frequencies (0.02, 0.08, 0.32 cycles per degree (cpd)) and two temporal frequencies (3, 8 cycles per second (cps)). We first compared temporal frequency (TF) preferences for each response type in sSC, and found no significant difference in TF between clusters (Figure 5D). This suggests that sSC gaze shift responsive cells do not have significant temporal frequency preference differences between response types (p = 0.317, Kruskal-Wallis test, n = 11, n =77 cells).

We next wanted to compare SF preferences in sSC. We find a significant difference in SF preferences between early positive and negative response types in sSC (Figure 5C). We found the earliest responding cells prefer low SF and the latest responding cells prefer high SF. These results show a weak coarse-to-fine visual processing pattern, in which first the coarse features of the visual scene are processed followed by more fine details It is important to note that the SF preferences in sSC differ when compared to V1 responses (Parker et al., 2023). The SF preferences in sSC do not share exact SF preference values with V1 nor do sSC cells have a gradually increasing SF preference across all response types. Together, these results suggest that sSC processes visual input following gaze-shifts differently than V1.

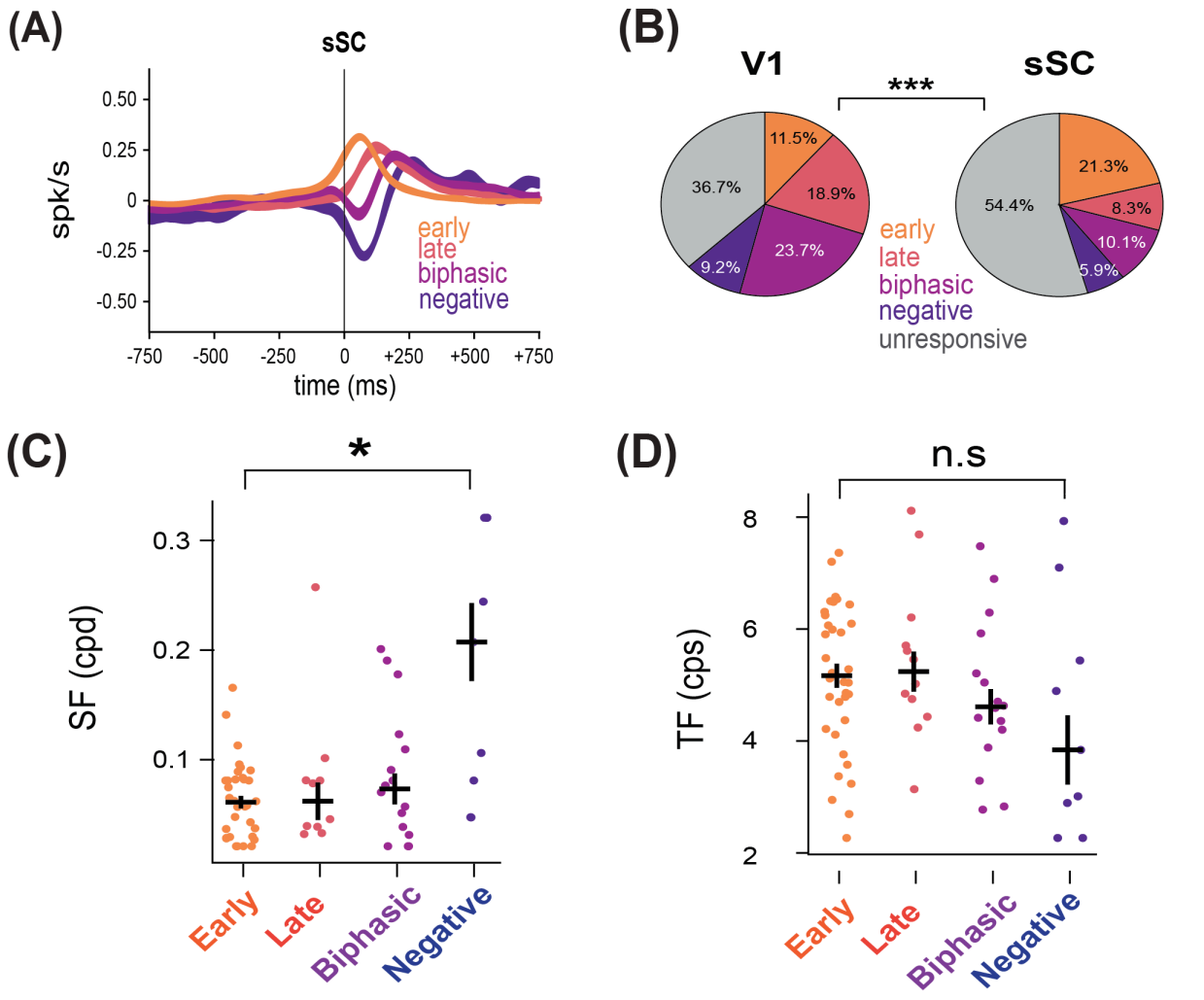


Figure 5. sSC response types and coarse-to-fine processing (A) Mean PETHs of sSC response types clustered using the same metrics as in Parker et al., 2023. **(B)** Proportion of cells in each response type between V1 and sSC showing significant difference in proportion of cells in each response group. (Chi-squared test $X^2_{(1)}= 43.96$, $p<0.001$) **(C)** Spatial frequency preference for gratings-responsive cells in each gaze shift response cluster, calculated as weighted mean of responses. Median and standard error are shown for each cluster. Bar and asterisk represent statistical significance at $p<0.05$ (Kruskal-Wallis test, $n = 11$ animals, $n = 77$ cells; Wilcoxon Rank-sum test early versus negative, $p = 0.005$, early versus biphasic $p = 0.020$, biphasic versus negative $p = 0.029$).**(D)** Temporal frequency preference for gratings-responsive cells in each gaze shift response cluster, calculated as weighted mean of responses.Bar above indicates no significance at $p = 0.317$ (Kruskal-Wallis test, $n =11$ animals, $n =77$ cells).

4.5 DISCUSSION

We recorded the activity of SC neurons across depth in freely moving mice to investigate how active vision, in terms of head and eye movements, influences neural activity in the superior colliculus. A wide breadth of previous research has shown the importance of SC in visuo-motor behaviors. Additionally, a wide scope of research has shown that SC is organized in a topographic manner with sSC responding to visuosensory input and dSC responding to multimodal input and producing motor output. However, the integration of sSC and dSC during active sensing is largely understudied. We found distinct gaze-shifting responses for sSC and dSC, with sSC responding to the visual aspect of gaze shifts and dSC responding to the motor aspect of gaze shifts. Specifically, in sSC we find two main response types, early and negative, to gaze-shifting head/eye movements; however, they are not responsive to compensatory head/eye movements. Additionally, when animals were placed in an arena lacking any visual input (complete darkness) sSC neurons' response dynamic to gaze-shifts was eliminated. Further, we found that gaze shift responsive sSC neurons have strong contrast preferences to white noise stimuli presented during head-fixation. These findings suggest that sSC responses to gaze-shifting head/eye movements likely result from the change in visual input during these saccadic movements as opposed to motor efference signals of the combined head/eye movement. We believe this is the first time sSC responses during free movement have been measured in darkness and these results further support the established notion that sSC encodes visual information rather than motor signals.

We also compared sSC gaze shift responses to established coarse-to-fine gaze shift dynamics in V1 (Parker et al., 2023). We found that sSC neurons that responded earliest following the onset of gaze shift preferred low SF, and neurons responding latest to gaze shift onset preferred higher SF. These results correspond with a weak form of the coarse-to-fine visual processing we observed in V1. Coarse-to-fine visual processing postulates that the visual system responds first to coarse aspects of the visual scene before responding to fine details (Hedgé, 2008) sSC spatial frequency preferences clearly differ when compared to V1 which suggests that sSC and V1 process the input of visual input following a gaze shift differently. While we are not sure what exactly these SF preference differences represent in terms of overall visual processing differences, one possible explanation is that sSC and V1 are primed to respond differently to the gaze shift induced abrupt change in visual input. For instance, sSC may be tuned to respond

preferentially to the coarse, low spatial frequency, features of the new visual input whereas V1 may be tasked with higher order processing of filling in and making sense of the new visual input which it does through coarse-to-fine processing.

Additionally, the temporal sequence of sSC responses, early positive, late positive, and negative, are linked to the 'saccade-and-fixate' pattern of natural vision where rapid saccadic eye movements are interspersed with slower fixational movements. Together our findings showcase a unique response dynamic in sSC following gaze-shifting head and eye movements and highlight the need to study visual processing in sSC in unrestrained conditions.

Furthermore, we found that neurons in dSC respond to the motor aspect of gaze-shifting eye movements rather than the abrupt change in visual input. This is to be expected given the role of dSC in generation of head and eye movements. We also found dSC responds to compensatory movements although the response amplitudes are lesser than gaze-shifting responses. Lastly, cells in dSC are significantly modulated by head movement, and have strong direction of head movement preferences.

