REPRESENTATIONS OF ACTIVE VISION

by

ELLIOTT TAYLOR TSUYOSHI ABE

A DISSERTATION

Presented to the Department of Biology
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctorate of Philosophy

March 2023

DISSERTATION APPROVAL PAGE

Student: Elliott Taylor Tsuyoshi Abe

Title: Representations of Active Vision

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctorate of Philosophy degree in the Department of Biology by:

Luca Mazzucato	Chair
Cristopher Niell	Advisor
James Murray	Advisor
Timothy Gardner	Core Member
Margeret Sereno	Core Member
Matthew Smear	Institutional Representative

and

Krista Chronister	Vice Provost of Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded March 2023

DISSERTATION ABSTRACT

Elliott Taylor Tsuyoshi Abe

Doctorate of Philosophy

Department of Biology

March 2023

Title: Representations of Active Vision

This dissertation focuses on the interplay between visual processing and motor action during natural behaviors, which has previously been limited due to technological constraints in experimental paradigms. However, recent technological innovations have improved the data collection process, enabling a better understanding of visual processing under naturalistic conditions. This dissertation lays out the foundational experimental methods, data analysis, and theoretical modeling to study visual processing during natural behaviors.

Chapter II of the dissertation establishes and characterizes how mice change their gaze during prey capture behavior using a miniaturized camera system to simultaneously record the eyes and head as the mice captured crickets. The study finds that there are two types of eye movements during prey capture. The majority of eye movements are compencatory, however there is a subset that shift the gaze of the mouse and are initiated due to head movements in a 'saccade and fixate' strategy.

Chapter III, expands upon the previous methods and records neural activity, eye position, head orientation, and visual scene simultaneously while mice freely explore an arena. The data is used to create a model to correct the visual scene for gaze position, enabling the mapping of the first visual receptive fields in a free-moving animal. The study discovers neurons in the primary visual cortex that are tuned to eye position and head orientation, with most cells integrating positional and visual information through a multiplicative gain modulation mechanism.

Chapter IV explores mechanisms for computing higher-order visual representations, like distance estimation, from predictions. The study creates a simulated environment where an agent records visual scene, depth maps, and positional information while navigating an arena. A deep convolutional recurrent neural network is trained on the visual scene and tasked with predicting future visual input. Post-training, the study is able to linearly decode the pixel-wise distance of the visual scene without explicit distance information. This work establishes that predictive processing is a viable mechanism for the visual system to learn to create higher-order visual representations without explicit training.

This dissertation consists of previously published co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Elliott Taylor Tsuyoshi Abe

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR, USA
University of Washington, Seattle, WA, USA

DEGREES AWARDED:

Doctor of Philosophy, Biology, 2023, University of Oregon
Bachelor of Science, Physics, 2017, University of Washington

AREAS OF SPECIAL INTEREST:

Computational and Theoretical Neuroscience
Deep Learning
Reinforcement Learning
Vision

PROFESSIONAL EXPERIENCE:

Teaching Assistent, University of Oregon, 2017-2022
Research Assistent, University of Oregon, 2017-2023

GRANTS, AWARDS AND HONORS:

Donald Wimber Fund, University of Oregon, 2020

WRF Innovation Fellowship in Neuroengineering, Washington Research Foundation, University of Washington, 2016

Computational Neuroscience Research Stipend, Computational Neuroscience Training, University of Washington, 2016

Washington Research Foundation Fellowship, Washington Research Foundation, University of Washington 2015

PUBLICATIONS:

*Parker, P. R. L., *Martins, D. M., *Leonard, E. S. P., Casey, N. M., Sharp, S. L., **Abe, E. T. T.**, Smear, M. C., Yates, J. L., Mitchel, J. F., Niell, C. M. (2022). A dynamic sequence of visual processing initiated by gaze shifts. BioRxiv. https://doi.org/10.1101/2022.08.23.504847

*Parker, P. R. L., ***Abe, E. T. T.**, Leonard, E. S. P., Martins, D. M., Niell, C. M. (2022). Joint coding of visual input and eye/head position in V1 of freely moving mice. Neuron. https://doi.org/10.1016/j.neuron.2022.08.029

Parker, P. R. L., **Abe, E. T. T.**, Beatie, N. T., Leonard, E. S. P., Martins, D. M., Sharp, S. L., Wyrick, D. G., Mazzucato, L., Niell, C. M. (2021). Distance estimation from monocular cues in an ethological visuomotor task. eLife 11:e74708. https://doi.org/10.7554/eLife.74708

Michaiel, A. M., **Abe, E. T. T.**, Niell, C. M. (2020). Dynamics of gaze control during prey capture in freely moving mice. ELife, 9, 1–18. https://doi.org/10.7554/elife.57458

*Duffy, A., ***Abe, E.**, Perkel, D. J., Fairhall, A. L. Variation in sequence dynamics improves maintenance of stereotyped behavior in an example from bird song. Proc. Natl. Acad. Sci. 201815910 (2019). doi:10.1073/pnas.1815910116

ACKNOWLEDGEMENTS

My graduate career has been enriched by the guidance and support of truly wonderful mentors. I would like to first and foremost thank my mentor Cris Niell for his constant support and guidance throughout my Ph.D. His natural curiosity and generosity to explore fascinating scientific questions has been an inspiration. Cris is always ready to hear my crazy scientific ideas and encourage me to pursue, explore and refine my research interests. Under his guidance, I have become a more rigorous, creative, and knowledgeable researcher. I would also like to thank James Murray, who for the last couple years during my Ph.D. has been co-advising me. With his careful, methodical, and patient guidance, my learning in theoretical modeling has been greatly enhanced, deepened and expanded. Additionally, I would like to express appreciation to Yashar Ahmadian, my former theory advisor. He guided me during the beginning of my Ph.D. and I learned many fascinating techniques and methods from him. A very grateful and enthusiastic Thank you to the other members of my committee as well – Lucca Mazzucato, Matt Smear, Tim Gardner, and Margaret Sereno – for their patient support throughout my learning process.

I would like to extend a special thank you to Philip Parker, who has been a steadfast friend, mentor, and scientific role-model for me, throughout my graduate career. His rigorousness and attention to detail, and the knowledge he exemplifies through his research, has greatly shaped my scientific thinking. Whether during those disappointing times of research when everything seems to be failing or enlivening times when there is immense creativity, Phil's encouraging support and mentorship throughout my graduate studies helped me develop my confidence as researcher. Outside of the lab he also introduced me to brewing beer and making wine, floating the Willamette, and fishing. He has become a lifelong friend and academic colleague. Additionally, a heartfelt thank you to the other members of the Niell lab over the years – Kristen Chauvin, Emmalyn Leonard, Dylan Martins, Michael Sidikpramana, Mea Songco, Nate Casey, Angie Michael, Mandi Servsons, Ian Etherington, Hannah Bishop, Johanna Tomorsky, Joseph Wekselblatt, Judit Pungor and Denise Piscopo - who have been exceptionally fun and supportive to be around. From mushroom hunting to margaritas, my involvement in the Niell lab both academically and socially, will always represent cherished times I shall carry with me into the future.

As part of the neurotheory group involving joint lab meetings, journal clubs, and social outings, I would like to acknowledge Lucca Mazzucato and the members of his lab – Lia Papodopoulos, Daniel Burnham, Nicu Istrate, and David Wyrick - for always being willing to give feedback and advice. A special thank you to David Wyrick for always challenging my thinking and exploring interesting philosophical questions with me. To members of the Murray lab – Christian Schmid, Ben Lemberger, and Matthew Trappett – thank you for the great conversations about math, physics and computer science. Additionally, thank you to all the previous members of the Ahmadian lab – Caleb Holt, Gabriel Barello, and Takafumi Arakaki – because when first starting my Ph.D., their guidance helped me identify and navigate my research interests with more clarity. I am immensely grateful to have been a part of the supportive joint neurotheory group throughout my graduate career.

Finally, I would like to express heartfelt appreciation and gratitude to my friends supporting and encouraging me as I sometimes stumbled through graduate school. For the amazing support, friendship, inspiration and for being a genuinely amazing human being, I would like to extend my gratitude to Rachel Lukowicz for, over these years, expanding my view of neuroscience. Thank you for always being willing to walk me through the complexities of genetics, biology, and developmental neuroscience, and to grow and extend my understanding of what it means to advocate for equality. Rachel will be a lifelong friend and I look forward to enjoying and supporting the unfolding of her future, as she has done with me in the past.

To my family, a special thank you for always believing in and supporting me. I would not have been able to engage this journey as meaningfully as I have, without their steady and positive presence.

This work is dedicated to my first academic mentors, William Gore and Meryl Tsukiji, for teaching me how to learn, grow and think critically. They inspired and pushed me to fulfill my potential, first as student and a person, and then as a researcher. It is not an understatement to say I would not be the person I am today, if they had not taken me under their wing and mentored me.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

The ability to learn, change, and grow is a fundamental aspect of life. By interacting with the world around us, we can gather new information and experiences helping us to adapt and improve. This process is essential for the development of our individual and collective knowledge, as well as for the advancement of science and technology. To transform information into knowledge, whether socially or scientifically construed, we rely on observations, measurements, and theories to interpret our surrounding environment. By weaving common patterns together, we develop frameworks of understanding to help utilize the information we gather.

From a philosophical point of view, conceptual frameworks provide a structure for organizing and understanding the vast amount of information and knowledge we acquire through our interactions with the world. These frameworks help us make sense of the world and connect seemingly disparate pieces of information into a coherent whole. However, as we continue to learn and grow, our understanding of the world can evolve and change. Eventually, this process leads to the limits of a conceptual framework whereby an evolving complexity of new observations and ideas renders the current framework outdated. In such instances, the previous framework breaks down so the latest information can be incorporated into a new framework, better able to explain and understand the world within an improved context of current knowledge. This process of building, breaking down, expanding, and rebuilding is not limited to the development of scientific knowledge, but also has a neurobiological implementation.

From a neuroscience perspective, a conceptual framework can be quantified as the development of an internal model of the environment. By interacting with and observing physical phenomena to understand the dynamics of the world, we can build internal representations of the external world which can be combined, and integrated, for decision-making, motor action, and behavior. For example, visual navigation involves transducing light that hits the retina, a two-dimensional surface, and transforming this information into a three-dimensional representation like a map to navigate from location to location. By interacting with the environment, this map can increase in depth and detail or can change due to unexpected observations.

How do observations about physical phenomena become abstract conceptual representations that build an internal model of the world? One starting point for the generation of an internal model begins by building abstractions of sensory information. Vision is a fundamental sense animals can use to interact with the world at a distance. By processing the visual environment, animals can make abstractions about the space they live in to make decisions. Visual neuroscience has had remarkable success in explaining phenomena by developing models where simple features of sensory processing build into more complex representations (Hubel and Wiesel, 1962, 1959; Manassi et al., 2013). The canonical model of visual processing postulates that higher-order visual representations are built from stimulus responses of simple features that are successively combined in a hierarchical manner (Niell, 2015; Niell and Scanziani, 2021) and are modulated by other activities like locomotion (Niell and Stryker, 2010) or prediction errors (Rao and Ballard, 1999). However, this model of visual processing is still limited since most experimental results have relied upon unnatural visual stimuli and restraining the movement of the animal by either anesthesia or head fixation. Additionally, this mode of visual processing typically assumes static visual input. During natural behavior, sensory information and motor action are continually interdependent in a dynamic manner. For example, visual information is used to inform where and how to move, and self-motion is used to sample new visual information.

Another area of visual processing that started to develop in the late 20th century is active vision. Active vision refers to the ability of an organism or artificial system to actively control its visual sensors to gather information (Blake and Yuille, 1992) . In the field of visual neuroscience, researchers often use computational models to represent and simulate active vision to better understand visual processing. These models typically involve representing the visual sensors and their movements, as well as the visual information that is gathered and processed by the system (Blake and Yuille, 1992). Unlike previous models of visual processing, this framework presents a unique opportunity to incorporate behavior within the capacity of the model. However, until recently due to technological barriers, this area of research within visual neuroscience has not been fully investigated. In recent years, the study of visual processing during natural behavior has received new momentum with the advent of new technologies like deep learning and neural networks. With new methods and experimental observations, the field of visual neuroscience is undergoing a dramatic expansion and conceptual paradigm shift from passive to active processing (Parker et al., 2020). The expansion of our experimental knowledge requires a more complex and comprehensive theoretical model of visual processing, accounting for not only the visual processing but also the motor signals that generate interaction with the environment.

Profound scientific progress occurs at the intersection and integration of experimental and theoretical research. With the recent revolution of new technologies, large datasets of complex data are being generated. As a result, there is a growing need for close collaboration in experimental and theoretical research. Related to the development and evolution of conceptual frameworks, experimental and theoretical research are two approaches to scientific inquiry that are often interdependent and complementary. Experimental research involves collecting data and evidence through controlled observations and experiments to test hypotheses and theories. Theoretical research, on the other hand, involves developing and analyzing theoretical models and frameworks to understand, explain and predict phenomena. These different approaches work in combination to enrich our knowledge about a subject. As a computational and theoretical neuroscientist, this dissertation integrates and shows examples of collaborating closely with experimentalists to develop and expand conceptual frameworks to generate knowledge about how the brain creates representations of the environment during naturalistic visual behaviors.

This dissertation consists of three projects where I developed novel data analysis models for eye tracking in freely behaving mice, a model of visual processing in freely moving mice, and a normative model of visual processing for distance estimation. In Chapter II, I will describe the collaborative effort to investigate how mice shift their gaze during prey capture of crickets. This research laid the foundation to interrogate what neurons in the primary visual cortex (V1) are responding to during free movement. In Chapter III, I will describe a new experimental paradigm where the visual scene and neural activity in V1 can be simultaneously recorded. With this new methodology, I present a novel encoding model to show neurons in V1 respond not only to visual information but also to positional information about the eyes and head during free movement. I will further show that in most neurons, these two signals are integrated through a nonlinear gain modulation mechanism. Finally, in Chapter IV, I will present a new theoretical framework for learning higher-order visual representations. Taken together, this research lays the foundation for an emerging paradigm shift to more closely integrate experimental and theoretical frameworks to expand our understanding of visual processing during naturalistic behaviors. This chapter concludes with the abstracts for Chapters II-IV.

1.1 Chapter II: Dynamics of gaze control during prey capture in freely moving mice

Many studies of visual processing are conducted in constrained conditions such as head- and gaze-fixation, and therefore less is known about how animals actively acquire visual information in natural contexts. To determine how mice target their gaze during natural behavior, we measured head and bilateral eye movements in mice performing prey capture, an ethological behavior that engages vision. We found that the majority of eye movements are compensatory for head movements, thereby serving to stabilize the visual scene. During movement, however, periods of stabilization are interspersed with non-compensatory saccades that abruptly shift gaze position. Notably, these saccades do not preferentially target the prey location. Rather, orienting movements are driven by the head, with the eyes following in coordination to sequentially stabilize and recenter the gaze. These findings relate eye movements in the mouse to other species, and provide a foundation for studying active vision during ethological behaviors in the mouse.

1.2 Chapter III: Joint Coding of Visual Input and Eye/Head Position in V1 of Freely Moving Mice

Visual input during natural behavior is highly dependent on movements of the eyes and head, but how information about eye and head position is integrated with visual processing during free movement is unknown, since visual physiology is generally performed under head-fixation. To address this, we performed single-unit electrophysiology in V1 of freely moving mice while simultaneously measuring the mouse's eye position, head orientation, and the visual scene from the mouse's perspective. From these measures, we mapped spatiotemporal receptive fields during free movement based on the gaze-corrected visual input. Furthermore, we found a significant fraction of neurons tuned for eye and head position, and these signals were integrated with visual responses through a multiplicative mechanism in the majority of modulated neurons. These results provide new insight into coding in mouse V1, and more generally provide a paradigm for performing visual physiology under natural conditions, including active sensing and ethological behavior.