4.5.1 Temporal dynamics and gaze shift responses in sSC

Our work provides insight into how sSC encodes head and eye movements during spontaneous exploration of a visually enriched environment. To our knowledge this is the first time neural responses to head/eye movements have been investigated during free movement in mouse sSC. By measuring head and eye movements simultaneously with neural recordings we found that most neurons in sSC respond to movements when the animal's head and eyes move together to change the direction of gaze, resulting in new visual input. Neurons in sSC do not respond to compensatory movements in which the animal's head and eyes move in opposite directions to stabilize their gaze, consistent with the fact that there is no change in visual input. Additionally, sSC's gaze shift response dynamic is visually dependent given that gaze shifts in the absence of visual input (complete darkness) do not evoke similar responses. Specifically, in darkness sSC cells do not seem to respond to the gaze-shifting head and eye movement at all, with neither activation nor suppression of responses. This result suggests that cells in sSC are responding to the abrupt change in visual information that occurs during the gaze shift rather than to its motor component.

Recent work in mouse V1 has shown a dynamic temporal sequence after gaze shift initiation that corresponds with coarse-to-fine visual processing (Parker et al., 2023). We wanted

to determine if gaze shift temporal responses in sSC share similarities with mouse V1. Using the same clustering analysis as Parker et al., 2023, we find that sSC shares response types with V1. However, the prominent response type in sSC is early positive as compared to biphasic in V1. It is worth noting that while all four response types appear in sSC this is likely due to using the same *k*-means clustering metrics as were used for V1 experiments. If sSC cells' responses were to be clustered independently it is likely that fewer response types may emerge or perhaps new clusters would appear but in order to make direct comparisons between gaze shift responses in both regions we had to use the same metrics for sSC analysis. Cells in sSC respond with a unique response to gaze shifts that shares response type similarities with V1 but differs in many ways. The distribution of response types across the gaze shift responsive population differs significantly between sSC and V1. This suggests that while sSC and V1 respond reliably to the rapid change in visual input that occurs during a gaze shift, their detailed response dynamics to gaze shifts are unique.

4.5.2 Coarse-to-fine processing in sSC

A wide breadth of research from human psychophysics supports coarse-to-fine processing as a visual processing framework (Peyrin, 2010; Neri, 2011; Petras et al., 2019). Recently this framework has been reported in V1 in both mice and marmoset (Skyberg et al., 2022; Parker et al., 2023). Coarse-to-fine processing is the idea that first lower SF components are recognized/processed prior to higher SF components of the visual scene. Here we examined coarse-to-fine processing in mouse sSC during active vision for the first time, which revealed a weak progression of low SF to high SF that corresponds with early and late gaze-shift initiated responses.

4.5.3 Conclusion and future work

Together our work shows a unique pattern of processing during active vision across depth in SC. We find that gaze shifts initiate a temporal sequence in sSC that supports coarse-to-fine visual processing. In dSC we find a robust motor response dynamic to head and eye movements during both gaze-shifting and compensatory movements. This work highlights the importance of studying visual processing in natural conditions that allow for unrestrained movement to best understand the neural computations performed by the visual system.

This work sets a foundation for future studies exploring circuit dynamics within SC during unrestrained active vision. One potential next step would be to investigate gaze shift

responses on a cell type specific level. Previous work has highlighted the importance of SC wide-field (WF) vertical and narrow-field (NF) vertical neurons in prey capture (Hoy et al., 2019), a visuomotor task heavily reliant on SC. These cell types respond preferentially to different visual stimuli during head-fixation and prey capture, making them ideal to explore during gaze shift dynamics. WF cells respond preferentially to moving stimuli and NF cells are direction selective with direct inputs to dSC that influence head/eye direction (Schiller & Stryker, 1972;Doubell et al., 2003). These cell types have different visual tuning properties and projections within SC which suggests they may have different neural dynamics during gaze-shifting head/eye movements. One could utilize optogenetics to target WF or NF cell populations and perturb them during unrestrained gaze shifts to calculate response dynamics in the absence of input from either cell type. This experiment would determine if WF and NF cells respond similarly to gaze shifts.

Additionally, this work bisected responses from SC into two major subdivisions, sSC or dSC. However, many studies also consider an intermediate SC (iSC) layer, at the intersection of sSC and dSC, where visual and multimodal inputs from cortex enter. In primates limited evidence suggests visually enhanced connections between sSC and iSC (Wurtz & Mohler, 1976) but these layer interactions are vastly understudied, and have not been investigated in rodents. Additionally, it is largely unknown if iSC sends visual signals to dSC to initiate head/eye movements but there is recent evidence of numerous synaptic connections between these layers (Doykos et al., 2020). Targeted iSC and dSC electrophysiological recordings during free movement in a visually complex environment could begin to deduce the role of iSC during active vision.

Lastly, it would be of great interest to explore gaze shift dynamics during task related active vision. Our work thus far has studied gaze shift responses during spontaneous exploration of a visually enriched arena. It is reasonable to hypothesize gaze shift response dynamics may differ in a rewarded visual task. In mice and primates reward has been shown to influence visual processing (Henschke et al., 2020; Soltani et al., 2021). It is unknown if similar effects would be seen during rewarded gaze shifts during free movement.

Ultimately, this work highlights the importance of studying visual processing in unrestrained conditions to reveal the complex dynamics of visual processing; while also

providing a cornerstone for future studies to investigate gaze shift dynamics in SC during many behaviors at the cellular and circuit level.

4.6 METHODS

4.6.1 Animals

All procedures were conducted in accordance with National Institutes of Health guidelines and were approved by the University of Oregon Institutional Animal Care and Use Committee. The 5-10 month-old mice (Mus musculus, C57BL/6J, Jackson Laboratories and bred-in house) were kept on a 12-h light/dark cycle. In total, five male and six female mice were used for this study. Mice were housed with sibling cagemates until habituation and then they were singly housed. Humidity was between 40-60% and temperature was 21±1°C. Data collection and analysis were not performed blind due to the condition of the experiments. No animal or data points were excluded from analysis. Data collection was not randomized as there was only one experimental group.

4.6.2 Surgery and habituation

Mice were initially implanted with a titanium headplate over SC to allow for head-fixation and attached of head-mounted experimental hardware. After 3 days of recovery a miniature connecter (Mill-Max 853-93-100-10-001000) was secured to the skull to allow for repeated, reversible attachment of a camera arm, eye/world cameras, and IMU (Michaiel et al., 2020; Parker et al., 2022, Parker et al., 2023). To simulate the weight of the real electrophysiology implant for habituation, a ‘dummy’ electrophysiology drive was glued to the headplate. Animals were handled for several days before surgical procedures, and habituated (~60 mins) to the head-fixed spherical treadmill and freely moving arena while tethered for several days before experiments.

The electrophysiology implant was performed once animals moved comfortably for an extended period of time in the arena. A craniotomy was performed over monocular SC, and linear silicon probe (128 channels, Diagnostic Biochips P64-10D, P128-6) mounted in a custom three-dimensionally printed drive (Yuta Senzai, UCSF) was lowered into the brain using a stereotax to an approximate tip depth of 1500 μm from the pial surface. The surface of the craniotomy was covered with artificial dura (Dow DOWSIL 3-4680) and the drive was secured to the headplate using light-curable dental acrylic (Unifast LC). A second craniotomy was performed above the left frontal cortex and a reference wire was inserted into the brain. The

opening of the craniotomy was covered with a small amount of sterile ophthalmic ointment and the wire was glued in place with UV light cure glue (Loctite AA3972). Animals recovered overnight and experiments began the following day.

4.6.3 Hardware and recording

The camera arm was oriented approximately 90 deg to the right of the nose and included an eye-facing camera (iSecurity101 1000TVL NTSC, 30 frames per s (fps) interlaced), an infrared light-emitting diode to illuminate the eye (Chanzon, 3-mm diameter, 940-nm wavelength), a wide-angle camera oriented toward the mouse's point of view (BETAFPV C01, 30 fps interlaced) and an inertial measurement unit acquiring three-axis gyroscope and accelerometer signals (Rosco Technologies; acquired at 30 kHz, downsampled to 300 Hz and interpolated to camera data). Fine-gauge wire (Cooner, 36 AWG, no. CZ1174CLR) connected the IMU to its acquisition box, and each of the cameras to a USB video capture device (Pinnacle Dazzle or StarTech USB3HDCAP). A top-down camera (FLIR Blackfly USB3, 60 fps) recorded the mouse in the arena.

The Diagnostic Biochips electrophysiology headstage (built into the silicon probe package) was connected to an Open Ephys acquisition system via an ultra-thin cable (Intan no. C3216). Electrophysiology data were acquired at 30 kHz and bandpass filtered between 0.01 Hz and 7.5 kHz. We first used the Open Ephys GUI (https://github.com/open-ephys/plugin-GUI) to assess the quality of the electrophysiology data, then recordings were performed in Bonsai20 using custom workflows (https://github.com/nielllab/FreelyMovingEphys). System timestamps were collected for all hardware devices and later used to align data streams through interpolation.

During experiments, animals were first head-fixed on a spherical treadmill to measure visual tuning properties using traditional metrics, then immediately transferred to an arena where they could move freely and explore the visually enriched environment. Recording duration was approximately 40 min head-fixed and 1 hr freely moving. For head-fixed experiments, a 70-cm monitor (BenQ GW2780) was placed approximately 27.5 cm away from the mouse's right eye, and visual stimuli were presented using Pyschtoolbox-3 (Kleiner et al., 2007). Head-fixed stimuli were recorded using the head-mounted world camera. We first presented 15 min of a band-limited Gaussian noise stimulus (Piscopo et al, 2013) (spatial frequency spectrum 0.05 cpd to 0.12 cpd, flat temporal frequency spectrum with a low-pass cutoff at 4 Hz) to confirm SC targeting based on spike-triggered average receptive fields. We next presented flashed sparse

noise stimuli consisting of full- and minimum-luminance circular spots on a gray background played for 5 min. Spots were a range of sizes, 2,3,8,16, and 32 degrees in diameter and were presented so that each size made up an equal fraction of the screen, approximately 15% on average. Each stimulus frame was presented for 250 ms immediately followed by the previous frame without ISI (Piscopo et al., 2013). Next, sinusoidal gratings were presented at eight directions of motion for three spatial frequencies (0.02, 0.08, 0.32 cpd) and two temporal frequencies (2, 8 cps) for 10 min, with a 1 sec stimulus duration and 0.5 sec gray ISI with stimulus conditions randomly interwoven (12 presentations per stimulus). Finally, we presented a contrast-reversing square-wave checkerboard stimulus for 5 min with a spatial frequency of 0.04 cpd and temporal frequency of 0.5 Hz.