1.3 Chapter IV: Emergence of Distance Estimation in Predictive Neural Networks

Vision allows the brain to form internal models of distant surroundings. These representations are traditionally thought to arise from the bottom-up processing of retinal input. However, when naturally behaving animals move through their environment, and self-motion information is sent via efference copies into visual cortex, this is combined with visual input to calculate objective information about the surroundings. To probe such computations, and inspired by ongoing experiments in mouse vision, we focused on monocular depth estimation through motion parallax, where animals combine self-motion and visual signals to calculate absolute distances to environmental objects. We simulated a camera-agent performing random walks in a 3D environment and fed the recorded 2D camera frames into a multi-layer recurrent neural network (RNN), which was trained to predict future frames. Accurate prediction of future sensory inputs requires learning an internal model for the dynamics of the agent and environment. These dynamics are best modeled in terms of natural dynamical variables (such as 3D coordinates of objects), which may not be explicit in the sensory inputs (such as 2D retinotopic inputs), but can be extracted from them. Therefore, we hypothesized that unsupervised predictive learning results in an explicit representation of depth by providing self-motion signals to the RNN, would further improve this representation. To test our hypothesis, we trained linear readouts from neural activations in different layers of the trained RNN to match ground-truth depth maps. We found that the RNN does indeed form depth representations, without being explicitly tasked to do so. Moreover, depth representations become more explicit and accurate in deeper layers of the network. These results suggest internal representations of depth can arise from learning to predict future monocular visual inputs and potentially, by integrating visual and self-motion signals.

CHAPTER II

DYNAMICS OF GAZE CONTROL DURING PREY CAPTURE IN FREELY MOVING MICE

2.1 JOURNAL STYLE INFORMATION

Originally published as Michaiel, A. M., Abe, **E. T. T.**, Niell, C. M. (2020). Reproduced from eLife, 9, 1–18. https://doi.org/10.7554/elife.57458

2.2 AUTHOR CONTRIBUTIONS

Angie M Michaiel, Conceived the project, Developed methodology and designed experiments, Analyzed data, Wrote the manuscript, Created figures; Elliott TT Abe, Wrote camera calibration software, Contributed to data analysis, Manuscript preparation; Cristopher M Niell, Conceived the project, Designed experiments, Analyzed data, Wrote manuscript, Provided resources

2.3 INTRODUCTION

Across animal species, eye movements are used to acquire information about the world and vary based on the particular goal (Yarbus, 1967). Mice, a common model system to study visual processing due to their genetic accessibility, depend on visual cues to successfully achieve goal-directed tasks in both artificial and ethological freely-moving behavioral paradigms, such as the Morris water maze, nest building, and prey capture; (MORRIS, 1981; Clark et al., 2006; Hoy et al., 2016). It is unclear, however, how mice regulate their gaze to accomplish these goals. Previous studies in both freely moving rats and mice have shown that eye movements largely serve to compensate for head movements (Wallace et al., 2013; Payne and Raymond, 2017; Meyer et al., 2018, 2020), consistent with the vestibulo-ocular reflex (VOR) present in nearly all species (Straka et al., 2016). While such compensation can serve to stabilize the visual scene during movement, it is not clear how this stabilization is integrated with the potential need to shift the gaze for behavioral goals, particularly because mice lack a specialized central fovea in the retina, and also have laterally facing eyes resulting in a relatively limited binocular field (roughly 40° as opposed to 135° in humans (Drager, 1978)). In addition, because eye movements are altered in head-fixed configurations due to the lack of head movement (Payne and Raymond, 2017; Meyer et al., 2020), understanding the mechanisms of gaze control and active visual search benefits from studies in freely moving behaviors.

Prey capture can serve as a useful paradigm for investigating visually guided behavior. Recent studies have shown that mice use vision to accurately orient towards and pursue cricket prey (Hoy et al., 2016), and have begun to uncover neural circuit mechanisms that mediate both the associated sensory processing and motor output (Hoy et al., 2019; Shang et al., 2019; Zhao et al., 2019; Han et al., 2017). Importantly, prey capture also provides a context to investigate how mice actively acquire visual information, as it entails identifying and tracking a localized and ethological sensory input during freely moving behavior. Here, we asked whether mice utilize specific eye movement strategies, such as regulating their gaze to maximize binocular overlap, or actively targeting and tracking prey. Alternatively, or in addition, mice may use directed head movements to target prey, with eye movements primarily serving a compensatory role to stabilize the visual scene.

Predators typically have front-facing eyes which create a wide binocular field through the overlap of the two monocular fields, allowing for depth perception and accurate estimation of prey location (Carandini et al., 2005). Prey species, in contrast, typically have laterally facing eyes, and as a result, have large monocular fields spanning the periphery, which allow for reliable detection of approaching predators. Though mice possess these characteristics of prey animals, they also act as predators in pursuing cricket prey (Hoy et al., 2016). How then do animals with laterally placed eyes target prey directly in front of them, especially when these targets can rapidly move in and out of the narrow binocular field? This could require the modulation of the amount of binocular overlap, through directed lateral eye movements, to generate a wider binocular field, such as in the case of cuttlefish (Feord et al., 2020), fish (Bianco et al., 2011), many birds (Martin, 2009), and chameleons (Katz et al., 2015). In fact, these animals specifically rotate their eyes nasally before striking prey, thereby creating a larger binocular zone. However, it is unknown whether mice use a similar strategy during prey capture. Alternatively, they may use coordinated head and eye movements to stabilize a fixed size binocular field over the visual target.

Foveate species make eye movements that center objects of interest over the retinal fovea, in order to use high acuity vision for complex visual search functions including identifying and analyzing behaviorally relevant stimuli (Hayhoe and Ballard, 2005). Importantly, afoveate animals (those lacking a fovea) represent a majority of vertebrate species, with only some species of birds, reptiles, and fish possessing distinct foveae (Harkness and Bennet-Clark, 1978), and among mammals, only simian primates possessing foveae (Walls, 1942). It remains unclear whether mice, an afoveate mammalian species, actively control their gaze to target and track moving visual targets using directed eye movements, or whether object localization is driven by head movements. We therefore aimed to determine the oculomotor strategies that allow for effective targeting of a discrete object, cricket prey, within the context of a natural behavior.

Recent studies have demonstrated the use of head-mounted cameras to measure eye movements in freely moving rodents (Wallace et al., 2013; Meyer et al., 2018; **?**). Here, we used miniature cameras and an inertial measurement unit (IMU) to record head and bilateral eye movements while unrestrained mice performed a visually guided and goal-directed task, approach and capture of live insect prey. We compared the coordination of eye and head movements, as well as measurements of gaze relative to the cricket prey during approach and non-approach epochs, to determine the oculomotor strategies that mice use when localizing moving prey.

2.4 RESULTS

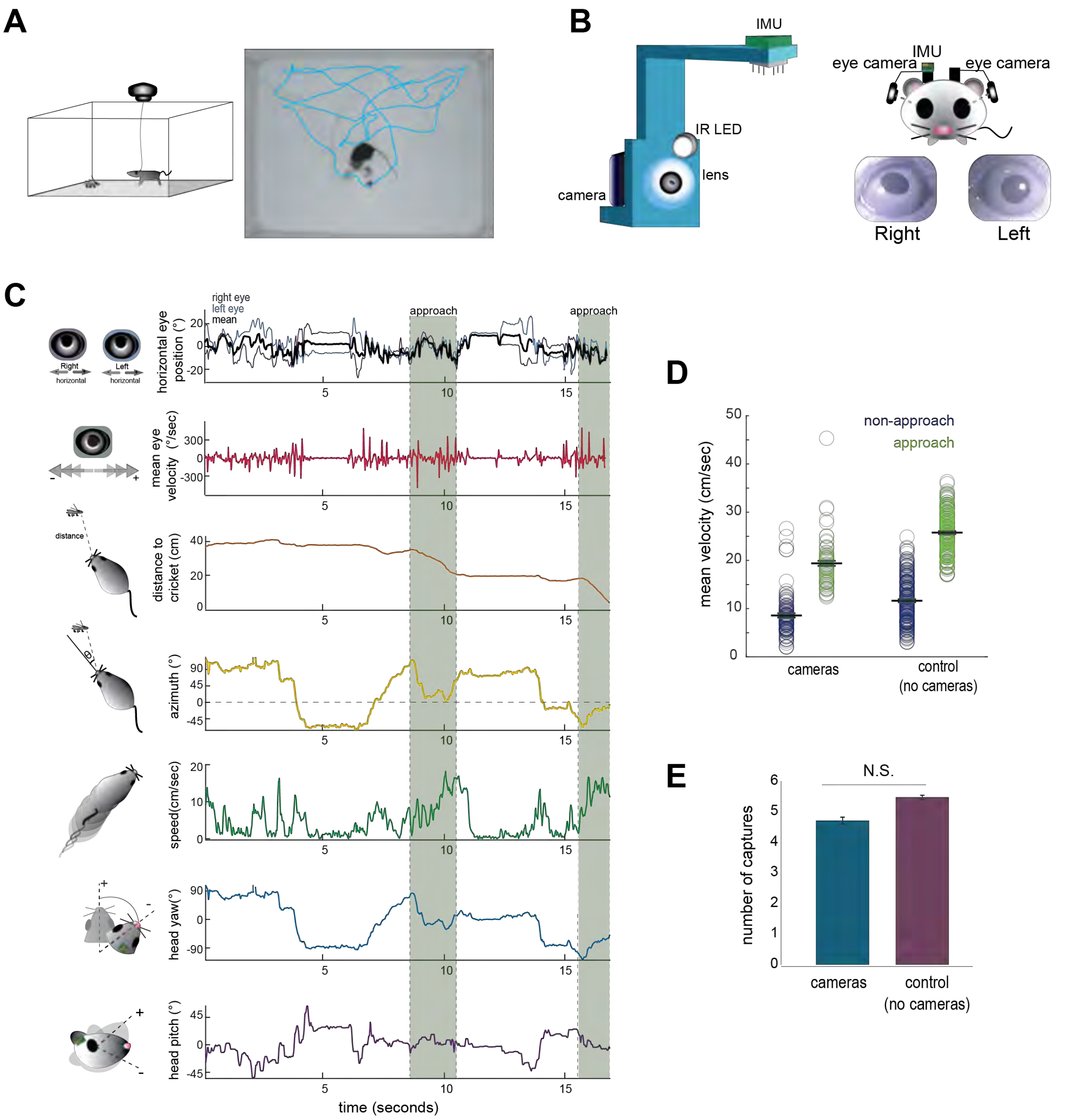
2.4.1 Tracking eye and head movements during prey capture

Food-restricted mice were habituated to hunt crickets in an experimental arena, following the paradigm of (Hoy et al., 2016). To measure eye and head rotations in all dimensions, mice were equipped with two reversibly attached, lateral head-mounted cameras and an inertial measurement unit (IMU) board with an integrated 3-dimensional accelerometer and gyroscope (Figures 2.1A, 2.1B; Supplemental Movie 1). In addition, we recorded the behavior of experimental animals and the cricket prey with an overhead camera to compute the relative position of the mouse and cricket, as well as orientation of the head relative to the cricket. Following our previous studies (Hoy et al., 2016, 2019), we defined approaches based on the kinematic criteria that the mouse was oriented towards the cricket and actively moving towards it (see Methods). Together, these recordings and analyses allowed us to synchronously measure eye and head rotations along with cricket and mouse kinematics throughout prey capture behavior (Figure 2.1C; Supplemental Movie 1). The cameras and IMU did not affect overall mouse locomotor speed in the arena or total number of crickets caught per 10-minute session (paired t-test, p=0.075; Figure 2.1D/E), suggesting that placement of the cameras and IMU did not significantly impede movement or occlude segments of the visual field required for successful prey capture behavior.

2.4.2 Eye vergence is stabilized during approach periods

To determine whether mice make convergent eye movements to enhance binocular overlap during approaches, we first characterized the coordination of bilateral eye movements. We defined central eye position, i.e. 0°, as the average pupil location for each eye, across the recording duration. Measurement of eye position revealed that freely moving mice nearly constantly move their eyes, typically within a ± 20 degree range (Figure 2.1C, 2.2A), as shown previously. Figure 2.2C shows example traces of the horizontal position of the two eyes (top), along with running speed of the mouse (bottom). As described previously (Payne and Raymond, 2017; Wallace et al., 2013; Meyer et al., 2018) and analyzed below (Figure 2.3D), the eyes are generally stable when the mouse is not moving. In addition, the raw traces reveal a pattern of eye movement wherein rapid correlated movements of the two eyes are superimposed on slower anti-correlated movements. The pattern of rapid congruent movements and slower incongruent movements was also reflected in the time-lagged cross-correlation of the change in horizontal position across the two eyes (Figure 2.2E), which was positive at short time lags and negative at longer time lags.

Figure 2.1. (Next page) Tracking eye and head movements during prey capture. **A)** Unrestrained mice hunted live crickets in a rectangular plexiglass arena (45x38x30 cm). Using an overhead camera, we tracked the movement of the mouse and cricket. Example image with overlaid tracks of the mouse (cyan). **B)** 3D printed holders house a miniature camera, collimating lens, an IR LED, and an IMU, and are reversibly attached to implants on the mouse's head, with one camera aimed at each eye. **C)** Synchronized recordings of measurements related to bilateral eye position, mouse position relative to cricket (distance and azimuth), mouse speed, and head rotation in multiple dimensions (analysis here focuses on yaw and pitch). **D)** Average mouse locomotor speed did not differ across experimental and control experiments (no camera and IMU) for both non-approach and approach periods. Individual dots represent the average velocity per trial. **E)** Average number of captures per 10 minute session did not differ between experimental and control sessions (control N=7 animals, 210 trials; cameras N=7 animals, 105 trials; two-sample t-test, p=0.075).

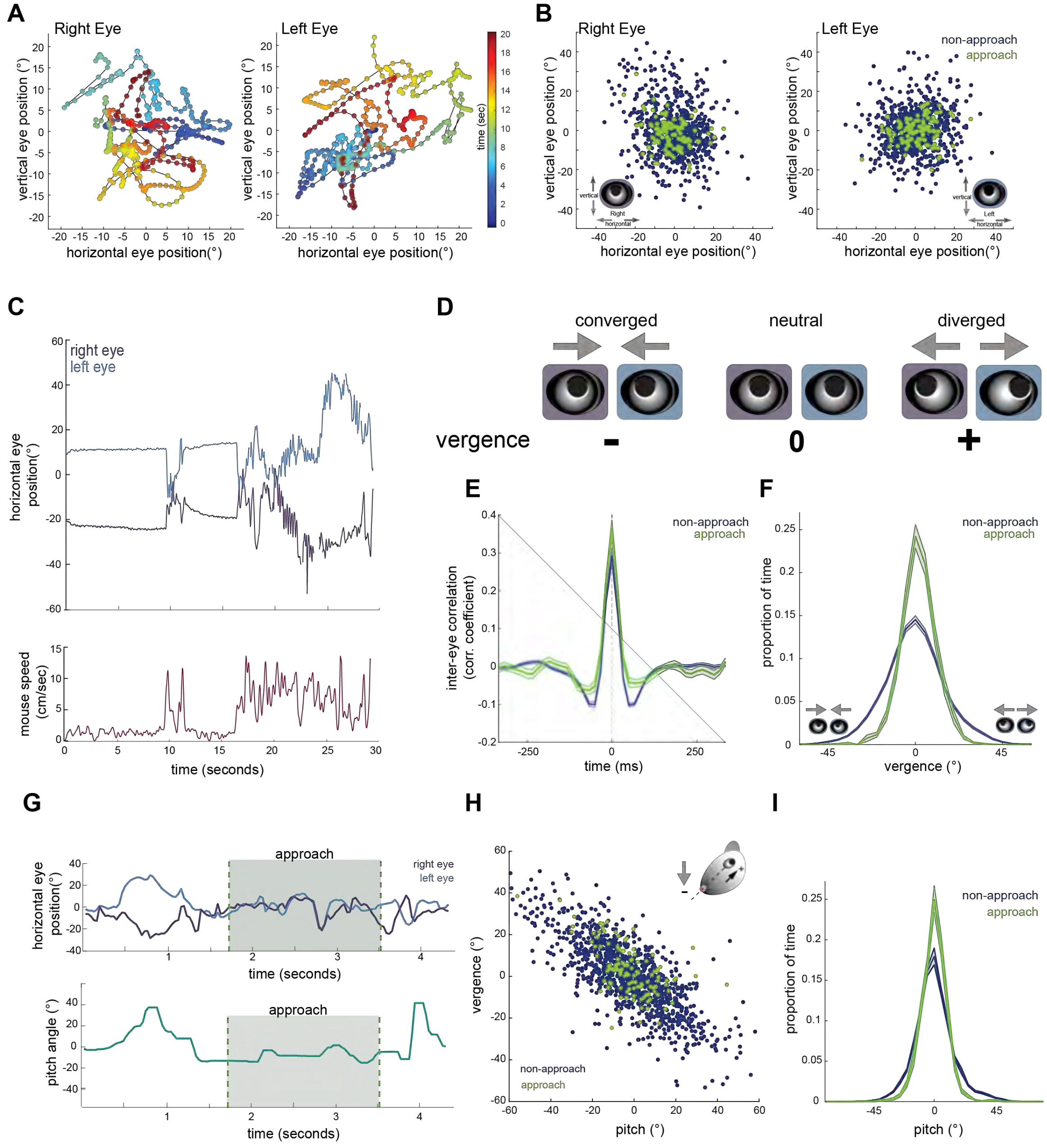


We next calculated the vergence angle, which is the difference in the horizontal position of the two eyes (Figure 2.2D). The range of vergence angles was broadly distributed across negative (converged) and positive (diverged) values during non-approach periods, but became more closely distributed around zero (neutral vergence) during approaches (Figure 2.2F; paired t-test, p=1.9x10-13). This can be observed in the individual trace of eye movements before, during, and after an approach (Figure

2.2G, top), showing that while the eyes converge and diverge outside of approach periods, they move in a more coordinated fashion during the approaches. Thus, mice do not converge their eyes nasally to create a wider binocular field during approaches; rather the eyes are more tightly aligned, but at a neutral vergence, during approaches relative to non-approach periods.

Previous studies have demonstrated that eye vergence varies with head pitch (Wallace et al., 2013; Meyer et al., 2018, 2020). As the head tilts downwards, the eyes move outwards; based on the lateral position of the eyes, this serves to vertically stabilize the visual scene relative to changes in head pitch (Wallace et al., 2013). We therefore sought to determine whether the stabilization of horizontal eye vergence we observed during approaches reflects corresponding changes in head pitch. Consistent with previous studies, we also found eye vergence to covary with head pitch (Figure 2.2H), such that when the head was vertically centered, the eyes no longer converged or diverged, but were aligned at a neutral vergence (i.e., no difference between the angular positions across the two eyes, see schematic in Figure 2.2D).

Figure 2.2. **(Next page) Eye position is more aligned across the two eyes during approach periods. A)** Example eye movement trajectory for right and left eyes for a 20 second segment, with points color-coded for time. **B)** Horizontal and vertical position for right and left eyes during approach and non-approach times. N=7 animals, 105 trials, 805 time pts (non-approach), 105 time pts (approach), representing a random sample of 0.4% of non-approach and 1% of approach time points. **C)** Example trace of horizontal eye positions (top) and running speed (bottom) for a 30 second segment. **D)** Schematic demonstrating vergence eye movements. **E)** Cross correlation of horizontal eye position across the two eyes for non-approach and approach periods. **F)** Histogram of vergence during non-approach and approach. **G)** Example trace of horizontal eye position (top) and head pitch (bottom) before, during, and after an approach. **H)** Scatter plot of head pitch and eye vergence. As head pitch tilts downwards, the eyes move temporally to compensate (as in schematic). N=7 animals, 105 trials, 1252 time points (non-approach), 123 time points (approach), representing a sample of 0.7% of non-approach and 1.2% of approach time points. **I)** Histogram of head pitch during approach and non-approach periods, across all experiments.



Strikingly, we found that while the relationship between head pitch and vergence was maintained during approaches (Figure 2.2H), the distribution of head pitch was more centered during approach periods (Figure 2.2H, 2.2I; paired t-test, p=1.5x10-13), indicating a stabilization of the head in the vertical dimension. This can also be seen in the example trace in Figure 2.2G, where the head pitch becomes maintained around zero during pursuit. These data show that the increased alignment of the two eyes observed during approaches largely represents the stabilization of up/down head rotation during active pursuit, consequently reducing the need for compensatory vergence movements.

2.4.3 Coordinated horizontal eye movements are primarily compensatory for horizontal head rotations

Next, we aimed to understand the relationship between horizontal head movements (yaw) and horizontal eye movements during approach behavior. In order to isolate the coordinated movement of the two eyes, removing the compensatory changes in vergence described above, we averaged the horizontal position of the two eyes for the remaining analyses (Figure 2.3A). Changes in head yaw and mean horizontal eye position were strongly negatively correlated at zero time lag (Figure 2.3B), suggesting rapid compensation of head movements by eye movements, as expected for VOR-stabilization of the visual scene. The distribution of head and eye movements at zero lag (Figure 2.3C) shows that indeed changes in head yaw were generally accompanied by opposing changes in horizontal eye position, represented by the points along the negative diagonal axis. However, there was also a distinct distribution of off-axis points, representing a proportion of non-compensatory eye movements in which the eyes and head moved in the same direction (Figure 2.3C).

Many studies have reported a limited range (Niell and Stryker, 2010; Payne and Raymond, 2017; Samonds et al., 2018; Stahl, 2004), consistent with the idea that eye movements are generally driven by head movement. Correspondingly in the freely moving context of the prey capture paradigm, we found greatly reduced eye movements when the animals were stationary versus when the animals were running (Figure 2.3D; Kolmogorov-Smirnov test, p=0.032).

We next compared the distribution of mean eye position during approaches and non-approach periods. In contrast to the stabilization of head pitch described above, the distribution of head yaw velocities was not reduced during approaches as shown (Figure 2.3E; paired t-test p=0.889), consistent with the fact that mice must move their heads horizontally as they continuously orient to pursue prey. For both non-approach and approach periods, eye position generally remained within a range less than the size of the binocular zone (± 20 degrees; Figure 2.3F, paired t-test, p=0.044), suggesting that the magnitude of eye movements would not shift the binocular zone to an entirely new location. Comparison of horizontal eye velocity between non-approach and approach epochs revealed that the eyes move with similar dynamics across both behavioral periods (Figure 2.3G, panel 1; paired t-test, p=.072). Additionally, at times when head yaw was not changing, horizontal eye position also did not change (Figure 2.3G, panel 2; paired t-test, p=0.13). Together, these observations suggest that most coordinated eye movements in the horizontal axis correspond to changes in head yaw, and that the eyes do not scan the visual environment independent of head movements or when stationary.

2.4.4 Non-compensatory saccades shift gaze position

Gaze position - the location the eyes are looking in the world - is a combination of the position of the eyes and the orientation of the head. Compensatory eye movements serve to prevent a shift in gaze, whereas non-compensatory eye movements (i.e., saccades) shift gaze to a new position. Although the vast majority of eye movements are compensatory for head movements, as demonstrated by strong negative correlation in Figure 2.3B/C, a significant number of movements are not compensatory, as seen by the distribution of off-axis points in Figure 2.3C. These eye movements will therefore shift the direction of the animal's gaze relative to the environment. We next examined how eye movements, and particularly non-compensatory movements, contribute to the direction of gaze during free exploration and prey capture. In particular, are these gaze shifts directed at the target prey?

We segregated eye movements into compensatory versus gaze-shifting by setting a fixed gaze velocity threshold of ±180 /sec, based on the gaze velocity distribution (Figure 2.4A), which shows a transition between a large distribution around zero

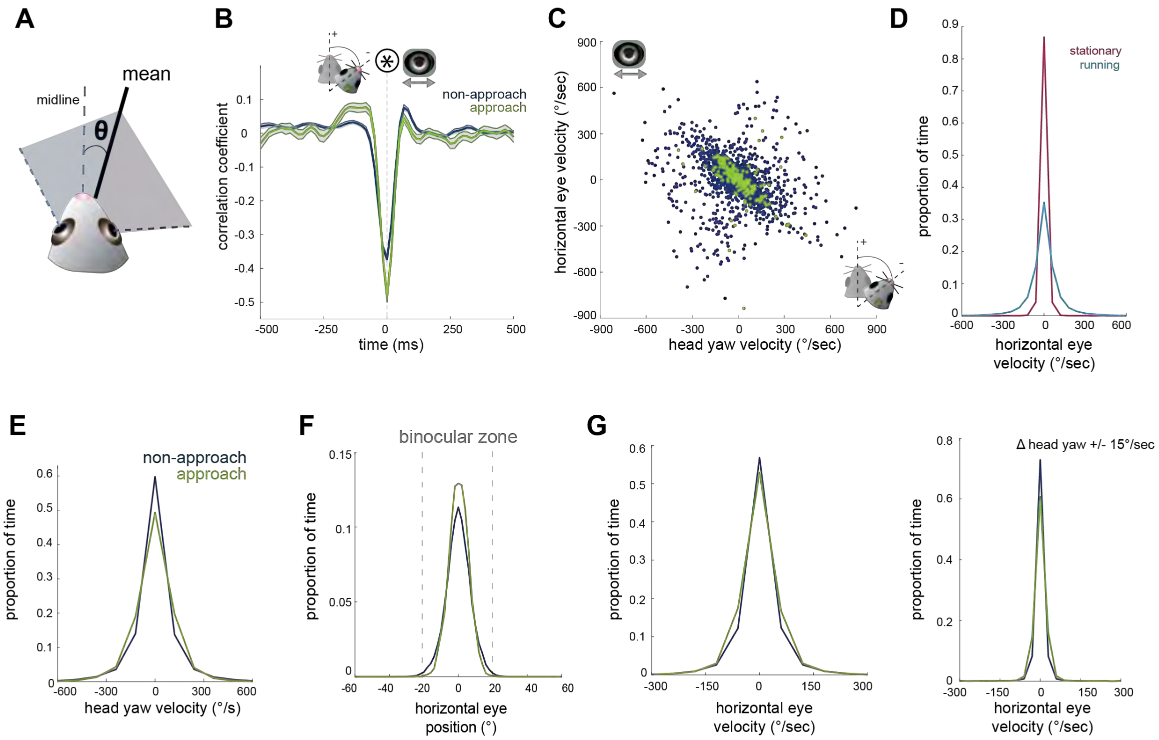


Figure 2.3. **Horizontal eye movements are mostly compensatory for yaw head rotations.** **A)** To remove the effect of non-conjugate changes in eye position (i.e. vergence shifts), we compute the average angular position of the two eyes. **B)** Cross-correlation of change in head yaw and horizontal eye position. **C)** Scatter plot of horizontal rotational head velocity and horizontal eye velocity. N=7 animals, 105 trials, 3604 (non-approach) and 201 (approach) timepoints, representing 2% of non-approach and 2% of approach timepoints. **D)** Distribution of horizontal eye position during stationary and running periods (defined as times when mouse speed is greater than 1 cm/sec; Kolmogorov-Smirnov test, p=0.032). **E)** Distribution of head angle velocity (paired t-test, p=0.889). **F** Distribution of mean absolute eye position (paired t-test, p=0.044). **G)** Distribution of horizontal eye velocity (paired t-test, p=0.072) and distribution of eye velocity when head yaw is not changing (change in head yaw between ± 15 deg/sec; paired t-test, p=0.13; N=7 animals, 105 trials).

(stabilized gaze) and a long tail of higher velocities (rapid gaze shifts). This also provides a clear segregation in the joint distribution of eye and head velocity (Figure 2.4B), with a large number of compensatory gaze-stabilizing movements (black points) where eye and head motion are anti-correlated, and much smaller population of gaze shifts (red). This classification approach provides an alternative to standard primate saccade detection (Andersen and Mountcastle, 1983; Stahl, 2004; Mathis et al., 2018), which is often based on eye velocity rather than gaze velocity, since in the freely moving condition, particularly in afoveate species, rapid gaze shifts (saccades) often result from a combination of head and non-compensatory eye movements, rather than eye movements alone (Land, 2006).

We next determined how compensatory and non-compensatory eye movements contribute to the dynamics of gaze during ongoing behavior, by computing the direction of gaze as the sum of eye position and head position. Strikingly, the combination of compensatory and non-compensatory eye movements (Figure 2.4C, top) with continuous change in head orientation (Figure 2.4C, middle) results in a series of stable gaze positions interspersed with abrupt shifts (Figure 2.4C, bottom). This pattern of gaze stabilization interspersed with rapid gaze-shifting movements, known as "saccade-and-fixate," is present across the animal kingdom and likely represents a fundamental mechanism to facilitate visual processing during movement (Land, 1999). These results demonstrate that the mouse oculomotor system also engages this fundamental mechanism.

Durations of fixations between saccades showed wide variation, with a median of 230 ms (Figure 2.4D). To quantify the degree of stabilization achieved, we compared the root mean square (RMS) deviation of gaze position and head yaw during stabilization periods (Figure 2.4E). This revealed that the gaze is approximately three times less variable than the head (Figure 2.4F; median head=3.91 deg; median gaze=1.58 deg; paired t-test; p=0), resulting in stabilization to within nearly 1 degree over extended periods, even during active pursuit.