The arena for free moving experiments was 48-cm long by 37-cm wide by 30-cm high. Three of the arena walls were custom wallpaper covered by acrylic including black and white high- and low- spatial frequency gratings and white noise, the other wall was covered by a 61-cm monitor (BenQ GW2480). The monitor displayed moving sparse noise where full- and minimum-luminance spots were 4, 8, and 16 deg in diameter. Each spot moved in one of eight evenly spaced directions at one of five speeds (10,20,40,80,120 deg s^{-1}). Spots appeared on the edge of the screen and moved across until they disappeared on the opposing side. The arena floor was a gray silicone mat (Gartful) and was covered with a dense layer of black and white Lego bricks to provide 3-D visual contrast. To encourage foraging during the recording small pieces of tortilla chips (Juanita's) were lightly scattered on the arena floor; animals however, were not food or water restricted.

For dark recordings the entire experimental enclosure was sealed in light-blocking material, all potential light sources within the enclosure and all external light sources were turned off, to ensure complete darkness. Prior to darkness experiments one drop of 2% pilocarpine HCl ophthalmic solution was applied to the animal's right eye to constrict the pupil and allow for accurate eye tracking as described in Parker et al., 2023. Once the pupil was restricted enough for tracking in the dark (3-5 min) the animal was moved from the spherical treadmill to the dark arena for the recording. The recording lasted approximately 20 min or until the effect of pilocarpine wore off followed immediately by the light recording.

4.6.4 Data preprocessing

Electrophysiology data was largely preprocessed following the same methodology as Parker et al., 2023. Raw electrophysiology data from same session head-fixed and free moving recordings were concatenated into a single file for spike sorting, allowing us to track single units across the entire experiment. Common-mode noise was removed by subtracting the median across all channels at each timepoint. Spike sorting was performed using Kilosort 2.5 (https://github.com/MouseLand/Kilosort), and single units were isolated and selected using Phy 2.0 (Phy 2.0 beta 5; https://github.com/cortex-lab/phy) based on numerous parameters: contamination (<10%), firing rate (mean>0.5 Hz across entire recording), autocorrelogram, and waveform shape. Following spike sorting the data were then split back out into individual recordings for further analysis.

Laminar depth was calculated from the multi-unit local field potential (LFP) of head-fixed neural responses during presentation of flash checkerboard stimulus which generates strong LFP responses in sSC. To identify the surface of sSC we followed methods from Ito et al., 2017. Briefly, we averaged evoked LFP responses to reversal of checkerboard stimulus and chose the time point when LFP has the maximum negative amplitude to represent the top of sSC. Similar analysis is done when looking across depth of SC to identify most superficial responses in a given recording. Additionally, cells were split into sSC or dSC, cells that may be in intermediate SC (iSC) were grouped with dSC for simplicity.

Response_ch = mean(LFP_ch)

Pupil position was extracted from the eye camera data (Michaiel et al., 2020). Briefly, eye camera data were deinterlaced to achieve 60-fps video, then eight points around the pupil were tracked with DeepLabCut (Mathis et al., 2018). We then fit an ellipse to these points and computed pupil position by angular rotation. The head-mounted world camera was also deinterlaced to achieve 60-fps video and distortion from the camera lens were corrected with OpenCV. The position of the mouse in the arena was recorded using a top-down camera and tracked using DeepLabCut. Running periods were defined as timepoints when the animal's neck point had a velocity of > 2cm s^{-1}, and stationary states were periods where the animal's neck point had a velocity of < 2cm s^{-1}.

Horizontal head rotation velocity was extracted from the IMU, converted to deg s^{-1} and interpolated to eyecam frame timestamps. We defined leftward and rightward directions as the direction of head movement from the animal's perspective. For example, a movement of the right eye in the nasal direction would be a leftward movement. To select eye/head movement onsets, we first determined if the head velocity was >60 deg s^{-1} in the leftward or rightward direction. If a sufficiently large head movement was made, we then separated it into either gaze-shifting or compensatory movements using gaze velocity, where gaze is defined as the sum of horizontal eye and head velocities (Michaiel et al., 2020). If there was a high gaze velocity (>240 deg s^{-1}) concurrent with the head movement, the movement was considered gaze-shifting. Those that resulted in low gaze velocity (<120 deg s^{-1}) were considered compensatory. Eye/head movements that resulted in intermediate gaze velocities (>120 deg s^{-1} and <240 deg s^{-1}) were excluded from our analysis to avoid contamination between the two categories of movements. For eye/head movements spanning multiple eye camera frames, only the first time point, representing the onset of the movement, was used to calculate the onset of the PETH. Compensatory movements which occurred 250 ms before or after a gaze-shifting movement were excluded to avoid contamination by gaze shifts. For head-fixed recordings, only high eye velocity (>240 deg s^{-1}) was used to identify saccades, since the head was immobilized.

To maintain internal consistency with eye cameras and obviate the need for a separate video stimulus synchronization signal, we determined stimulus onsets in head-fixed recordings directly from the head-mounted cameras. For the reversing checkerboard stimulus that updated every 500 ms, we identified frame transitions from the head-mounted camera video based on *k*-means clustering ($k = 2$) of the video into the two separate contrasts, and selected transitions between the clusters. For the flashed sparse noise stimulus that updated every 250 ms, we determined the timestamps of stimulus onset based on the root mean square pixel-wise change in the image, which showed clear peaks at stimulus transition. For drifting sinusoidal gratings, we determined direction of motion by computing optic flow from the worldcam video, SF based on the mean gradient magnitude from paired Sobel operators and TF from the mean Fourier transform of each pixel over time.

4.6.5 Analysis of neural responses

All analyses were performed in Python (v.3.8, python.org). Neural responses were calculated as a PETH from electrophysiology spike times using kernel density estimation with a

gaussian kernel with bandwidth (standard deviation) of 10 ms, sampled at 1-ms intervals. The PETH included neural activity from −1,000 ms to 1,000 ms around the event onset for all stimuli/events (unless noted otherwise). A modulation index was calculated for PETHs as the peak of the firing rate (R) from the event onset at -20 ms until 250 ms after the event, minus the baseline before the event (R_b), which was calculated as the mean of R from −750 ms to -450 ms and +450ms to +750ms around the event onset (unless noted otherwise).

$$\text{mod} = \max(\text{abs}(R - R_b))$$

For the drifting gratings stimulus, the PETH was calculated similarly to other stimuli, but from −1,500 ms before until 1,500 ms after stimulus onset to capture the full stimulus and ISI interval.

The modulation index for gaze-shift responses in the leftward and rightward directions were compared to determine a preferred direction. A direction selectivity index (DSI) was calculated as the difference between the preferred and nonpreferred maximum modulations divided by their sum.

$$\text{DSI} = (\text{mod}_{\text{pref}} - \text{mod}_{\text{nonpref}})/(\text{mod}_{\text{pref}} + \text{mod}_{\text{nonpref}})$$

To normalize PETHs, we subtracted the baseline R_b from the PETH (R) to give an evoked firing rate, and then divided by the maximum of R during a response window (R_{rw}) within a response range of −250 ms before to 250 ms after the event. The denominator includes responses up to 250 ms before the event to ensure that spurious increases in firing rate that are not time-locked to an event were not amplified by the normalization (this was extremely rare and only occurred in unresponsive cells).

$$R_{norm} = (R - R_b)/(\max(R_{rw}))$$

When normalizing the PETH of compensatory eye/head movements, we used that neuron's R_{rw} from gaze shifts in the preferred left/right direction so that all eye/head responses were normalized relative to the neuron's best response. Because the head-fixed sparse noise stimulus presentation was of short duration (250 ms), we used the spike rate at the time of the

stimulus onset, 0 ms, as the baseline firing rate (R_b) for normalization to avoid including responses to previously presented flashed stimuli.

Neurons were considered responsive to a stimulus or eye/head movement onset if they changed their firing rate by at least 10% and 1 spike per second (sp s^{-1}) in the 250 ms following the onset of the event. For the gratings stimulus, suppressed-by-contrast cells were not considered responsive, and removed by requiring that gratings PETHs have a peak firing rate >0.5 normalized sp s^{-1} during the 1 s of stimulus presentation.

The latency of a neuron's peak response was calculated as the timepoint of the maximum firing rate of its normalized PETH in the period between 25 ms to 250 ms after the onset of the event. For temporal sequence plots, normalized PETHs were sorted by the latency of their peak for the preferred direction of gaze-shifting eye/head movements. Peak latency was not calculated for units unresponsive to gaze shifts. Peak latency sorting was cross-validated by randomly assigning gaze-shift events into a train set and a test set, calculating a PETH for each half of the data, and sorting the test set by the peak times of the training set.