2.4.5 Targeting of gaze relative to cricket during pursuit

Saccade-and-fixate serves as an oculomotor strategy to sample and stabilize the visual world during free movement. In primates, saccades are directed towards specific targets of interest in the visual field. Is this true of the non-compensatory movements

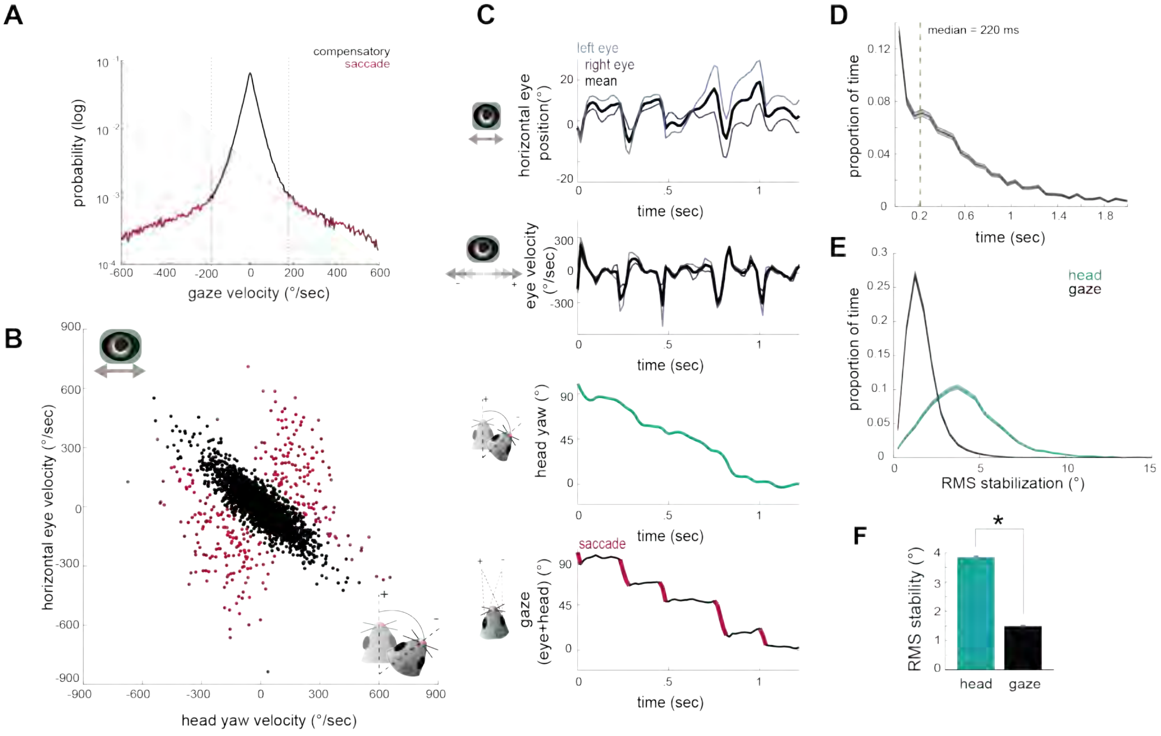


Figure 2.4. **Head movements and subsequent saccades target the cricket during prey capture.** **A)** Joint distributions of head yaw and horizontal eye velocity were clustered into compensatory eye movements (black) and non-compensatory eye movements (red). Clustering was done on all approach timepoints (N=10026). Points shown are a random sample of 2005 points, 20% of total approach time points. **B)** Histograms of gaze velocity for compensatory and saccade eye movement distributions, based on clustering shown in A. **C)** Example traces of horizontal eye position, head yaw, and gaze demonstrate a saccade-and-fixate pattern in gaze. **D)** Histogram of fixation duration; fixations N=8761, 105 trials. **E)** Root Mean Squared (RMS) stabilization histograms for head yaw and gaze. **F)** Bar graphs are medians of RMS stabilization distributions (median head=3.91 deg; median gaze=1.58 deg; paired t-test, p=0).

in the mouse? In other words, do saccades directly target the cricket? To address this, we next analyzed the dynamics of head and gaze movements relative to the cricket position during hunting periods, to compare how accurately the direction of the gaze and the head targeted the cricket during saccades. Figure 2.5A shows example traces of head and eye dynamics across a pursuit period (see also Supplemental Movie 2). Immediately before pursuit, the animal begins a large head turn towards the target, thereby reducing the azimuth angle (head relative to cricket). This head turn is accompanied by a non-compensatory eye movement in the same direction (Figure 2.5A, 3rd panel, see mean trace in black) that accelerates the shift in gaze. Then during the pursuit, the eyes convert the continuous tracking of the head into a series of stable locations of the gaze (black sections in Figure 2.5A, bottom). Note also the locking of the relative position of the two eyes (Figure 2.5A, 3rd panel, blue and purple), as described above in Figure 2.2.

To determine how head and eye movements target the prey, we computed absolute value traces of head and gaze angle relative to cricket (head and gaze azimuth), and aligned these to the onset of each non-compensatory saccadic eye movement. The average of all traces during approaches revealed that saccades are associated with a head turn towards the cricket, as shown by a decrease in the azimuth angle (Figure 2.5B). Immediately preceding a saccade, the gaze is stabilized while the head turns, and the saccade then abruptly shifts the gaze. Notably, following the saccade, the azimuth of gaze is the same as the azimuth of the head, suggesting that eye movements are not targeting the cricket more precisely, but simply 'catching up' with the head, by re-centering following a period of stabilization.

To further quantify this, we assessed the accuracy of the head and gaze at targeting cricket position before and after saccades. Preceding saccades, the distribution of head angles was centered around the cricket, while the gaze less accurately targeted and offset to the left or right (Figure 2.5C/5D top; paired t-test $p=2x10^{-5}$), due to compensatory stabilization. After the saccade, however, gaze and head were equally targeted towards the cricket (Figure 2.5C/D bottom; p=0.4), as the saccade recentered the eyes relative to the head and thereby the cricket. This pattern

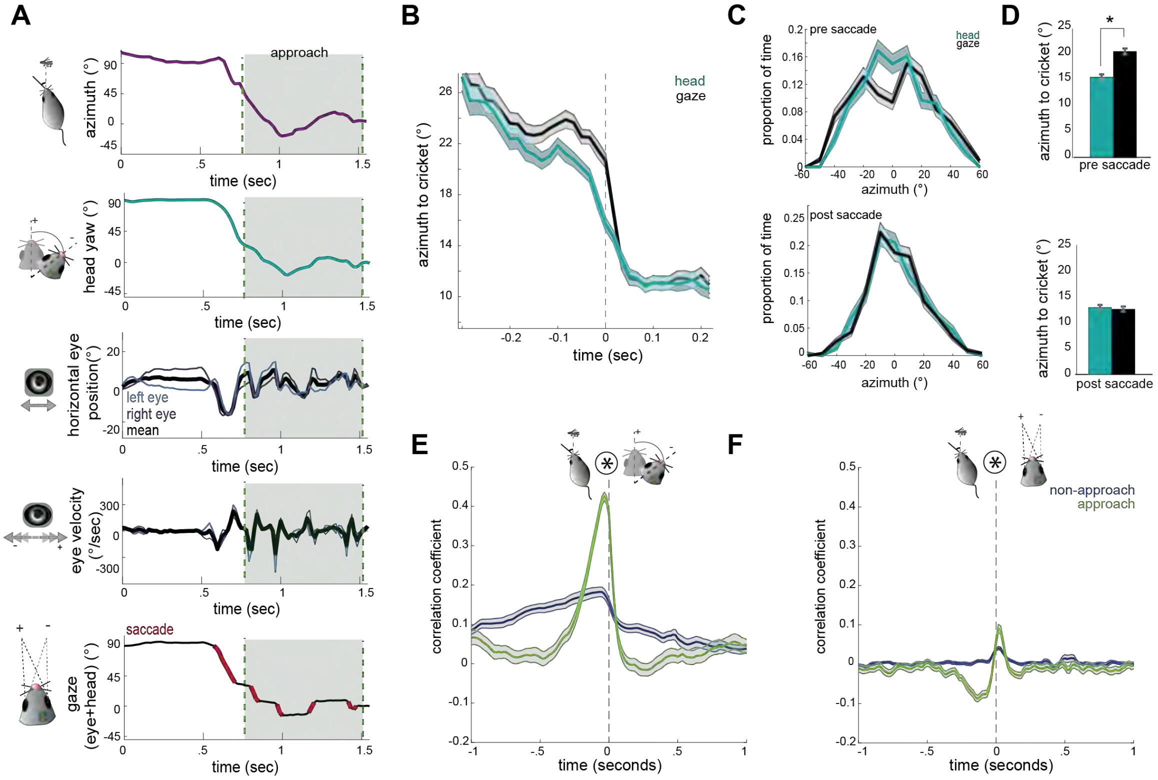


Figure 2.5. **Head angle tracks cricket position more accurately than gaze position.** **A)** Example traces of horizontal eye position, azimuth to cricket, head yaw, and gaze demonstrate a saccade-and-fixate pattern in gaze before and during an approach period. The head is pointed directly at the cricket when azimuth is 0°. Note the rapid decrease in azimuth, head yaw, and mean horizontal eye position creating a saccade immediately preceding the start of approach. **B)** Average head yaw and gaze around the time of saccade as a function of azimuth to the cricket. Time = 0 is the saccade onset. **C)** Histograms of head yaw and gaze position before and after saccades occur. **D)** Medians of yaw and gaze distributions from C (paired t-test, $p_{\text{pre saccade}}=2x10^{-5}$; $p_{\text{post saccade}}=0.4$). **E)** Cross correlation of azimuth and change in head yaw for non-approach and approach periods. **F)** Cross correlation of azimuth and change in gaze for non-approach and approach periods. N=105 trials, 7 animals.

of stabilizing the gaze and then saccading to recenter the gaze repeats whenever the head turns until capture is successful (see Supplemental Movie 2).

Further supporting a strategy where the head guides targeting, with the eyes following to compensate, we examined how both head and eye movements are correlated with the cricket's position. At short latencies, the change in head angle relative to the location of the cricket was highly correlated (Figure 2.5E), indicating that during pursuit the animal is rapidly reorienting its head towards the cricket. However, the change in gaze with the azimuth instead showed only a weak correlation because the eyes themselves are not always aligned with the azimuth due to stabilization periods (Figure 2.5F). Together, these results suggest that in mice, tracking of visual objects in freely moving contexts is mediated through directed head movements, and corresponding eye movements that stabilize the gaze and periodically update to recenter with the head as it turns.

2.5 DISCUSSION

Here we investigated the coordination of eye and head movements in mice during a visually guided ethological behavior, prey capture, that requires the localization of a specific point in the visual field. This work demonstrates that general principles of coordinated eye and head movements, observed across species, are present in the mouse. Additionally, we address the potential targeting of eye movements towards behaviorally relevant visual stimuli, specifically the moving cricket prey. We find that tracking is achieved through directed head movements that accurately target the cricket prey, rather than directed, independent eye movements. Together, these findings define how mice move their eyes to achieve an ethological behavior and provide a foundation for studying active visually-guided behaviors in the mouse.

One potential limitation of our eye tracking system is the 60Hz framerate of the miniature cameras. This temporal resolution is significantly lower than traditional eye tracking paradigms using videography or eye-coil systems in head-restrained humans, non-human primates, and rodents (Payne and Raymond, 2017; Sakatani and Isa, 2007), though similar to recent video-based tracking in freely moving rodents (Meyer et al., 2018, 2020) and humans (Mathis et al., 2018). We do not expect that this would significantly alter our findings, as the basic parameters of eye movements (amplitude and speed) that we found (Figures 2.2B, 2.3C, 2.3F) were similar to measurements made in both head-fixed mice with high-speed videography (Sakatani and Isa, 2007) and freely moving mice with a magnetic sensor (Payne and Raymond, 2017). However, although we are able to detect peak velocities over 300°/sec, we may still be underestimating the peak velocity during saccades. Therefore increasing the temporal resolution further could lead to more robust detection of rapid gaze shifts and would potentially enhance classification of saccadic eye movements.

We found a pattern of gaze stabilization interspersed with abrupt, gaze-shifting saccades during both non-approach and approach epochs. This oculomotor strategy has been termed 'saccade-and-fixate' (reviewed in (Land, 1999)), and is present in most visual animal species, from insects to primates, and was recently demonstrated in mice (Meyer et al., 2020). In primates, gaze shifts can be purely driven by eye movements, but in other species saccades generally correspond to non-compensatory eye movements during head rotation, suggesting transient disengagement of VOR mechanisms. These saccadic movements are present in invertebrates and both foveate and non-foveate vertebrates (reviewed in (Land, 1999)), and work to both recenter the eyes and relocate the position of gaze as animals turn. We found that these brief congruent head and eye movements are interspersed with longer duration (median 200 ms) periods of compensatory movements, which stabilize the gaze to within nearly 1 as the head continues to rotate. Together these eye movements function to create a stable series of images on the retina even during rapid tracking of prey.

However, the saccade-and-fixate strategy raises the question of whether mice actively target a specific relevant location with saccadic eye movements. We examined this during periods of active pursuit to determine whether the eyes specifically target the cricket, relative to head orientation. During pursuit, most saccades occur during corrective head turns toward the cricket location. While saccades do bring the gaze closer to the cricket, they do not do so more accurately than the head direction. In fact, prior to the saccade, mice sacrifice centering of the gaze on the target to instead achieve visual scene stability. The eyes then ‘catch up’ to the head as it is rotating (Figure 2.5B/C). Thus, these eye movements serve to reset eye position with the head, rather than targeting the cricket specifically. Combined with the fact that mice do not make significant eye movements in the absence of head movements (Figure 2.3F), this suggests that mice do not perform either directed eye saccades or smooth pursuit, which are prominent features of primate vision. On the other hand, the fact that they use a saccade-and-fixate strategy makes it clear that they are still actively controlling their retinal input, despite low visual acuity and the common perception that mice are not a highly visual species. Indeed, the saccade-and-fixate strategy makes mouse vision consistent with the vast majority of species across the animal kingdom.

We also examined whether mice make specific vergence eye movements that could serve to modulate the binocular zone, as in some other species with eyes located laterally on the head. We find that rather than moving the eyes nasally to expand the binocular zone, during approach toward the cricket the two eyes become stably aligned, but at a neutral vergence angle that is neither con- verged or diverged (Figure 2E). While several species with laterally-placed eyes use convergent eye movements during prey capture to create a wider binocular field (Feord et al., 2020; Bianco et al., 2011; Martin, 2009; Katz et al., 2015), our results show that mice do not utilize this strategy during prey capture. However, vergence eye movements in rodents have previously been shown to compensate for head tilt (Wallace et al., 2013), and correspondingly we find that during approach periods mice stabilize head tilt. Thus, the stable relative alignment of the two eyes during approach likely reflects stabilization of the head itself. These results suggest that the 40 degree binocular zone is sufficient for tracking centrally located objects, as the eyes to not move to expand this during approaches. This is consistent with previous work showing that during active approach the mouse's head is oriented within ±15 degrees relative to the cricket (Hoy et al., 2016), meaning that even the resting binocular zone would encompass the cricket. However, it remains to be determined whether mice actually use binocular disparity for depth estimation during prey capture. A recent study demonstrated that mouse V1 encodes binocular disparities spanning a range of 3–25 cm from the mouse's head (Land, 2019), suggesting that disparity cues are available at the typical distances during approach (interquartile range 14.6 cm to 27.6 cm). Alternatively, mice may use retinal image size or other distance cues, or may simply orient to the azimuthal position of the cricket regardless of distance.

The finding that mice do not specifically move their eyes to target a location does not preclude the possibility that different regions of retinal space are specialized for certain processing. In fact, as a result of targeting head movements, the cricket prey is generally within the binocular zone during approach, so any mechanisms of enhanced processing in the binocular zone or lateral retina would still be behaviorally relevant. Anatomically, there is a gradient in density of different retinal ganglion cell types from medial to lateral retina (Bleckert et al., 2014). Likewise behavioral studies have shown enhanced contrast detection when visual stimuli are located in the binocular field, rather than the monocular fields (Speed et al., 2019). Based on the results presented here, in mice these specializations are likely to be engaged by head movements that localize stimuli in the binocular zone in front of the head, as opposed to primates, which make directed eye movements to localize stimuli on the fovea.

Together, the present findings suggest that orienting relative to visual cues is driven by head movements rather than eye movements in the mouse. This is consistent with the general finding that for animals with small heads it is more efficient to move the head, whereas animals with large heads have adapted eye movements for rapid shifts to overcome the inertia of the head (Land, 2019). From the experimental perspective, this suggests that head angle alone is an appropriate measure to determine which visual cues are important during study of visually guided, goal-directed behaviors in the mouse. However, measurements of eye movements will be essential for computing the precise visual input animals receive (i.e., the retinal image) during ongoing freely moving behaviors, and how this visual input is processed within visual areas of the brain. The saccade-and-fixate strategy generates a series of stable visual images separated by abrupt changes in gaze that shift the visual scene and location of objects on the retina. How then are these images, interleaved with periods of motion blur, converted into a continuous coherent percept that allows successful natural behaviors to occur? Anticipatory shifts in receptive field location during saccades, as well as gaze position-tuned neural populations, have been proposed as mechanisms in primates to maintain coherent percepts during saccades, while corollary discharge, saccadic suppression, and visual masking have been proposed to inhibit perception of motion blur during rapid eye movements (Higgins and Rayner, 2015; Wurtz, 2008). However, the mechanisms that might mediate these, at the level of specific cell types and neural circuits, are poorly understood. Studying these processes in the mouse will allow for investigation of the neural circuit basis of these perceptual mechanisms through the application of genetic tools and other circuit dissection approaches (Huberman and Niell, 2011; Luo et al., 2008). Importantly, most of our visual perception occurs during active exploration of the environment, where the combined dynamics of head and eye movements create a dramatically different image processing challenge than typical studies in stationary subjects viewing stimuli on a computer monitor. Examination of these neural mechanisms will extend our understanding of how the brain performs sensory processing in real-world conditions.

2.6 MATERIALS AND METHODS

2.6.1 Key Resource Table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Strain, strain background (Mus musculus)	C57Bl/6J	JAX	JAX: 000664	Wild type animals
Software, algorithm	Matlab	Matlab	Matlab R2020a	
Software, algorithm	DeepLabCut	Mathis et al., 2018		
Software, algorithm	Bonsai	Lopes et al., 2015		

Table 1. Chapter II: Key Resource Table

2.6.2 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 (protocol number 17–27). Animals used for this study were wild-type (C57 Bl/6J) males and females (3 males and four females) aged 2–6 months.

2.6.3 Prey capture behavior

Prey capture experiments were performed following the general paradigm of (Hoy et al., 2016). Mice readily catch crickets in the homecage without any training or habituation, even on the first exposure to crickets. However, we perform a standard habituation process to acclimate the mice to being handled by the experimenters, hunting within the experimental arena, and wearing cameras and an IMU while hunting. Following six 3 min sessions (over 1–2 days) of handling, the animals were placed in the prey capture arena to explore with their cagemates. The duration of this group habitu- ation was at least six 10 min sessions over 1–2 days. One cricket (Rainbow mealworms, 5 week old) per mouse was placed in the arena with the mice for the last half of the habituation sessions. For the subsequent habituation step, the mice were placed in the arena alone with one cricket for 7–10 min. This step was repeated for 2–3 training days (6–9 sessions) until most mice successfully caught crick- ets within the 10 min period.

Animals were then habituated to head-fixation above a spherical Styrofoam treadmill (Dombeck et al., 2007). Head fixation was only used to fit, calibrate, and attach cameras before experiments. Cameras were then fitted to each mouse (described below) and mice were habituated to wearing the cameras while walking freely in the arena, which took 1–2 sessions lasting 10 min. After the animals were comfortable with free locomotion with cameras, they were habituated to hunting with cameras attached. This took roughly one to two e hunting sessions of 10 min duration for each mouse. The animals were then food deprived for a period of 12–18 hr and then run in the prey capture assay for three 10 min sessions per data collection day. Although animals will hunt crickets without food restriction, this allowed for more trials within a defined experimental period.

The rectangular prey capture arena was a white arena of dimensions 38 x 45x30 cm Hoy et al. (2016). The arena was illuminated with a 15 Watt, 100 lumen incandescent light bulb placed roughly one meter above the center of the arena to mimic lux during dawn and dusk, times at which mice naturally hunt (Bailey and Sperry, 1929). Video signal was recorded from above the arena using a CMOS camera (Basler Ace, acA2000–165 umNIR, 30 Hz acquisition). Following the habituation process, cameras were attached and mice were placed in the prey capture arena with one cricket. Experimental animals captured and consumed the cricket before a new cricket was placed in the arena. The experimenters removed any residual cricket pieces in the arena before the addition of the next cricket. A typical mouse catches and consumes between 3–5 crickets per 10 min session. Control experiments were performed using the same methods, but with no cam- eras or IMU attached.

2.6.4 Surgical procedure

To allow for head-fixation during initial eye camera alignment, before the habituation process mice were surgically implanted with a steel headplate, following (Niell and Stryker, 2010). Animals were anesthetized with isoflurane (3% induction, 1.5%–2% maintenance, in O_2) and body temperature was maintained at 37.5°C using a feedback-controlled heating pad. Fascia was cleared from the surface of the skull following scalp incision and a custom steel headplate was attached to the skull using

Vetbond (3M) and dental acrylic. The headplate was placed near the back of the skull, roughly 1 mm anterior of Lambda. A flat layer of dental acrylic was placed in front of the headplate to allow for attachment of the camera connectors. Carprofen (10 mg/kg) and lactated Ringer's solution were administered subcutaneously and animals were monitored for three days following surgery.

2.6.5 Camera assembly and head-mounting

To measure eye position, we used miniature cameras that could be reversibly attached to the mouse's head via a chronically implanted Millmax connector. The cameras (1000 TVL Mini CCTV Camera; iSecurity101) were 5 x 6 x 6 mm with a resolution of 480x640 pixels and a 78 degree viewing angle, and images were acquired at 30Hz. Some of the cameras were supplied with a built in NIR blocking filter. For these cameras, the lens was unscrewed and the glass IR filter removed with fine forceps. A 200 Ohm resistor and 3mm IR LED were integrated onto the cameras for uniform illumination of the eyes. Power, ground, and video cables were soldered with lightweight 36 gauge FEP hookup wire (Cooner Wire; CZ 1174). A 6 mm diameter collimating lens with a focal distance of 12 mm (Lilly Electronics) was inserted into custom 3D printed housing and the cameras were then inserted and glued behind this (see Figure 2.1 for schematic of design). The inner side of the arm of the camera holder housed a male Mill-Max connector (Mill-Max Manufacturing Corp. 853-93-100-10-001000) cut to 5mm (2 rows of 4 columns), used for reversible attachment of the cameras to the implants of experimental animals. A custom IMU board with integrated 3-dimensional accelerometer and gyroscopes (Rosco Technologies) was attached to the top of one of the camera holders (see Figure 2.1B). The total weight of the two cameras together, with the lenses, connectors, 3D printed holders, and IMU was 2.6 grams. Camera assemblies were fitted onto the head by attaching them to corresponding female Mill-Max connectors. Cameras were located in the far lateral periphery of the mouse's visual field, roughly 100 degrees lateral of the head midline and 40 degrees above the horizontal axis. When the camera was appropriately focused on the eye, the female connectors were glued onto the acrylic implant using cyanoacrylate adhesive (Loctite). Because the connectors were each positioned during this initial procedure and permanently fixed in place, no adjustment of camera alignment was needed for subsequent experimental days.

2.6.6 Mouse and cricket tracking

Video data with timestamps for the two eyes and overhead camera were acquired at 30 frames per second using Bonsai (Lopes et al., 2015). We used DeepLabCut (Mathis et al., 2018) for markerless estimation of mouse and cricket position from overhead videos. For network training, we selected 8 points on the mouse head (nose, two camera connectors, two IR LEDs, two ears, and center of the head between the two ears), and two points for the cricket (head and body). Following estimation of the selected points, analysis was performed with custom MATLAB scripts.

Position and angle of the head were computed by fitting the 8 measured points on the head for each video frame to a defined mean geometry plus and x-y translation and horizontal rotation. The head direction was defined as the angle of this rotation, referenced to the line between the nose and center of the head. Following (Hoy et al., 2016), we defined approaches as times at which the velocity of the mouse was greater than 1 cm/sec, the azimuth of the mouse was between -45 and 45 degrees relative to cricket location, and the distance to the cricket was decreasing at a rate greater than 10 cm/sec.

Analog voltage signals from the IMU were recorded using a LabJack U6 at 50Hz sampling rate. Voltages from the accelerometer channels were median filtered with a window of 266.7 ms to remove rapid transients and converted to m/sec2, providing angular head orientation. Voltages from the gyroscope channels were converted to radians/sec without filtering, providing head rotation velocity.

2.6.7 Eye tracking and eye camera calibration

Video data with timestamps for the two eyes were acquired at 30fps using Bonsai. The video data are delivered by the camera in NTSC format, an interlaced video format in which two sequential images (acquired at 60fps) are interdigitated into each frame on alternate horizontal lines. We there- fore de-interlaced the video in order to restore the native 60fps resolution by separating out alter- nate lines of each image. We then linearly downsampled the resolution along the horizontal axis by a factor of two, to match spatial resolution in horizontal and vertical dimensions.

To track eye position, we used DeepLabCut (Mathis et al., 2018) to track eight points along the edge of the pupil. The eight points were then fit to an ellipse using the least-squares criterion. In order to convert pupil displacement into angular rotation, which cannot be calibrated by directed fixation as in primates, we followed the methods used in (Wallace et al., 2013). This approach is based on the principle that when the eye is pointed directly at the camera axis, the pupil is circular, and as the eye rotates, the circular shape flattens into an ellipse depending on the direction and degree of angular rotation from the center of the camera axis. To calculate the transformation of a circle along the camera axis to the ellipse fit, two pieces of information are needed: the camera axis center position and the scale factor relating pixels of displacement to angular rotation. To find the camera axis, we used the constraint that the major axis of the pupil ellipse is perpendicular to the vector from the pupil center to the camera axis center. This defines a set of linear equations for all of the pupil observations with significant ellipticity, which are solved directly with a least-squares solution. Next, the scale factor was estimated based on the equation defining how the ellipticity of the pupil changes with the corresponding shift from the camera center in each video frame. Based on the camera center and scale factor for each video, we calculated the affine transformation needed to transform the circle to the ellipse fit of the pupil in each frame, and the angular displacement from the camera axis was then used for subsequent analyses. Mathematical details of this method are presented in (Wallace et al., 2013).

Following computation of kinematic variables (mouse, cricket, and eye position/rotation), these values were linearly interpolated to a standard 30Hz timestamp to account for differences in acquisition timing across the multiple cameras and the IMU.

2.6.8 Quantification and Statistical Analyses

To cluster types of eye and head movements into compensatory and saccadic movements, we fit data from joint distributions of eye and head velocity to a Gaussian mixture model (Matlab). We used all recorded approach timepoints across animals and experiments (N=7 animals, 105 trials, 10026 timepoints) for this clustering. Using a model with three clusters revealed two compensatory groups (both clustering along the negative diagonal, which we merged) and one non-compensatory, which was used to define saccades for subsequent analysis. To determine periods when the animal was moving versus stationary, head movement speed was median filtered across a window of 500 ms and a threshold of 1cm/sec was applied..

Two-tailed paired t-tests or Wilcoxon Rank sum tests were used to compare data between non-approach and approach epochs. For comparisons between experimental and control groups, two-sample tests (Kolmogorov-Smirnov or two-sample two-tailed t-test) were used. Significance was defined as $p<0.05$, although p-values are presented throughout. In all figures, error bars represent ± the standard error of the mean or median, as appropriate.

2.7 ACKNOWLEDGMENTS

We thank Yichen Fan and Alyssa Fuentez for their help with behavioral data collection; the Wehr lab for assistance in implementing head-mounted cameras; members of the Niell lab for helpful discussions; and Dr. Marina Garrett, Dr. Clifford Keller, Dr. Jude Mitchell, and Dr. Matthew Smear for feedback on the manuscript. This work was supported by NIH R34NS111669 (C.M.N.) and the University of Oregon Promising Scholar Award (A.M.M.).

2.8 DECLARATION OF INTERESTS

The authors declare no competing interests.

2.9 SUPPLEMENTAL MATERIAL

Supplemental Movie 1 Video of mouse performing prey capture with reversibly attached eye cameras, demonstrating synchronized measurement of bilateral pupil positions and mouse/cricket behavior. The direction of each eye is superimposed on the head (purple and light blue lines) based on calculated pupil position and head angle.

Supplemental Movie 2 Video of mouse performing prey capture, demonstrating dynamics of head orienting (dark blue) and gaze direction (cyan). Note that during head turns the gaze is transiently offset from the head angle vector, due to compensatory eye movements, creating a stable image for the animal. Then, non-compensatory saccades shift the gaze position such that it aligns with the head to accurately target the cricket.

2.10 BRIDGE TO CHAPTER III

In this chapter, we investigated the strategies which mice use to target their gaze during prey capture. To study how visual processing occurs during natural behavior, we first needed to establish how the eye and head movements occur and change the visual field while mice are freely moving. We developed novel experimental and data analysis methods to accurately track the eye, which laid the foundation for the study of visual coding in freely moving mice. In Chapter III, we extend these methods to record neural activity and the visual scene simultaneously, and connect the different streams of information with a computational encoding model for the primary visual cortex. By building upon methods defined in this chapter, we then also showed how the integration of the neural activity, visual scene, and the eye/head position occurs in V1.

CHAPTER III

JOINT CODING OF VISUAL INPUT AND EYE/HEAD POSITION IN V1 OF FREELY MOVING MICE

3.1 JOURNAL STYLE INFORMATION

Originally published as *Parker, P. R. L., ***Abe, E. T. T.**, Leonard, E. S. P., Martins, D. M., Niell, C. M. (2022). Reproduced from Neuron https://doi.org/10.1016/j.neuron.2022.08.029

*Authors contributed equally

3.2 AUTHOR CONTRIBUTIONS

E.T.T.A., P.R.L.P., and C.M.N. conceived the project. E.T.T.A. led the design and implementation of computational analysis, and P.R.L.P. led the design and implementation of experiments. E.S.P.L. contributed to data collection. D.M.M. contributed to data pre-processing. E.T.T.A. and P.R.L.P. generated figures. E.T.T.A., P.R.L.P., and C.M.N. wrote the manuscript.

3.3 INTRODUCTION

A key aspect of natural behavior is movement through the environment, which has profound impacts on the incoming sensory information (Gibson, 1979). In vision, movements of the eyes and head due to locomotion and orienting transform the visual scene in ways that are potentially both beneficial, by providing additional visual cues, and detrimental, by introducing confounds due to self-movement. By accounting for movement, the brain can therefore extract more complete and robust information to guide visual perception and behavior. Accordingly, a number of studies have demonstrated the impact of movement on activity in cortex (Parker et al., 2020; Froudarakis et al., 2019; Busse et al., 2017). In head-fixed mice, locomotion on a treadmill increases the gain of visual responses (Niell and Stryker, 2010) and modifies spatial integration (Ayaz et al., 2013) in V1, while passive rotation generates vestibular signals (Bouvier et al., 2020; Vélez-Fort et al., 2018). Likewise, in freely moving mice and rats, V1 neurons show robust responses to head and eye movements and head orientation tuning (Guitchounts et al., 2020b,a; Meyer et al., 2018).

However, it is unknown how information about eye and head position is integrated into visual processing during natural movement, since studies of visual processing are generally performed during head-fixation to allow presentation of controlled stimuli, while natural eye and head movements require a mouse to be freely moving. Quantifying visual coding in freely moving animals requires determining the visual input, which is no longer under the experimenter's control and is dependent on both the visual scene from the mouse's perspective and its eye position. In addition, natural scenes, particularly during free movement, pose difficulties for data analysis since they contain strong spatial and temporal correlations and are not uniformly sampled because they are under the animal's control. Whether V1 receptive fields show similar properties under freely moving and restrained conditions is a question that goes back to the origins of cortical visual physiology (Hubel and Wiesel, 1962).

To address the experimental challenge, we combined high density silicon probe recordings with miniature head-mounted cameras (Michaiel et al., 2020; Meyer et al., 2018; Sattler and Wehr, 2021), with one camera aimed outwards to capture the visual scene from the mouse's perspective ("world camera"), and a second camera aimed at the eye to measure pupil position ("eye camera"), as well as an inertial measurement unit (IMU) to quantify head orientation. To address the data analysis challenge, we implemented a paradigm to correct the world camera video based on measured eye movements with a shifter network (Yates et al., 2021; Walker et al., 2019) and then use this as input to a generalized linear model (GLM) to predict neural activity (Pillow et al., 2008).