SF and TF preferences were based on the evoked firing rate for each stimulus condition, computed as the mean rate from 25 to 1,000 ms following stimulus onset, minus the mean baseline rate in the 500 ms before stimulus onset. The mean SF and TF tuning curves for all cells within gaze-shift clusters were calculated using each cell's mean evoked firing rate for each SF/TF normalized by the maximum response to its best SF/TF.

Each cell's weighted TF response, W_{TF}, was determined using the mean evoked responses, R, for each of the two presented TFs and at the cell's optimal SF and preferred orientation.

$$W_{TF} = (2 \times R_{low} + 8 \times R_{high})/(R_{low} + R_{high})$$

Likewise, weighted SF response was calculated at the cell's weighted TF response and preferred orientation using the mean evoked responses for each of the three presented SFs.

$$W_{SF} = (0.02 \times R_{low} + 0.08 \times R_{mid} + 0.32 \times R_{high})/(R_{low} + R_{high})$$

4.6.6 Response clusters

To compare visual response properties of sSC to V1 we clustered units by following methods laid out in Parker et al., 2023. Briefly, we clustered sSC units with similar gaze-shift responses by performing *k*-means cluster ($k = 5$), resulting in 4 clusters of responsive units and one unresponsive cluster. We used the same clustering metrics when comparing sSC light/dark experiments.

4.6.7 Statistical analysis

Statistical tests across cell types were performed using ANOVA with correction for multiple comparisons. Correlations between latency and response were tested using Spearman's correlation coefficient. Positive and negative SF and TF responses were tested using Wilcoxon Rank-Sum tests. Comparison between response populations in V1 and sSC were tested using chi-squared test. No statistical methods were used to predetermine sample sizes but our sample size is similar to previous publications (Liska et al., 2023; Parker et al., 2023).

4.7 SUPPLEMENTAL FIGURES

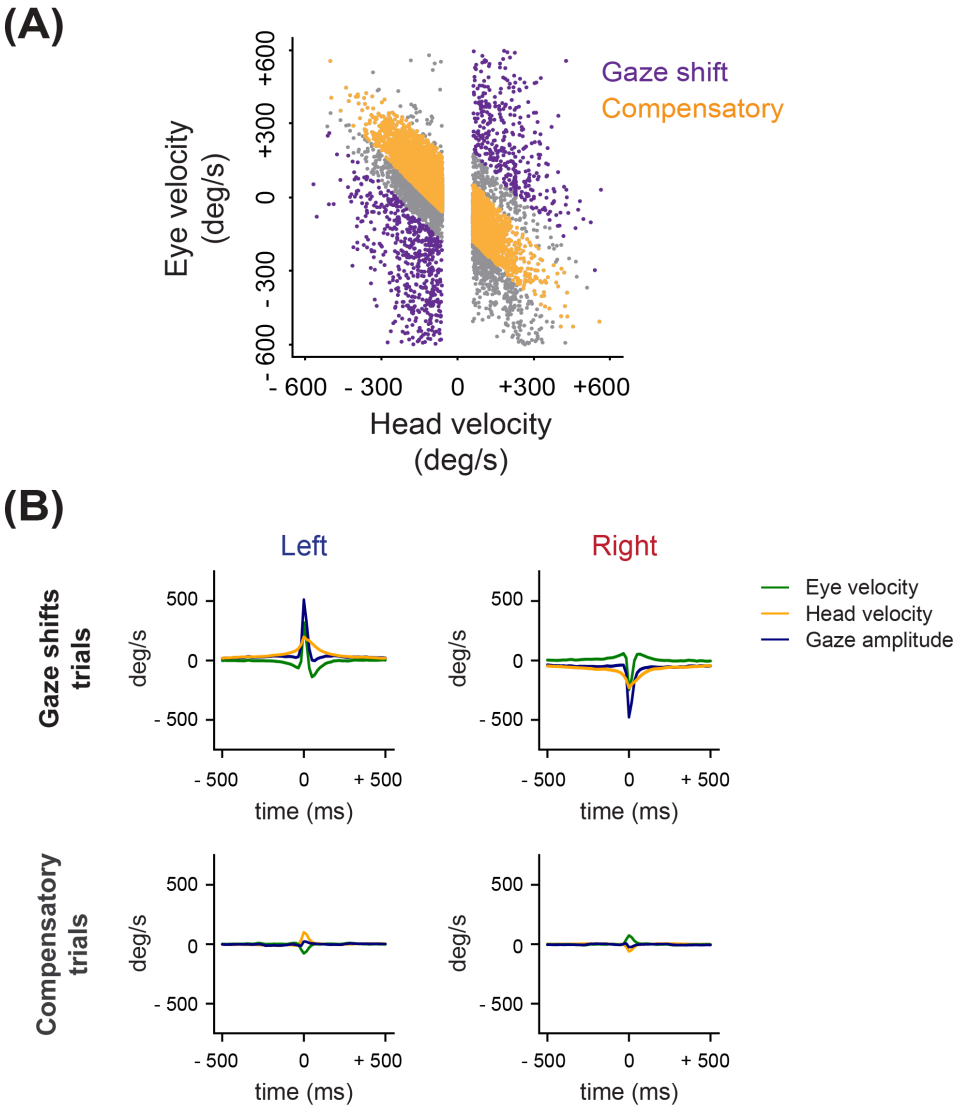


Figure S1. Characteristics of free movement. (A) Scatter plot of eye and head velocities subsampled (25x) from recordings showing compensatory, gaze-shifting, and intermediate movements, the latter of which are excluded from analysis in main text. **(B)** Mean head and eye velocity traces for left and right movements in compensatory and gaze shifting trials. Mean gaze amplitude responses are also plotted for compensatory and gaze shifting trials.

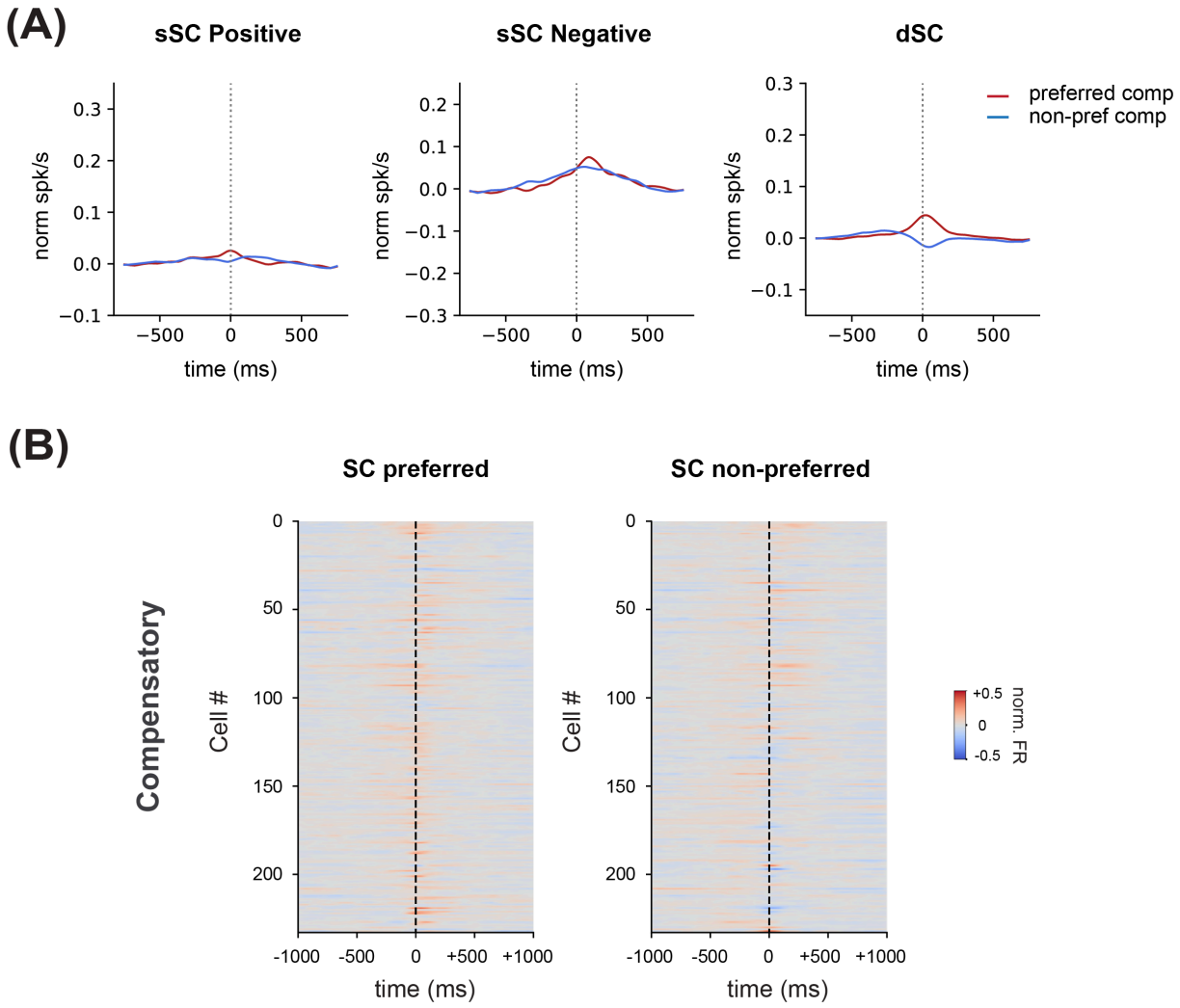


Figure S2. Compensatory responses in SC. **(A)** Mean PETHs for each response type in SC during preferred (red) and non-preferred (blue) compensatory head/eye movement. **(B)** Temporal sequence of SC responses to compensatory head/eye movements sorted by depth. Dashed line indicates onset of compensatory movement.