Using this approach, we first quantified the visual encoding alone during free movement, in terms of linear spatiotemporal receptive fields (RFs) from the GLM fit. For many units, the RF measured during free movement is similar to the RF measured with standard white noise stimuli during head-fixation within the same experiment, providing confirmation of this approach. We then extended the encoding model to incorporate eye position and head orientation, and found that these generally provide a multiplicative gain on the visual response. Together, this work provides new insights into the mechanisms of visual coding in V1 during natural movement, and opens the door to studying the neural basis of behavior under ethological conditions.

3.4 RESULTS

3.4.1 A generalized linear model accurately estimates spatiotemporal receptive fields during free movement

In order to study how visual processing in V1 incorporates self-motion, we developed a system to perform visual physiology in freely moving mice (Figure 3.1A). To estimate the visual input reaching the retina, a forward-facing world camera recorded a portion (∼120 deg) of the visual scene available to the right eye. A second miniature camera aimed at the right eye measured pupil position, and an IMU tracked head orientation. Finally, a driveable linear silicon probe implanted in left V1 recorded the activity of up to 100+ single units across layers. The same neurons were first recorded under head-fixed conditions to perform white noise RF mapping, and then under conditions of free movement (Figure 3.1B). Well isolated units were highly stable across the two conditions (Figure S3.1 and Methods). Figure 3.1C and Video S3.1 show example data obtained using this system in a freely moving animal. Mice were allowed to explore a visually complex arena containing black and white blocks (three-dimensional sparse noise), static white noise and oriented gratings on the walls, and a monitor displaying moving spots. After several days of habituation, mice were active for a majority of the time spent in the arena (82%), with an average movement speed of 2.6 cm/s, which is comparable to other similar studies (see Methods; (Juavinett et al., 2019; Meyer et al., 2018).

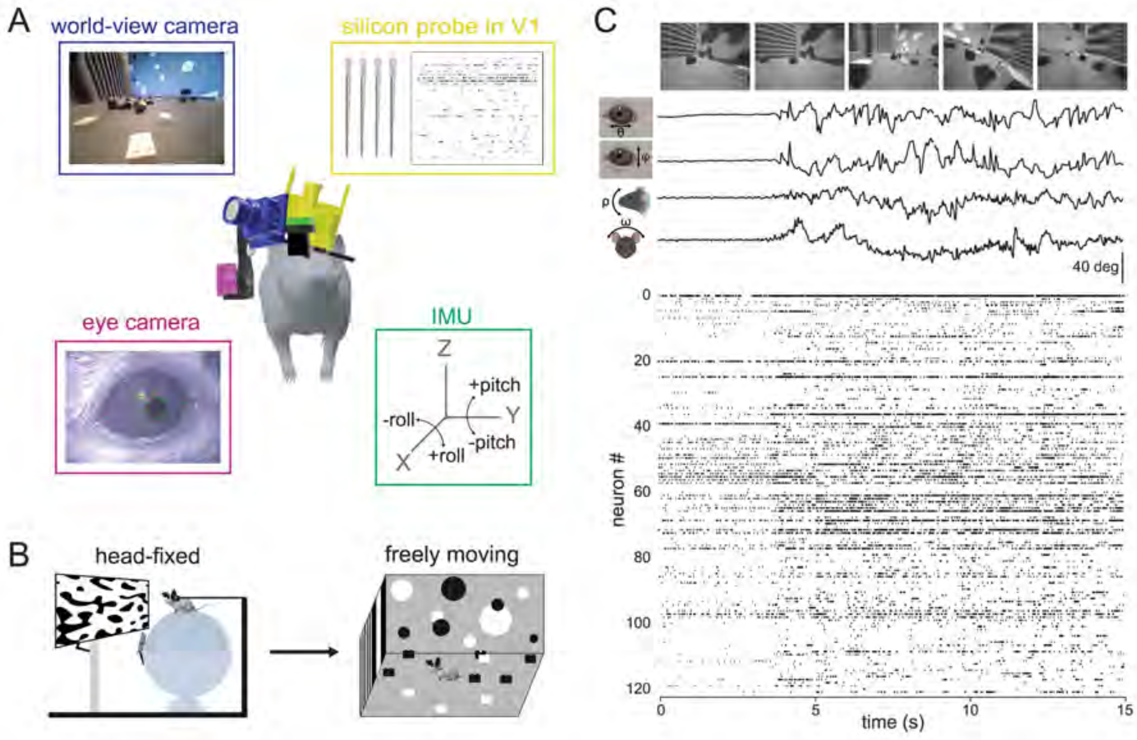


Figure 3.1. **A)** Schematic of recording preparation including 128-channel linear silicon probe for electrophysiological recording in V1 (yellow), miniature cameras for recording the mouse's eye position (magenta) and visual scene (blue), and inertial measurement unit for measuring head orientation (green). **B)** Experimental design: controlled visual stimuli were first presented to the animal while head-fixed, then the same neurons were recorded under conditions of free movement. **C)** Sample data from a fifteen second period during free movement showing (from top) visual scene, horizontal and vertical eye position, head pitch and roll, and a raster plot of over 100 units. Note that the animal began moving at ~4 secs, accompanied by a shift in the dynamics of neural activity.

To quantify visual coding during free movement, both the neural activity and the corresponding visual input are needed. The world camera captures the visual scene in a head-centric point of view, while the visual input needed is in a retinocentric perspective. To tackle this problem, we used a shifter network to correct the world camera video for eye movements (Walker et al., 2019; Yates et al., 2021). The shifter network takes as input the horizontal (theta) and vertical (phi) eye angle, along with the vertical head orientation (pitch) to approximate cyclotorsion (Wallace et al., 2013), and outputs the affine transformation for horizontal and vertical translation and rotation, respectively (Figure S3.2). We trained the shifter network and a GLM end-to-end with a rectified linear activation function to determine the camera correction parameters that best enable prediction of neural activity for each recording session (Figure 3.2A). All GLM fits in this study were cross-validated using train-test splits (see Methods for details). This analysis draws on the relatively large numbers of simultaneously recorded units as it determines the best shift parameters by maximizing fits across all neurons, thereby determining the general parameters of the eye camera to world camera transformation rather than being tailored to individual neurons.

The outputs of the shifter network (Figure S3.2A-C) show that it converts the two axes of eye rotation (in degrees) into a continuous and approximately orthogonal combination of horizontal and vertical shifts of the worldcam video (in pixels), as expected to compensate for the alignment of the horizontal and vertical axes of the eye and world cameras. These outputs were also consistent in cross-validation across subsets of the data (coefficient of determination R^2, dx=0.846, dy=0.792, $d\alpha$=0.945; Figure S2A-C). When the shifts were applied to the raw world camera video it had the qualitative effect of stabilizing the visual scene in between rapid gaze shifts, as would be expected from the vestibulo-ocular reflex and "saccade-and-fixate" eye movement pattern described previously in mice (Video S3.2; (Meyer et al., 2020; Michaiel et al., 2020). We quantified this by computing the total horizontal and vertical displacement of the raw and shifted world camera video based on image registration between sequential frames. When corrected for eye position, continuous motion of the image

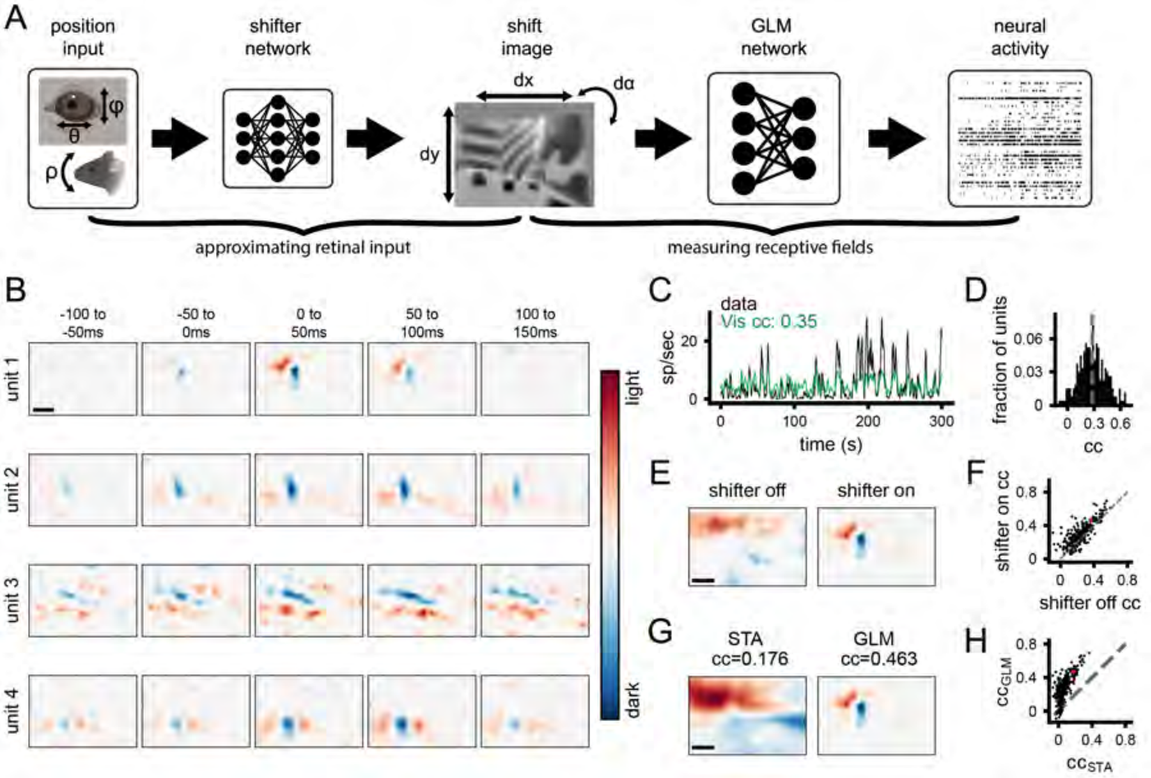


Figure 3.2. **A)** Schematic of processing pipeline. Visual and positional information is used as input into the shifter network, which outputs parameters for an affine transformation of the world-camera image. The transformed image frame is then used as the input to the GLM network to predict neural activity. **B)** Four example freely moving spatiotemporal visual receptive fields. Scale bar for RFs represents 10 degrees. **C)** Example actual and predicted smoothed (2 s window) firing rates for unit 3 in B. **D)** Histogram of correlation coefficients (cc) for the population of units recorded. Average cc shown as gray dashed line. **E)** Example of a freely moving RF with the shifter network off (left) and on (right) at time lag 0 ms. Colormap same as B. **F)** Scatter plot showing cc of predicted versus actual firing rate for all units with the shifter network off vs on. Red point is the unit shown in E. **G)** Example receptive field calculated via STA (left) versus GLM (right). **H)** Scatter plot showing cc of predicted vs actual firing rate for all units, as calculated from STA or GLM. Red point is the unit shown in G.

is converted into the step-like pattern of saccade-and-fixate (Figure S3.2D) and the image is stabilized to within 1 deg during the fixations (Figure S3.2E,F; (Michaiel et al., 2020). This eye-corrected retinocentric image was then used as input for the GLM network to predict neural activity in subsequent analysis.

We estimated spatiotemporal RFs during free movement using a GLM to predict single-unit activity from the corrected world camera data. Single-unit RFs measured during free movement had clear on and off sub-regions and a transient temporal response (Figure 3.2B). To our knowledge, these are the first visual receptive fields measured from a freely moving animal. It should be noted that the temporal response is still broader than would be expected, which likely reflects the fact that the GLM cannot fully account for strong temporal correlations in the visual input. Furthermore, the GLM predicted the continuous time-varying firing rate of units during free movement (Figure 3.2C). Across the population of neurons recorded (N=268 units, 4 animals), neural activity predicted from the corrected world camera data was correlated with the actual continuous firing rate (CC mean 0.28, max 0.69; Figure 3.2D). These values are on par with those obtained from mapping V1 RFs in awake and anesthetized head-fixed animals (Carandini et al., 2005).

To demonstrate the impact of correcting the visual input for eye movements, we computed RFs from the raw, uncorrected world camera data. This resulted in single-unit RFs becoming blurred, and reduced the ability to predict neural activity (Figure 3.2E,F; shifter on vs. off p=8.17e-23, paired t-test). Nonetheless, it is notable that the overall improvement was modest (mean increase in cc=0.06) and although some units required the shifter network, many units maintained a similar ability to predict firing rate even without the shifter. This is perhaps due to the large size of receptive fields relative to the amplitude of eye movements in the mouse (see

Discussion). To determine the relative benefit of the GLM approach relative to a simpler reverse correlation spike-triggered average (Chichilnisky, 2001), we compared receptive fields and ability to predict firing rate from these two methods (Figure 3.2G-H). Receptive fields from the STA were much broader and appeared to reflect structure from the environment (Figure 3.2G), as expected since the STA will not account for spatiotemporal correlations in the input. Correspondingly, the STA performed much worse than the GLM in predicting neural activity (Figure 3.2H; p=2e-93). Finally, as an additional verification that the GLM method is able to accurately reconstruct RFs from limited data and that natural scene statistics are not biasing the RF estimates, we simulated neural activity based on Gabor RFs applied to the world camera data. The results demonstrate that the GLM can reconstruct simulated RFs with high accuracy, resulting in reconstructed RFs that are both qualitatively and quantitatively similar to the original (Figure S3.2F,G).

3.4.2 Comparison of receptive fields measured under freely moving versus head-fixed conditions

To determine whether RFs measured during free movement were comparable to those measured using traditional visual physiology methods, we compared them to RFs measured using a white noise stimulus under head-fixed conditions. The large majority of units were active (mean rate >1Hz) during each of these conditions (Figure 3.3A) and in each condition roughly half the units had a fit that significantly predicted neural activity, with slightly more in freely moving (Figure 3.3A). Overall, many neurons that had a clear white noise RF also had a clear RF from freely moving data (Figure 3.3B), which closely matched in spatial location, polarity, and number of sub-regions. To quantitatively compare RFs, we calculated the pixel-wise correlation coefficient between them. To provide a baseline for this metric, we first performed a cross-validation test-retest by comparing the RFs from the first and second half of each recording separately (Figure S3.3). The mean test-retest cc was 0.46 for head-fixed and 0.58 and freely moving. We considered a unit to have a robust test-retest RF if this pixel-wise cc was greater than 0.5 (Figure S3.3C), and then evaluated the similarity of RFs for units that had robust fits in both conditions. The distribution of correlation coefficients between head-fixed and freely moving RFs for these units (Figure 3.3C) shows a strong correspondence for RFs across the two conditions (Figure 3.3C; 74% of units had a significant cc versus shuffled data). Taken together, these results show that for the units that had clearly defined RFs in both conditions, RFs measured with freely moving visual physiology are similar to those measured using traditional methods, despite the dramatically different visual input and behavior between these two conditions.

3.4.3 V1 integrates visual and position signals

Studies in head-fixed mice have shown a major impact of locomotion and arousal on activity in visual cortex (Busse et al., 2017; Niell and Stryker, 2010; Ayaz et al., 2013; Vinck et al., 2015). However, the impact of postural variables such as head

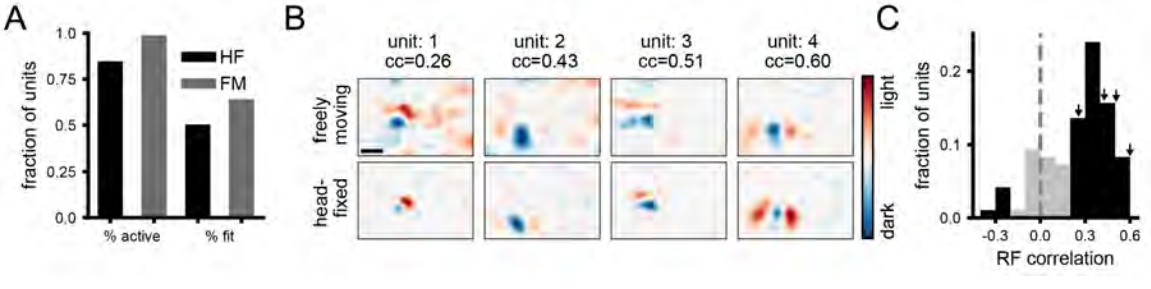


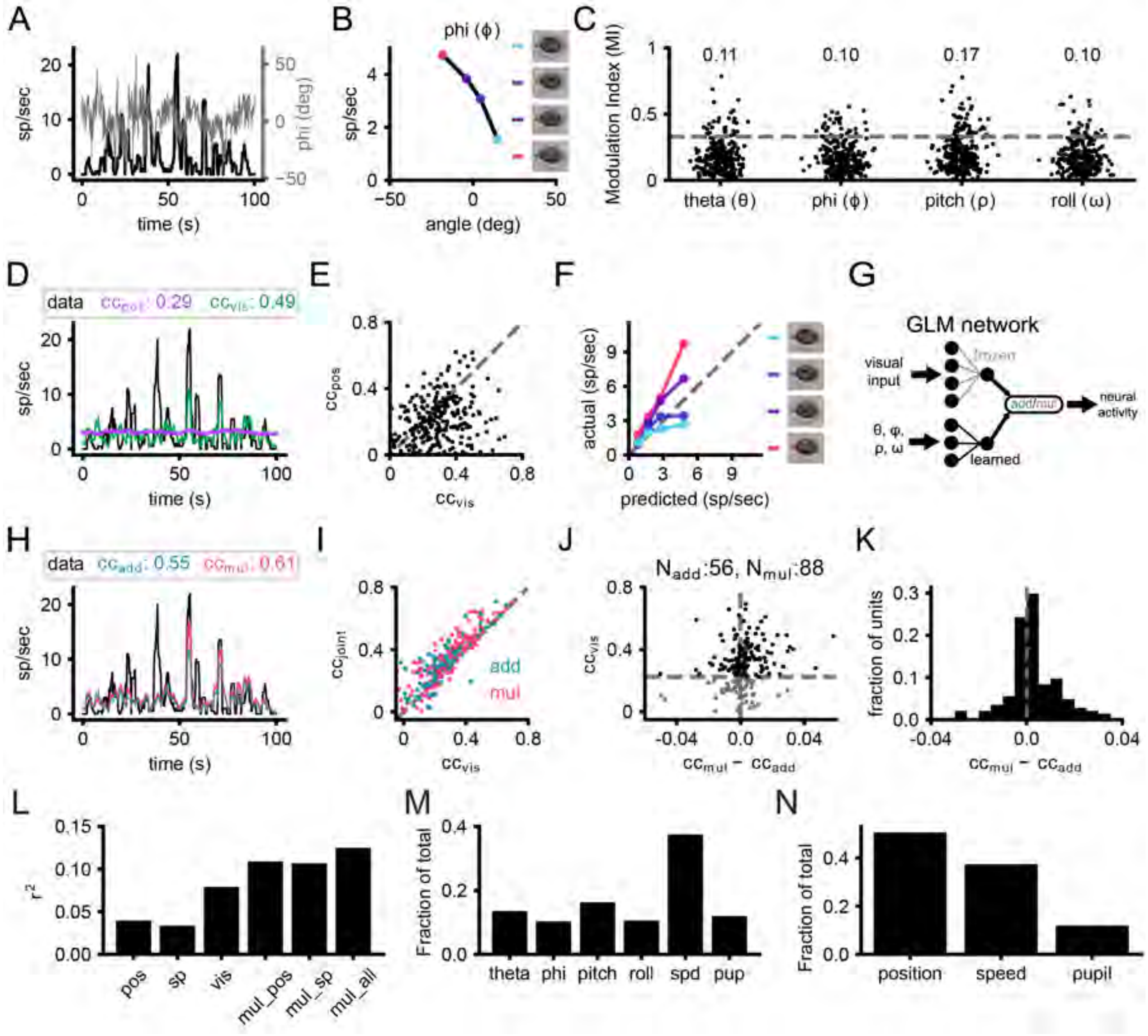
Figure 3.3. **A)** Fraction of units that were active (>1 Hz firing rate) and that had significant fits for predicting firing rate, in head-fixed and freely moving conditions. **B)** Example spatial receptive fields measured during free movement (top) and using a white noise mapping stimulus while head-fixed (bottom) at time lag 0 ms. Scale bar in top left is 10 deg. **C)** Histogram of correlation coefficients between freely moving and head-fixed RFs. Black color indicates units that fall outside two standard deviations of the shuffle distribution. Arrows indicate locations in the distribution for example units in A.

position and eye position are not easily studied in head-fixed conditions, particularly since eye movements are closely coupled to head movement (Meyer et al., 2020; Michaiel et al., 2020). We therefore sought to determine whether and how eye/head position modulate V1 neural activity during free movement, based on measurement of pupil position from the eye camera and head orientation from the IMU. Strikingly, many single units showed tuning for eye position and/or head orientation, with 25% (66/268) of units having a modulation index ($MI = \frac{rate_{max} - rate_{min}}{rate_{max} + rate_{min}}$) greater than 0.33 for at least one position parameter, which equates to a two-fold change in firing rate (Figure 3.4A-C). To determine whether single-unit activity was better explained by visual input or eye/head position, we fit GLMs using either one as input. For most units (189/268 units, 71%), firing rate was better explained by a visual model, although the activity of some units was better explained by eye/head position (Figure 3.4D,E; 78/268 units, 29%). It should be noted that the units that were better fit by position model might nonetheless be better described by a more elaborate visual model.

To gain a qualitative understanding of how V1 neurons might combine visual and position information, we plotted predicted firing rates from visual-only GLM fits against the actual firing rates binned into quartiles based on eye/head position (example in Figure 3.4F). While the data should lie on the unity line in the absence of position modulation, additive integration would shift the entire curve up or down, and multiplicative integration would cause a slope change. Across the population of recorded neurons, many units showed evidence of gain modulation that tended to appear more multiplicative than additive.

To directly quantify the integration of visual and eye/head position information, and in particular to test whether this was additive or multiplicative, we trained two additional models: additive and multiplicative joint-encoding of visual and position information. To train the joint fit of visual and position signals, we froze the weights of the initial visual fits and trained positional weights that either added to or multiplied the visual signal for each unit (Figure 3.4G). Incorporating eye position and head orientation enables the model to more accurately predict large changes in the firing rate (Figure 3.4H). The inclusion of positional information almost universally improved predicted neural activity compared to visual fits alone (Figure 3.4I). For units that had a significant visual fit ($cc > 0.22$, cross-validated, N=173 units), incorporating positional information resulted in an average fractional increase in correlation of 34% (0.07 average increase in cc). Multiplicatively combining visual and positional signals generated predictions that more closely matched actual firing rates than an additive combination in a majority of units (Figure 3.4J,K; p=0.0005,

Figure 3.4. **A)** Overlay of vertical eye angle (phi; gray) and the smoothed firing rate of an example unit (black). **B)** Example tuning curve for head pitch. Colored points denote the quartiles of phi corresponding to panel F. **C)** Scatter of the modulation indices for eye position and head orientation (N=268 units, 4 animals). Numbers at top of the plot represent the fraction of units with significant tuning. **D)** Same unit as A. Example trace of smoothed firing rates from neural recordings and predictions from position-only and visual-only fits. **E)** Scatter plot of cc for position-only and visual-only fits for all units. **F)** Gain curve for the same unit in A and C. Modulation of the actual firing rates based on phi indicated by color. **G)** Schematic of joint visual and position input training. **H)** Same unit as A, C, and E. Smoothed traces of the firing rates from the data, additive and multiplicative fits. **I)** Correlation coefficient for visual-only versus joint fits. Each point is one unit, color coded for the joint fit that performed best. **J)** Comparison of additive and multiplicative fits for each unit. Units characterized as multiplicative are to the right of the vertical dashed line, while additive ones are to the left. Horizontal dashed line represents threshold set for the visual fit, since in the absence of a predictive visual fit, a multiplicative modulation will be similar to an additive modulation. **K)** Histogram of the difference in cc between additive and multiplicative models. The visual threshold from I was applied to the data. **L)** Explained variance (r^2) for position only (pos), speed and pupil only (sp), visual only (vis), multiplicative with eye/head position (mul_{pos}), multiplicative with speed and pupil (mul_{sp}), and multiplicative with eye/head position, speed and pupil (mul_{all}). **M)** The fraction of contribution of the weights for multiplicative fits with eye/head position, speed (spd) and pupil (pup). **N)** Same as M but summing together the contribution for eye/head position.



one sample t-test $cc_{mult} - cc_{add}$ for units with significant visual-only fits versus gaussian distribution with mean=0), suggesting visual and position signals in mouse V1 are more often integrated nonlinearly, consistent with previous studies in primate visual and parietal cortex (Andersen and Mountcastle, 1983; Morris and Krekelberg, 2019).

To further characterize the head and eye position modulations, we performed additional experiments recording V1 activity during free movement in nearly total darkness, followed by recording in the standard light condition. A significant fraction of neurons were modulated by at least 2:1 in the dark (Figure S3.4A,B; dark: 17%, 41/241; light: 31%, 75/241 units). Comparing the degree of modulation in the light vs dark for individual units revealed that the degree of tuning often shifted (Figure S3.4C), with some increasing their position tuning (consistent with an additive modulation that has a proportionally larger effect in the absence of visual drive) and others decreasing their position tuning (consistent with a multiplicative modulation that is diminished in the absence of a visual signal to multiply). In addition, to test whether position modulation might result from the abrupt transition from head-fixed recordings to free movement, we compared the degree of modulation during the first and second half of free movement sessions, and found no consistent change (Figure S3.4D). Finally, to test whether there was a bias in tuning for specific eye/head positions (e.g., upward versus downward pitch), we examined the weights of the position fits, which showed distributions centered around zero (Figure S3.4E), indicating that tuning for both directions was present for all position parameters, across the population.

Many response properties have been shown to vary across the cell types and layers of mouse V1 (Niell and Scanziani, 2021). Separating recorded units into putative excitatory and inhibitory, based on spike waveform as performed previously (Niell and Stryker, 2008), demonstrated that the visual fit performed better than than head/eye position for putative excitatory neurons, while the contributions were roughly equal for putative inhibitory cells (Figure S3.4F). This may be explained by the fact that putative excitatory neurons in mouse V1 have more linear visual responses (Niell and Stryker, 2008). We also examined whether the contribution of visual versus position information varied by laminar depth, and found no clear dependence (Figure S3.4G,H).

Finally, we examined the role of two factors that are known to modulate activity in mouse V1: locomotor speed and pupil diameter (Vinck et al., 2015; Niell and Stryker, 2010; Reimer et al., 2014). It is important to note that our GLM analysis excludes periods when the head is completely still, since that leads to dramatic over-representation of specific visual inputs that presents a confound in fitting the data. Therefore, the results presented above do not include the dramatic shift from non-alert/stationary to alert/moving that has been extensively studied (McGinley et al., 2015). Furthermore, changes in locomotor speed during free movement are associated with other changes (e.g., optic flow) that do not (occur under head-fixed locomotion, thus the model weights may represent other factors besides locomotion per se. Nonetheless, we find that including speed and pupil in the fit does indeed predict a part of the neural activity (Figure 3.4L). However this does not occlude the contribution from head/eye position or visual input. Examination of the weights in a joint fit of all parameters together demonstrates that although the contribution of locomotor speed is greater than any one individual position parameter (Figure 3.4M), the summed weights of head/eye position parameters are still the largest contribution (Figure 3.4N). It is also interesting to note that although head and eye position are often strongly correlated in the mouse due to compensatory eye movements (Michaiel et al., 2020; Meyer et al., 2020), the weights for each of these parameters are roughly equal in the GLM fit that can account for these correlations (Figure 3.4M), demonstrating that both head and eye may contribute independently to coding in V1, in addition to known factors such as locomotion and arousal.

3.5 DISCUSSION

Nearly all studies of neural coding in vision have been performed in subjects that are immobilized in some way, ranging from anesthesia to head and/or gaze fixation, which greatly limits the ability to study the visual processing that occurs as an animal moves through its environment. One important component of natural movement is the integration of the incoming visual information with one's position relative to the scene. In order to determine how individual neurons in mouse V1 respond to visual input and eye/head position, we implemented an integrated experimental and model-based data analysis approach to perform visual physiology in freely moving mice. Using this approach, we demonstrate the ability to estimate spatiotemporal visual receptive fields during free movement, show that individual neurons have diverse tuning to head and eye position, and find that these signals are often combined through a multiplicative interaction.

3.5.1 Integration of visual input and eye/head position

The ongoing activity of many units in V1 was modulated by both eye position and head orientation, as demonstrated by empirical tuning curves (Figure 3.4B) and model-based prediction of neural activity based on these parameters (Figure 3.4D). Modulation of neural activity in V1 and other visual areas by eye position (Weyand and Malpeli, 1993; Trotter and Celebrini, 1999; Rosenbluth and Allman, 2002; Durand et al., 2010; Andersen and Mountcastle, 1983) and head orientation (Guitchounts et al., 2020b; Brotchie et al., 1995) has been observed across rodents and primates, and fMRI evidence suggests human V1 encodes eye position (Merriam et al., 2013). Similar encoding of postural variables was also reported in posterior parietal cortex and secondary motor cortex using a GLM-based approach (Mimica et al., 2018). Many of the position-tuned units we observed were also visually responsive, with clear spatiotemporal receptive fields.

In order to determine how these position signals were integrated with visual input, we used the GLM model trained on visual input only and incorporated either an additive or multiplicative signal based on a linear model of the eye/head position parameters. For neurons that had both a significant visual and position component, we found that the majority were best described by a multiplicative combination. This multiplicative modulation corresponds to a gain field, a fundamental basis of neural computation (Salinas and Abbott, 1996; Salinas and Sejnowski, 2001). Gain fields have been shown to serve a number of roles, including providing an effective mechanism for coordinate transformations as they enable direct readout of additive or subtractive combinations of input variables, such as the transformation from retinotopic to egocentric position of a visual stimulus. Studies in head-fixed primates have demonstrated gain fields for eye position (Morris and Krekelberg, 2019; Andersen and Mountcastle, 1983; Salinas and Sejnowski, 2001) and head orientation (Brotchie et al., 1995), and similar gain modulation for other factors such as attention (Salinas and Abbott, 1997). The demonstration of gain modulation by eye/head position in freely moving mice shows that this mechanism is engaged under natural conditions with complex movement.

Given the presence of gain fields in mouse visual cortex, two immediate questions arise: what are the sources of the position signals, and what are the cellular/circuit mechanisms that give rise to the gain modulation? Regarding sources, evidence suggests eye position signals arrive early in the visual system, perhaps even at the level of the thalamic lateral geniculate nucleus (Lal and Friedlander, 1990), while head orientation information could be conveyed through secondary motor cortex (Guitchounts et al., 2020b) retrosplenial cortex (Vélez-Fort et al., 2018) or from neck muscle afferents (Crowell et al., 1998). Regarding the mechanism, multiplicative interactions have been suggested to arise from synaptic interactions including active dendritic integration, recurrent network interactions, changes in input synchrony, balanced excitatory/inhibitory modulatory inputs, and classic neuromodulators (Salinas and Abbott, 1996; Salinas and Sejnowski, 2001; Silver, 2010). Future research could take advantage of genetic methods available in mice to determine the neural circuit mechanisms that implement this computation (O'Connor et al., 2009; Niell and Scanziani, 2021; Luo et al., 2008).