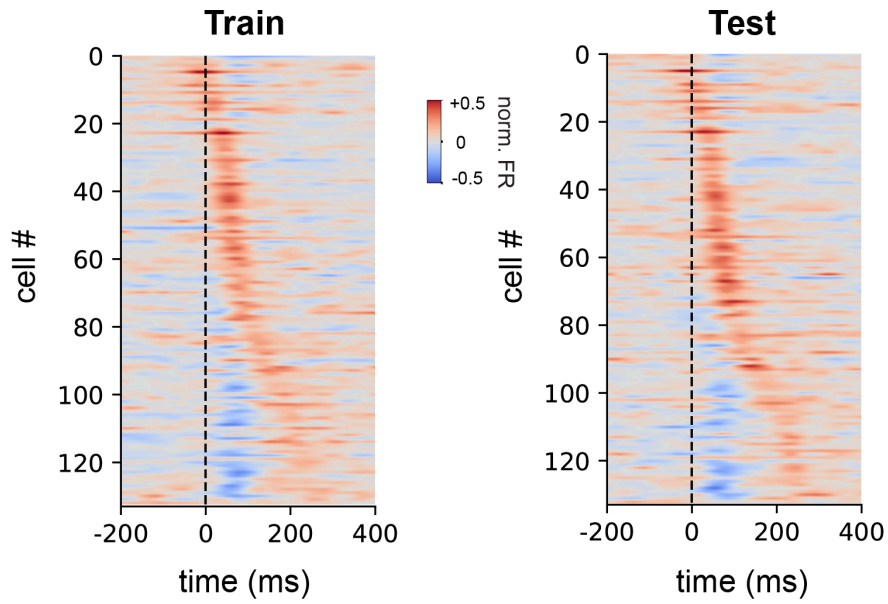


Figure S3. Cross-validation. Cross-validation of gaze shift PETHs of all responsive cells for all responsive cells in sSC. Gaze shift times were randomly divided into two sets used to calculate PETHs in both the train (left) and test (right) sets. The test set was sorted by the latency of the peak positive response in the train set.

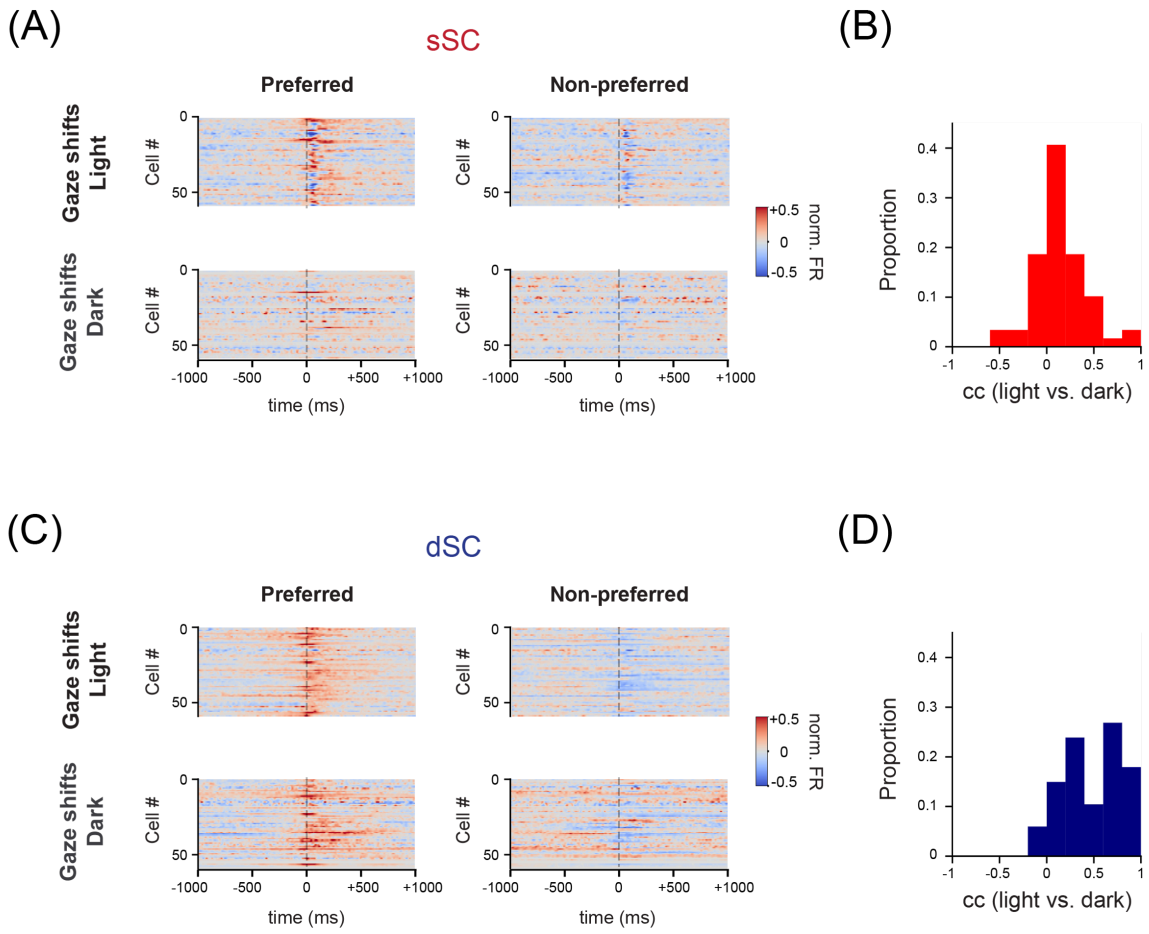


Figure S4. sSC and dSC preferred and non-preferred responses in light and dark (A) Temporal sequence plots for gaze shifts in light (top) and dark (bottom) for preferred and non-preferred gaze shift movements. **(B)** Cross correlation for sSC responses in light and dark showing little correlation between light and dark responses. **(C)** Same as **A** but for dSC cells. **(D)** Cross correlation of dSC responses between light and dark, showing high correlation between light and dark responses.

CHAPTER V
CONCLUSIONS

Together, my work has examined visual processing during active vision in free moving conditions during both task related and spontaneous visual exploration. Additionally, these projects have explored differences in active vision during gaze shifts between the two established parallel visual pathways: cortex and colliculus. As a whole the driving questions behind these projects were formulated to better our understanding of how naturalistic movement during unrestrained behavior influences visual processing. In Chapter II, we explored this utilizing an an ethologically relevant jumping task in which mice incorporated visual cues with self motion to accurately estimate distance. In Chapter III, we investigated the influence of gaze-shifting head and eye movements on visual processing in V1, where we found a dynamic temporal sequence response initiated by gaze shifts. Finally, in Chapter IV we studied the impact of head and eye movements during gaze shifts on a different brain region, SC, focused on uncovering gaze shift response dynamics during active vision across depth in SC. While all three of these projects were conducted separately, each contributed to answering a long standing question in the field of visual neuroscience: how does naturalistic movement impact visual processing. Altogether these results have created a deeper understanding of the neural underpinnings of active vision during free movement and highlight the importance of future work aimed at studying vision in unrestrained experimental paradigms.

This work lays the foundation for future studies of visual processing during naturalistic behaviors. In particular, our work in Chapter II shows that mice can perform complex tasks such as accurate gap-distance jumping with tethers and a head-mounted camera system. This is an important cornerstone that opens up the possibility to incorporate simultaneous neural recordings in tasks such as this one. Future studies involving chronic electrophysiology can dissect the role of V1 in distance estimation. Further, this variable gap-distance jumping task can be used to study neural underpinnings of many behaviors beyond distance estimation such as motor commands, reward based learning, and motivation.

Our work also opens up new areas of inquiry into neural coding of active vision. In Chapters III and IV we demonstrated novel principles of visual processing, based on the temporal dynamics and neural coding properties of neurons around gaze shifts, in two different

brain areas. This raises a number of new questions about the mechanisms and implications of these computational principles. For example, what circuitry underlies the gaze-shift initiated temporal sequence we find in V1 and sSC? One could begin to answer these questions by looking at cell type specific responses in V1 and sSC during free movement. Additionally, it is unknown how altering cell types responses through excitation or inhibition could influence these response dynamics. One particularly interesting finding from our work are the shared coarse-to-fine processing sequences in V1 and sSC despite the regions having significantly different response type distributions to gaze shifts. Further research into the circuit dynamics in each brain region are necessary to deduce why these brain regions appear to process visual input in similar ways but with different population dynamics. Additionally, our experiments in SC across depth probe future questions about layer specific roles in SC during naturalistic movements. For instance, what role does each layer of SC: superficial, intermediate, deep, play during visuomotor behaviors such as prey capture or escape? Using chronic electrophysiology one could measure layer specific responses during all steps of prey capture or escape behavior to dissect layer and cell type specific roles. Additionally, our experimental technique could be applied to study other visual tasks such as distraction or directed attention in a wide variety of brain regions such as V1, SC, dLGN, and higher visual areas.

Altogether these projects add to our understanding of dynamic visual processing during unrestrained movement and open the door for future works exploring vision in more naturalistic contexts. The influence of self-motion and free movement on visual processing is not well understood but by learning more about the neural computations that drive active vision we can better understand not only active sensation but how the brain functions as a whole.

REFERENCES CITED

Adey, W. R., & Noda, H. (1973). Influence of eye movements on geniculo‐striate excitability in the cat. *The Journal of Physiology*, *235*(3), 805-821.

Ahissar, E., & Arieli, A. (2001). Figuring space by time. *Neuron*, *32*(2), 185-201.

Allen, E. A., & Freeman, R. D. (2006). Dynamic spatial processing originates in early visual pathways. *Journal of Neuroscience*, *26*(45), 11763-11774.

Angelaki, D. E., & Cullen, K. E. (2008). Vestibular system: the many facets of a multimodal sense. *Annu. Rev. Neurosci.*, *31*(1), 125-150.

Barik, A., Thompson, J. H., Seltzer, M., Ghitani, N., & Chesler, A. T. (2018). A brainstem-spinal circuit controlling nocifensive behavior. *Neuron*, *100*(6), 1491-1503.

Basso, M. A., & May, P. J. (2017). Circuits for action and cognition: a view from the superior colliculus. *Annual review of vision science*, *3*(1), 197-226.

Basso, M. A., Bickford, M. E., & Cang, J. (2021). Unraveling circuits of visual perception and cognition through the superior colliculus. *Neuron*, *109*(6), 918-937.

Beck, J. M., Ma, W. J., Pitkow, X., Latham, P. E., & Pouget, A. (2012). Not noisy, just wrong: the role of suboptimal inference in behavioral variability. *Neuron*, *74*(1), 30-39.

Benardete, E. A., & Kaplan, E. (1999). The dynamics of primate M retinal ganglion cells. *Visual neuroscience, 16*(2), 355-368.

Bianco, I. H., Kampff, A. R., & Engert, F. (2011). Prey capture behavior evoked by simple visual stimuli in larval zebrafish. *Frontiers in systems neuroscience*, *5*, 101.

Biswas, D., Arend, L. A., Stamper, S. A., Vágvölgyi, B. P., Fortune, E. S., & Cowan, N. J. (2018). Closed-loop control of active sensing movements regulates sensory slip. *Current Biology*, *28*(24), 4029-4036.

Bittencourt, A. S., Nakamura-Palacios, E. M., Mauad, H., Tufik, S., & Schenberg, L. C. (2005). Organization of electrically and chemically evoked defensive behaviors within the deeper collicular layers as compared to the periaqueductal gray matter of the rat. *Neuroscience, 133*(4), 873-892.

Bleckert, A., Schwartz, G. W., Turner, M. H., Rieke, F., & Wong, R. O. (2014). Visual space is represented by nonmatching topographies of distinct mouse retinal ganglion cell types. *Current Biology*, *24*(3), 310-315.

Boi, M., Poletti, M., Victor, J. D., & Rucci, M. (2017). Consequences of the oculomotor cycle for the dynamics of perception. *Current Biology*, *27*(9), 1268-1277.

Boone, H. C., Samonds, J. M., Crouse, E. C., Barr, C., Priebe, N. J., & McGee, A. W. (2021). Natural binocular depth discrimination behavior in mice explained by visual cortical activity. *Current Biology*, *31*(10), 2191-2198.

Bouvier, G., Senzai, Y., & Scanziani, M. (2020). Head movements control the activity of primary visual cortex in a luminance-dependent manner. *Neuron*, *108*(3), 500-511.

Bredfeldt, C. E., & Ringach, D. L. (2002). Dynamics of spatial frequency tuning in macaque V1. *Journal of Neuroscience*, *22*(5), 1976-1984.

Burchfiel, J. L., & Duffy, F. H. (1974). Corticofugal influence upon cat thalamic ventrobasal complex. *Brain Research*, *70*(3), 395-411.

Campagner, D., Vale, R., Tan, Y. L., Iordanidou, P., Pavón Arocas, O., Claudi, F., ... & Branco, T. (2023). A cortico-collicular circuit for orienting to shelter during escape. *Nature*, *613*(7942), 111-119.

Cang, J., Savier, E., Barchini, J., & Liu, X. (2018). Visual function, organization, and development of the mouse superior colliculus. *Annual review of vision science, 4*(1), 239-262.

Carey, D. P., Goodale, M. A., & Sprowl, E. G. (1990). Blindsight in rodents: the use of a 'high-level'distance cue in gerbils with lesions of primary visual cortex. *Behavioural brain research*, *38*(3), 283-289.

Cazemier, J. L., Haak, R., Tran, T. L., Hsu, A. T., Husic, M., Peri, B. D., ... & Heimel, J. A. (2024). Involvement of superior colliculus in complex figure detection of mice. *eLife, 13,* e83708.

Chen, Y., & Qian, N. (2004). A coarse-to-fine disparity energy model with both phase-shift and position-shift receptive field mechanisms. *Neural computation*, *16*(8), 1545-1577.

Cisek, P. (1999). Beyond the computer metaphor: Behaviour as interaction. *Journal of Consciousness Studies*, *6*(11-12), 125-142.

Cloherty, S. L., Yates, J. L., Graf, D., DeAngelis, G. C., & Mitchell, J. F. (2020). Motion perception in the common marmoset. *Cerebral Cortex, 30*(4), 2659-2673.

Dayan, P., & Abbott, L. F. (2005). Theoretical neuroscience: computational and mathematical modeling of neural systems. *MIT press*.

DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *Journal of neurophysiology, 69*(4), 1091-1117.

DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1995). Receptive-field dynamics in the central visual pathways. *Trends in neurosciences, 18*(10), 451-458.

Doubell, T. P., Skaliora, I., Baron, J., & King, A. J. (2003). Functional connectivity between the superficial and deeper layers of the superior colliculus: an anatomical substrate for sensorimotor integration. *Journal of Neuroscience, 23*(16), 6596-6607.

Doykos, T. K., Gilmer, J. I., Person, A. L., & Felsen, G. (2020). Monosynaptic inputs to specific cell types of the intermediate and deep layers of the superior colliculus. *Journal of Comparative Neurology, 528*(13), 2254-2268.

Duffy, F. H., & Burchfiel, J. L. (1975). Eye movement-related inhibition of primate visual neurons. *Brain Research*, *89*(1), 121-132.

Durgin, F. H., Proffitt, D. R., Olson, T. J., & Reinke, K. S. (1995). Comparing depth from motion with depth from binocular disparity. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 679.

Eastman, K. M., & Huk, A. C. (2012). PLDAPS: a hardware architecture and software toolbox for neurophysiology requiring complex visual stimuli and online behavioral control. *Frontiers in neuroinformatics, 6*, 1.

Ellard, C. G., Goodale, M. A., & Timney, B. (1984). Distance estimation in the Mongolian gerbil: The role of dynamic depth cues. *Behavioural brain research*, *14*(1), 29-39.

Ellard, C. G., Goodale, M. A., Scorfield, D. M., & Lawrence, C. (1986). Visual cortical lesions abolish the use of motion parallax in the Mongolian gerbil. *Experimental brain research*, *64*, 599-602.

Ellard, C. G., & Sharma, L. (1996). The effects of cortical lesions on recognition of object context in a visuomotor task in the Mongolian gerbil. *Behavioural Brain Research*, *82*(1), 13-22.

Ellis, E. M., Gauvain, G., Sivyer, B., & Murphy, G. J. (2016). Shared and distinct retinal input to the mouse superior colliculus and dorsal lateral geniculate nucleus. *Journal of neurophysiology*, *116*(2), 602-610.

Fajen, B. R. (2021). Visual control of locomotion. *Cambridge University Press*.

Fortey, R. A., Briggs, D. E., & Wills, M. A. (1996). The Cambrian evolutionary 'explosion': decoupling cladogenesis from morphological disparity. *Biological Journal of the Linnean Society*, *57*(1), 13-33.

Gale, S. D., & Murphy, G. J. (2014). Distinct representation and distribution of visual information by specific cell types in mouse superficial superior colliculus. *Journal of Neuroscience*, *34*(40), 13458-13471.

Gallant, J. L., Connor, C. E., & Van Essen, D. C. (1998). Neural activity in areas V1, V2 and V4 during free viewing of natural scenes compared to controlled viewing. *Neuroreport*, *9*(7), 1673-1678.

Gandhi, N. J., & Katnani, H. A. (2011). Motor functions of the superior colliculus. *Annual review of neuroscience, 34*(1), 205-231.

Gibson, J. J. (1979). The ecological approach to visual perception: classic edition. *Psychology press*.

Giovannetti, E. A., & Rancz, E. (2024). Behind mouse eyes: the function and control of eye movements in mice. *Neuroscience & Biobehavioral Reviews*, 105671.

Gao, E., DeAngelis, G. C., & Burkhalter, A. (2010). Parallel input channels to mouse primary visual cortex. *Journal of Neuroscience*, *30*(17), 5912-5926.

Goodale, M. A., Ellard, C. G., & Booth, L. (1990). The role of image size and retinal motion in the computation of absolute distance by the Mongolian gerbil (Meriones unguiculatus). *Vision research*, *30*(3), 399-413.

Guitchounts, G., Masís, J., Wolff, S. B., & Cox, D. (2020). Encoding of 3D head orienting movements in the primary visual cortex. *Neuron*, *108*(3), 512-525.

Hegdé, J. (2008). Time course of visual perception: coarse-to-fine processing and beyond. *Progress in neurobiology*, *84*(4), 405-439.

Henschke, J. U., Dylda, E., Katsanevaki, D., Dupuy, N., Currie, S. P., Amvrosiadis, T., ... & Rochefort, N. L. (2020). Reward association enhances stimulus-specific representations in primary visual cortex. *Current Biology, 30*(10), 1866-1880.

Hippenmeyer, S., Vrieseling, E., Sigrist, M., Portmann, T., Laengle, C., Ladle, D. R., & Arber, S. (2005). A developmental switch in the response of DRG neurons to ETS transcription factor signaling. *PLoS biology*, *3*(5), e159.

Holmgren, C. D., Stahr, P., Wallace, D. J., Voit, K. M., Matheson, E. J., Sawinski, J., ... & Kerr, J. N. (2021). Visual pursuit behavior in mice maintains the pursued prey on the retinal region with least optic flow. *eLife*, *10*, e70838.

Hoy, J. L., Yavorska, I., Wehr, M., & Niell, C. M. (2016). Vision drives accurate approach behavior during prey capture in laboratory mice. *Current Biology*, *26*(22), 3046-3052.

Hoy, J. L., Bishop, H. I., & Niell, C. M. (2019). Defined cell types in superior colliculus make distinct contributions to prey capture behavior in the mouse. *Current Biology*, *29*(23), 4130-4138.

Huberman, A. D., & Niell, C. M. (2011). What can mice tell us about how vision works?. *Trends in neurosciences*, *34*(9), 464-473.

Hunt, N. H., Jinn, J., Jacobs, L. F., & Full, R. J. (2021). Acrobatic squirrels learn to leap and land on tree branches without falling. *Science*, *373*(6555), 697-700.

Hutson, K. A., & Masterton, R. B. (1986). The sensory contribution of a single vibrissa's cortical barrel. *Journal of neurophysiology*, *56*(4), 1196-1223.

Ibbotson, M., & Krekelberg, B. (2011). Visual perception and saccadic eye movements. *Current opinion in neurobiology*, *21*(4), 553-558.

Idrees, S., Baumann, M. P., Korympidou, M. M., Schubert, T., Kling, A., Franke, K., ... & Münch, T. A. (2022). Suppression without inhibition: how retinal computation contributes to saccadic suppression. *Communications Biology, 5*(1), 692.

Ito, S., Feldheim, D. A., & Litke, A. M. (2017). Segregation of visual response properties in the mouse superior colliculus and their modulation during locomotion. *Journal of Neuroscience, 37*(35), 8428-8443.

Kim, H. R., Angelaki, D. E., & DeAngelis, G. C. (2017). Gain modulation as a mechanism for coding depth from motion parallax in macaque area MT. *Journal of Neuroscience*, *37*(34), 8180-8197.

King, C. W., Ledochowitsch, P., Buice, M. A., & de Vries, S. E. (2023). Saccade-responsive visual cortical neurons do not exhibit distinct visual response properties. *eNeuro*, *10*(9).

Kleiner, M., Brainard, D. H., Pelli, D. G., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3? *Perception, 36* (ECVP Abstract Suppl.).), 14.

Kral, K. (2003). Behavioural–analytical studies of the role of head movements in depth perception in insects, birds and mammals. *Behavioural Processes*, *64*(1), 1-12.

Land, M. F. (1999). Motion and vision: why animals move their eyes. *Journal of Comparative Physiology A*, *185*, 341-352.

Land, M. (2019). Eye movements in man and other animals. *Vision research*, *162*, 1-7.

Legg, C. R., & Lambert, S. (1990). Distance estimation in the hooded rat: experimental evidence for the role of motion cues. *Behavioural brain research*, *41*(1), 11-20.

Leopold, D. A., & Logothetis, N. K. (1998). Microsaccades differentially modulate neural activity in the striate and extrastriate visual cortex. *Experimental Brain Research*, *123*, 341-345.

Leopold, D. A., & Park, S. H. (2020). Studying the visual brain in its natural rhythm. *Neuroimage*, *216*, 116790.

Li, N., Chen, S., Guo, Z. V., Chen, H., Huo, Y., Inagaki, H. K., ... & Svoboda, K. (2019). Spatiotemporal constraints on optogenetic inactivation in cortical circuits. *eLife*, *8*, e48622.

Linderman, S., Nichols, A., Blei, D., Zimmer, M., & Paninski, L. (2019). Hierarchical recurrent state space models reveal discrete and continuous dynamics of neural activity in C. elegans. *BioRxiv*, 621540.

Liska, J. P., Rowley, D. P., Nguyen, T. T., Muthmann, J. O., Butts, D. A., Yates, J. L., & Huk, A. C. (2022). Running modulates primate and rodent visual cortex via common mechanism but quantitatively distinct implementation. *bioRxiv*, 2022-06.

Liu, X., Huang, H., Snutch, T. P., Cao, P., Wang, L., & Wang, F. (2022). The superior colliculus: cell types, connectivity, and behavior. *Neuroscience Bulletin*, *38*(12), 1519-1540.

Lopes, G., Bonacchi, N., Frazão, J., Neto, J. P., Atallah, B. V., Soares, S., ... & Kampff, A. R. (2015). Bonsai: an event-based framework for processing and controlling data streams. *Frontiers in neuroinformatics*, *9*, 7.

Madisen, L., Mao, T., Koch, H., Zhuo, J. M., Berenyi, A., Fujisawa, S., ... & Zeng, H. (2012). A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. *Nature neuroscience*, *15*(5), 793-802.

Marr, D. (2010). Vision: A computational investigation into the human representation and processing of visual information. *MIT press*.

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature neuroscience, 21*(9), 1281-1289.

Mazer, J. A., Vinje, W. E., McDermott, J., Schiller, P. H., & Gallant, J. L. (2002). Spatial frequency and orientation tuning dynamics in area V1. *Proceedings of the National Academy of Sciences*, *99*(3), 1645-1650.

Medrea, I., & Cullen, K. E. (2013). Multisensory integration in early vestibular processing in mice: the encoding of passive vs. active motion. *Journal of neurophysiology*, *110*(12), 2704-2717.

Meyer, A. F., Poort, J., O'Keefe, J., Sahani, M., & Linden, J. F. (2018). A head-mounted camera system integrates detailed behavioral monitoring with multichannel electrophysiology in freely moving mice. *Neuron*, *100*(1), 46-60.

Meyer, A. F., O'Keefe, J., & Poort, J. (2020). Two distinct types of eye-head coupling in freely moving mice. *Current Biology*, *30*(11), 2116-2130.

Michaiel, A. M., Abe, E. T., & Niell, C. M. (2020). Dynamics of gaze control during prey capture in freely moving mice. *eLife*, *9*, e57458.

Mitchell, J. F., Reynolds, J. H., & Miller, C. T. (2014). Active vision in marmosets: a model system for visual neuroscience. *Journal of Neuroscience, 34*(4), 1183-1194.

Miura, S. K., & Scanziani, M. (2022). Distinguishing externally from saccade-induced motion in visual cortex. *Nature*, *610*(7930), 135-142.

Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S., & Churchland, A. K. (2019). Single-trial neural dynamics are dominated by richly varied movements. *Nature neuroscience*, *22*(10), 1677-1686.

Nadler, J. W., Angelaki, D. E., & DeAngelis, G. C. (2008). A neural representation of depth from motion parallax in macaque visual cortex. *Nature*, *452*(7187), 642-645.

Nadler, J. W., Nawrot, M., Angelaki, D. E., & DeAngelis, G. C. (2009). MT neurons combine visual motion with a smooth eye movement signal to code depth-sign from motion parallax. *Neuron*, *63*(4), 523-532.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive psychology*, *9*(3), 353-383.

Neri, P. (2011). Coarse to fine dynamics of monocular and binocular processing in human pattern vision. *Proceedings of the National Academy of Sciences, 108*(26), 10726-10731.

Niell, C. M., & Stryker, M. P. (2008). Highly selective receptive fields in mouse visual cortex. *Journal of Neuroscience, 28*(30), 7520-7536.

Niell, C. M., & Stryker, M. P. (2010). Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron*, *65*(4), 472-479.

Niell, C. M., & Scanziani, M. (2021). How cortical circuits implement cortical computations: mouse visual cortex as a model. *Annual Review of Neuroscience, 44*(1), 517-546.

niellab. (2022). Nlab-behavior. b0ea1fc. *GitHub*.

Nishimoto, S., Huth, A. G., Bilenko, N. Y., & Gallant, J. L. (2017). Eye movement-invariant representations in the human visual system. *Journal of vision*, *17*(1), 11-11.

Noda, H., & Adey, W. R. (1974). Retinal ganglion cells of the cat transfer information on saccadic eye movement and quick target motion. *Brain research*, *70*(2), 340-345.

Nummela, S. U., Coop, S. H., Cloherty, S. L., Boisvert, C. J., Leblanc, M., & Mitchell, J. F. (2017). Psychophysical measurement of marmoset acuity and myopia. *Developmental Neurobiology, 77*(3), 300-313.

Odoemene, O., Pisupati, S., Nguyen, H., & Churchland, A. K. (2018). Visual evidence accumulation guides decision-making in unrestrained mice. *Journal of Neuroscience*, *38*(47), 10143-10155.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive psychology*, *34*(1), 72-107.

Parker, P. R., Brown, M. A., Smear, M. C., & Niell, C. M. (2020). Movement-related signals in sensory areas: roles in natural behavior. *Trends in neurosciences*, *43*(8), 581-595.

Parker, P. R., Abe, E. T., Beatie, N. T., Leonard, E. S., Martins, D. M., Sharp, S. L., ... & Niell, C. M. (2022). Distance estimation from monocular cues in an ethological visuomotor task. *eLife*, *11*, e74708.

Parker, P. R., Abe, E. T., Leonard, E. S., Martins, D. M., & Niell, C. M. (2022). Joint coding of visual input and eye/head position in V1 of freely moving mice. *Neuron*, *110*(23), 3897-3906.

Parker, P. R., Martins, D. M., Leonard, E. S., Casey, N. M., Sharp, S. L., Abe, E. T., ... & Niell, C. M. (2023). A dynamic sequence of visual processing initiated by gaze shifts. *Nature Neuroscience*, *26*(12), 2192-2202.

Perry, V. H., & Cowey, A. (1984). Retinal ganglion cells that project to the superior colliculus and pretectum in the macaque monkey. *Neuroscience*, *12*(4), 1125-1137.

Petras, K., Ten Oever, S., Jacobs, C., & Goffaux, V. (2019). Coarse-to-fine information integration in human vision. *NeuroImage*, *186*, 103-112.

Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., ... & Vuilleumier, P. (2010). The neural substrates and timing of top–down processes during coarse-to-fine categorization of visual scenes: A combined fMRI and ERP study. *Journal of cognitive neuroscience, 22*(12), 2768-2780.

Piscopo, D. M., El-Danaf, R. N., Huberman, A. D., & Niell, C. M. (2013). Diverse visual features encoded in mouse lateral geniculate nucleus. *Journal of Neuroscience*, *33*(11), 4642-4656.

Purushothaman, G., Chen, X., Yampolsky, D., & Casagrande, V. A. (2014). Neural mechanisms of coarse-to-fine discrimination in the visual cortex. *Journal of Neurophysiology*, *112*(11), 2822-2833.

Richardson, F. (1909). A study of the sensory control in the rat. *The Psychological Review: Monograph Supplement*s, *12*(1), i.

Ringach, D. L., Sapiro, G., & Shapley, R. (1997). A subspace reverse-correlation technique for the study of visual neurons. *Vision research*, *37*(17), 2455-2464.

Rogers, B., & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception*, *8*(2), 125-134.

Rucci, M., Iovin, R., Poletti, M., & Santini, F. (2007). Miniature eye movements enhance fine spatial detail. *Nature*, *447*(7146), 852-855.

Sattler, N. J., & Wehr, M. (2021). A head-mounted multi-camera system for electrophysiology and behavior in freely-moving mice. *Frontiers in neuroscience*, *14*, 592417.

Savier, E. L., Chen, H., & Cang, J. (2019). Effects of locomotion on visual responses in the mouse superior colliculus. *Journal of Neuroscience*, *39*(47), 9360-9368.

Schiller, P. H., & Stryker, M. (1972). Single-unit recording and stimulation in superior colliculus of the alert rhesus monkey. *Journal of neurophysiology, 35*(6), 915-924.

Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Current opinion in neurobiology*, *20*(2), 172-176.

Schwab, I. R. (2018). The evolution of eyes: major steps. The Keeler lecture 2017: centenary of Keeler Ltd. *Eye*, *32*(2), 302-313.

Seabrook, T. A., Burbridge, T. J., Crair, M. C., & Huberman, A. D. (2017). Architecture, function, and assembly of the mouse visual system. *Annual review of neuroscience*, *40*(1), 499-538.

Senzai, Y., Fernandez-Ruiz, A., & Buzsáki, G. (2019). Layer-specific physiological features and interlaminar interactions in the primary visual cortex of the mouse. *Neuron, 101*(3), 500-513.

Senzai, Y., & Scanziani, M. (2022). A cognitive process occurring during sleep is revealed by rapid eye movements. *Science, 377*(6609), 999-1004.

Shang, C., Chen, Z., Liu, A., Li, Y., Zhang, J., Qu, B., ... & Cao, P. (2018). Divergent midbrain circuits orchestrate escape and freezing responses to looming stimuli in mice. *Nature communications*, *9*(1), 1232.

Shinkman PG. (1962). Visual depth discrimination in animals. *Psychological Bulletin 59*:489–501.

Skyberg, R., Tanabe, S., Chen, H., & Cang, J. (2022). Coarse-to-fine processing drives the efficient coding of natural scenes in mouse visual cortex. *Cell reports*, *38*(13).

Soltani, A., Rakhshan, M., Schafer, R. J., Burrows, B. E., & Moore, T. (2021). Separable influences of reward on visual processing and choice. *Journal of cognitive neuroscience, 33*(2), 248-262.

Sommer, M. A., & Wurtz, R. H. (2008). Brain circuits for the internal monitoring of movements. *Annu. Rev. Neurosci.*, *31*(1), 317-338.

Sousa, R., Brenner, E., & Smeets, J. B. (2011). Judging an unfamiliar object's distance from its retinal image size. *Journal of vision*, *11*(9), 10-10.

Spitler, K. M., & Gothard, K. M. (2008). A removable silicone elastomer seal reduces granulation tissue growth and maintains the sterility of recording chambers for primate neurophysiology. *Journal of neuroscience methods, 169*(1), 23-26.

Stamper, S. A., Madhav, M. S., Cowan, N. J., & Fortune, E. S. (2019). Using control theory to characterize active sensing in weakly electric fishes. *Electroreception: Fundamental Insights from Comparative Approaches*, 227-249.

Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C. B., Carandini, M., & Harris, K. D. (2019). Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, *364*(6437), eaav7893.

Toyama, K., Kimura, M., & Komatsu, Y. (1984). Activity of the striate cortex cells during saccadic eye movements of the alert cat. *Neuroscience Research*, *1*(4), 207-222.

van Beest, E. H., Mukherjee, S., Kirchberger, L., Schnabel, U. H., van der Togt, C., Teeuwen, R. R., ... & Self, M. W. (2021). Mouse visual cortex contains a region of enhanced spatial resolution. *Nature communications*, *12*(1), 4029.

Vélez-Fort, M., Bracey, E. F., Keshavarzi, S., Rousseau, C. V., Cossell, L., Lenzi, S. C., ... & Margrie, T. W. (2018). A circuit for integration of head-and visual-motion signals in layer 6 of mouse primary visual cortex. *Neuron*, *98*(1), 179-191.

Vreysen, S., Zhang, B., Chino, Y. M., Arckens, L., & Van den Bergh, G. (2012). Dynamics of spatial frequency tuning in mouse visual cortex. *Journal of neurophysiology*, *107*(11), 2937-2949.

Wallace, G. K. (1959). Visual scanning in the desert locust Schistocerca gregaria Forskål. *Journal of Experimental Biology*, *36*(3), 512-525.

Wallace, D. J., Greenberg, D. S., Sawinski, J., Rulla, S., Notaro, G., & Kerr, J. N. (2013). Rats maintain an overhead binocular field at the expense of constant fusion. *Nature*, *498*(7452), 65-69.

Wang, L., Sarnaik, R., Rangarajan, K., Liu, X., & Cang, J. (2010). Visual receptive field properties of neurons in the superficial superior colliculus of the mouse. *Journal of Neuroscience, 30*(49), 16573-16584.

Wang, L., Chen, I. Z., & Lin, D. (2015). Collateral pathways from the ventromedial hypothalamus mediate defensive behaviors. *Neuron*, *85*(6), 1344-1358.

Wekselblatt, J.B., Flister, E.D., Piscopo, D.M., & Niell, C.M. (2016). Large-scale imaging of cortical dynamics during sensory perception and behavior. *Journal of Neurophysiology. 11*(6), 2852-2866.

Wilson, J. J., Alexandre, N., Trentin, C., & Tripodi, M. (2018). Three-dimensional representation of motor space in the mouse superior colliculus. *Current Biology*, *28*(11), 1744-1755.

Wurtz, R. H., & Mohler, C. W. (1976). Organization of monkey superior colliculus: enhanced visual response of superficial layer cells. *Journal of Neurophysiology, 39*(4), 745-765.

Yates, J. L., Coop, S. H., Sarch, G. H., Wu, R. J., Butts, D. A., Rucci, M., & Mitchell, J. F. (2023). Detailed characterization of neural selectivity in free viewing primates. *Nature Communications, 14*(1), 3656.

Yilmaz, M., & Meister, M. (2013). Rapid innate defensive responses of mice to looming visual stimuli. *Current Biology*, *23*(20), 2011-2015.

Yu, Y., Hira, R., Stirman, J. N., Yu, W., Smith, I. T., & Smith, S. L. (2018). Mice use robust and common strategies to discriminate natural scenes. *Scientific reports*, *8*(1), 1379.

Zahler, S. H., Taylor, D. E., Wong, J. Y., Adams, J. M., & Feinberg, E. H. (2021). Superior colliculus drives stimulus-evoked directionally biased saccades and attempted head movements in head-fixed mice. *eLife*, *10*, e73081.

Zhao, Z. D., Zhang, L., Xiang, X., Kim, D., Li, H., Cao, P., & Shen, W. L. (2023). Neurocircuitry of predatory hunting. *Neuroscience Bulletin*, *39*(5), 817-831.