This multiplicative interaction can also be viewed as a form of nonlinear mixed selectivity, which has been shown to greatly expand the discriminative capacity of a neural code (Rigotti et al., 2013; Nogueira et al., 2021). The implications of nonlinear mixed selectivity have primarily been explored in the context of categorical variables, rather than continuous variables as observed here. In this context it is interesting to note that a significant number of units were nonetheless best described by an additive interaction. In an additive interaction the two signals are linearly combined, providing a factorized code where each signal can be read out independently. It may be that having a fraction of neurons using this linear interaction provides flexibility by which the visual input and position can be directly read out, along with the nonlinear interaction that allows computations such as coordinate transformations.

3.5.2 Methodological considerations

We estimated the visual input to the retina based on two head-mounted cameras – one to determine the visual scene from the mouse's perspective, and one to determine eye position and thereby correct the head-based visual scene to account for eye movements. Incorporation of eye position to correct the visual scene significantly improved the ability to estimate receptive fields and predict neural activity. Although head-fixed mice only make infrequent eye movements, freely moving mice (and other animals) make continual eye movements that both stabilize gaze by compensating for head movements and shift the gaze via saccades (Michaiel et al., 2020; Meyer et al., 2020). As a result, eye position can vary over a range of ±30 degrees (theta std: 16.5 deg, phi std: 17.8 deg in this study). Indeed, without eye movement correction many units did not have an estimated receptive field with predictive power (Figure 3.2F). Nonetheless, it is notable that some units were robustly fit even without correction – this likely reflects that fact that the eye is still within a central location a large fraction of the time (63% of timepoints within ±15 deg for theta, phi) and typical receptive fields in mouse V1 are on the order of 10-20 degrees (Niell and Stryker, 2008; Van den Bergh et al., 2010).

We estimated spatiotemporal receptive fields and predicted neural activity during free movement using a GLM – a standard model-based approach in visual physiology (Pillow et al. 2008). Despite its simplicity – it estimates the linear kernel of a cell's response – the GLM approach allowed us to estimate receptive fields in many neurons (39% of freely moving RFs significantly matched head-fixed white-noise RFs). These results are comparable to the fraction of units with defined STA receptive fields measured in head-fixed mice (64% of simple cells, 34% of total population in (Niell and Stryker, 2008); 49% of total population in (Bonin et al., 2011). The model fits were also able to predict a significant amount of ongoing neural activity (cc mean=0.29, max=0.73). Although this is still generally a small fraction of total activity, this is in line with other studies (Carandini et al., 2005; de Vries et al., 2020) and likely represents the role of additional visual features beyond a linear kernel, as well as other non-visual factors that modulate neural activity (Musall et al., 2019; Stringer et al., 2019; Niell and Stryker, 2010). A more elaborate model with nonlinear interactions would likely do a better job of explaining activity in a larger fraction of units; indeed, "complex" cells (Hubel and Wiesel, 1962) are not accurately described by a single linear kernel. However, for this initial characterization of receptive fields in freely moving animals, we chose to use the GLM since it is a well-established method, it is a convex optimization guaranteed to reach a unique solution, and the resulting model is easily interpretable as a linear receptive field filter. The fact that even such a simple model can capture many neurons' responses both shows the robustness of the experimental approach, and opens up the possibility for the use of more elaborate and nonlinear models, such as multi-component (Butts, 2019) or deep neural networks (Walker et al., 2019; Ukita et al., 2019; Bashivan et al., 2019). Implementation of such models may require extensions to the experimental paradigm such as longer recording times to fit a greater number of parameters.

3.5.3 Freely moving visual physiology

Visual neuroscience is dominated by the use of head-restrained paradigms, in which the subject cannot move through the environment. As a result, many aspects of how vision operates in the natural world remain unexplored (Parker et al., 2020; Leopold and Park, 2020). Indeed, the importance of movement led psychologist J. J. Gibson to consider the legs a component of the human visual system, and provided the basis for his ecological approach to visual perception (Gibson, 1979). The methods we developed here can be applied more broadly to enable a Gibsonian approach to visual physiology that extends beyond features that are present in standard head-fixed stimuli. While natural images and movies are increasingly used to probe responses of visual neurons in head-fixed conditions, these are still dramatically different from the visual input received during free movement through complex three-dimensional environments. This includes cues resulting from self-motion during active vision, such as motion parallax, loom, and optic flow that can provide information about the three-dimensional layout of the environment, distance, object speed, and other latent variables. Performing visual physiology in a freely moving subject may facilitate the study of the computations underlying these features.

Accordingly, a resurgent interest in natural behaviors (Juavinett et al., 2018; Datta et al., 2019; Dennis et al., 2021; Miller et al., 2022) provides a variety of contexts in which to study visual computations in the mouse. However, studies of ethological visual behaviors typically rely on measurements of neural activity made during head-fixation, rather than during the behavior itself (Hoy et al., 2019; Boone et al., 2021). Freely moving visual physiology is a powerful approach that ultimately can enable quantification of visual coding during ethological tasks to determine the neural basis of natural behavior.

3.6 ACKNOWLEDGEMENTS

We thank Drs. Michael Beyeler, Michael Goard, Alex Huk, David Leopold, Jude Mitchell, Matt Smear and members of the Niell lab for comments on the manuscript. We thank Geordi Helmick and the University of Oregon Technical Sciences Administration for assistance with graphic and hardware design/production, Dr. Yuta Senzai for assistance with chronic electrode implants and drive design, and Jonny Saunders for assistance with sensor fusion analysis. This work was supported by NIH grants R34NS111669-01, R01NS121919-01, and UF1NS116377 (C.M.N.).

3.7 DECLARATION OF INTERESTS

The authors declare no competing interests.

3.8 STAR METHODS

3.8.1 Key Resource Table

Reagent or Resource	**Source**	**Identifier**
Deposited data:		
Processed model data	This paper	link
Experimental Models: Organisms/Strains		
Mouse: C57BL/6J	Jackson Laboratories and bred in-house	Strain code: 027
Software and Algorithms		
Python 3.8	https://www.python.org/	RRID: SCR_008394
Open Ephys plugin-GUI	http://www.open-ephys.org/	link
Bonsai	https://open-ephys.org/bonsai	link
DeepLabCut	(Mathis et al. 2018)	link
Kornia	(Riba et al., 2019)	link
Data extraction and analysis code	This paper	link

Reagent or Resource	**Source**	**Identifier**
PyTorch	https://pytorch.org/	link
Other		
Open Ephys acquisition board	Open Ephys	link
Open Ephys I/O board	Open Ephys	link
P64-3 or P128-6 silicon probe	Diagnostic Biochips	link
RHD SPI interface cable, 6ft ultra-thin	Intan	link
3-D printed electrophysiology drive	Yuta Senzai (UCSF) / in-house design	custom
3-D printed camera arm	In-house design	custom
1000TVL NTSC miniature camera	iSecurity101	No longer available

Reagent or Resource	**Source**	**Identifier**
BETAFPV C01 miniature camera	BETAFPV	link
940nm 3mm IR LED	Chanzon	link
Animal head tracking device	Rosco Technologies	link
Mill-Max connector 853-93-100-10-001000	Digi-Key	link
FEP hookup wire 36 AWG CZ1174	Cooner	link
USB3HDCAP USB3 video capture device	Startech	link
Dazzle DVD recorder HD	Pinnacle	link
Black Fly S USB3 (BFS-U3-16S2M-CS)	Teledyne FLIR	link

Reagent or Resource	**Source**	**Identifier**
GW2780 OLED monitor	BenQ	link
GW2480 OLED monitor	BenQ	link
Mouse bungee (Version 1)	Razer	link
Unifast LC	GC America	link
DOWSIL 3-4680 silicone gel kit	Dow	link

Table 2. Chapter III: Key Resource Table

3.8.2 Resource availability

3.8.2.1 Lead Contact

Further information and requests for resources should be directed to and fulfilled by the lead contact, Dr. Cristopher M Niell (cniell@uoregon.edu).

3.8.2.2 Materials availability

This study did not generate new unique reagents.

3.8.2.3 Data and code availability

- All model data have been deposited at Data Dryad and are publicly available as of the date of publication. The DOI is listed in the key resources table.

- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table.

- Any additional information required to reanalyze the data reported in this work paper is available from the Lead Contact upon request.

3.8.3 Experimental model and subject details

3.8.3.1 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. Three- to eight-month old adult mice (C57BL/6J, Jackson Laboratories and bred in-house) were kept on a 12 h light/dark cycle. In total, 4 female and 3 male mice were used for this study (head-fixed/freely moving: 2 females, 2 males; light/dark: 3 females, 2 males).

3.8.4 Method details

3.8.4.1 Surgery and habituation

Mice were initially implanted with a steel headplate over primary visual cortex to allow for head-fixation and attachment of head-mounted experimental hardware. After three days of recovery, widefield imaging (Wekselblatt et al., 2016) was performed to help target the electrophysiology implant to the approximate center of left monocular V1. A miniature connector (Mill-Max 853-93-100-10-001000) was secured to the headplate to allow attachment of a camera arm (eye/world cameras and IMU; (Michaiel et al., 2020)). In order to simulate the weight of the real electrophysiology drive and camera system for habituation (6 g total), a 'dummy' system was glued to the headplate. Animals were handled by the experimenter for several days before surgical procedures, and subsequently habituated (~45 min) 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 3D-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 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.

3.8.4.2 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 fps interlaced), an infrared-LED 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 30 kHz, downsampled to 300 Hz and interpolated to camera data). Fine gauge wire (Cooner, 36 AWG, #CZ1174CLR) connected the IMU to its control 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 OpenEphys acquisition system via an ultra thin cable (Intan #C3216). The electrophysiology cable was looped over a computer mouse bungee (Razer) to reduce the combined impact of the cable and implant. We first used the OpenEphys GUI (https://open-ephys.org/gui) to assess the quality of the electrophysiology data, then recordings were performed in Bonsai (Lopes et al., 2015) using custom workflows. 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 receptive fields using traditional methods, then were transferred to an arena where they could freely explore. Recording duration was approximately 45 minutes head-fixed, and 1hr freely moving. For head-fixed experiments, a 27.5 in monitor (BenQ GW2780) was placed approximately 27.5 cm from the mouse's right eye. A contrast-modulated white noise stimulus (Niell and Stryker, 2008) was presented for 15 min, followed by additional visual stimuli, and the mouse was then moved to the arena. The arena was approximately 48 cm long by 37 cm wide by 30 cm high. A 24 in 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-spatial frequency gratings and white noise. A moving black and white spots stimulus (Piscopo et al., 2013) played continuously on the monitor while the mouse was in the arena. The floor was a gray silicone mat (Gartful) and was densely covered with black and white Legos. 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.

3.8.4.3 Data preprocessing

Electrophysiology data were acquired at 30 kHz and bandpass filtered between 0.01 Hz and 7.5 kHz. Common-mode noise was removed by subtracting the median across all channels at each timepoint. Spike sorting was performed using Kilosort 2.5

(Steinmetz et al., 2021), and isolated single units were then selected using Phy2 (https://github.com/cortex-lab/phy) based on a number of parameters including contamination (<10%), firing rate (mean >0.5 Hz across entire recording), waveform shape, and autocorrelogram. Electrophysiology data for an entire session were concatenated (head fixed stimulus presentation, freely moving period, or freely moving light and dark) and any sessions with apparent drift across the recording periods (based on Kilosort drift plots) were discarded. To check for drift between head-fixed and freely moving recordings, we compared the mean waveforms and noise level for each unit across the two conditions, based on a 2 ms window around the identified spike times in bandpass-filtered data (800-8000Hz). An example mean waveform, with its standard deviation across individual spike times, is shown in Figure S1A. To determine whether the waveform changed, indicative of drift, we calculated coefficient of determination (R^2) between the two mean waveforms for each unit, which confirms a high degree of stability as the waveforms are nearly identical across conditions (Figure S1B). To determine whether the noise level changed, we computed the standard deviation across spike occurrences within each condition, for each unit (Figure S3.1C). There was no change in this metric between head-fixed and freely moving, indicating that there was not a change in noise level that might disrupt spike sorting in one condition specifically.

World and eye camera data were first deinterlaced to achieve 60 fps video. The world camera frames were then undistorted using a checkerboard calibration procedure (Python OpenCV), and downsampled to 30 by 40 pixels to reduce dimensionality and approximate mouse visual acuity. In order to extract pupil position from the eye video, eight points around the pupil were tracked with DeepLabCut (Mathis et al. 2018). We then fit these eight points to an ellipse and computed pupil position in terms of angular rotation (Michaiel et al., 2020). Sensor fusion analysis was performed on the IMU data (Jonny Saunders, University of Oregon) to calculate pitch and roll of the head. Pitch and roll were then passed through a median filter with window size 550 ms. All data streams were aligned to 50 ms bins through interpolation using system timestamps acquired in Bonsai.

3.8.4.4 GLM Training

For all model fits, the data were partitioned into 10% groups, and were randomly sampled into cross-validation train and test sets (70%/30% split, respectively). Video frames were cropped by 5 pixels on each side to remove edge artifacts. Initially, a shifter network was trained on each recording session (see below) to estimate the appropriate horizontal shift, vertical shift, and rotation of the world camera video to correct for eye movements. The corrected eye camera data were then saved out and used for training. Eye and head position were z-scored and zero-centered before training and analysis. Four different networks were trained: 1) Eye position and head orientation signals only, 2) Visual input only, 3) Additive interaction between position and visual input, and 4) Multiplicative interaction between position and visual input. Units with a mean firing rate below 1 Hz in either head-fixed or freely moving were removed from the data set (17% of total units).

3.8.4.5 Network parameters

To train the model end-to-end and to speed up the computation we utilized the graphical processing unit (GPU) and pyTorch because the GLM is equivalent to a single-layer linear network. We then used a rectified linear activation function to approximate non-zero firing rates. Utilizing the GPU decreased training time for the model by multiple orders of magnitude (from over 500 hours down to 40 minutes for the entire dataset). L1 and L2 regularization was applied to the spatiotemporal filters of the visual model. The Adam optimization algorithm (Kingma and Ba, 2014) was used to update the parameters of the model to minimize prediction error. The loss and gradient of each neuron were computed independently in parallel so the full model represents the entire dataset. To account for the convergence of different parameters at different speeds as well as to isolate parameters for regularization, parameter groups were established within the optimizer with independent hyperparameters.

3.8.4.6 Shifter Network

In order to correct the world camera video for eye movements, we trained a shifter network to convert eye position and torsion into an affine transformation of the image at each time point. For each recording session, eye angle and head pitch (theta, phi, and rho) were used as input into a feedforward network with a hidden layer of size 50, and output representing horizontal shift, vertical shift, and image rotation. The output of the network was then used to perform a differentiable affine transformation (Riba et al., 2019) to correct for eye movements. Head pitch was used as a proxy of eye torsion (Wallace et al., 2013), and eye position was zero-centered based on the mean position during the freely moving condition. The transformed image was then used as input into the GLM network to predict the neural activity. The shifter network and GLM were then trained together to minimize the error in predicted neural activity. During the shifter training (2000 epochs) no L1 regularization was applied to ensure a converged fit. Horizontal and vertical shift was capped at 20 pixels and rotation was capped at 45 deg. The eye corrected videos were saved out to be used for the model comparison training. The shifter network was trained on freely moving data, since eye movements are greatly reduced during head-fixation, but was applied to both head-fixed and freely moving data to align receptive fields across the two conditions.

3.8.4.7 Tuning and gain curves

Tuning curves for eye and head position were generated by binning the firing rates into quartiles so the density of each point is equal and then taking the average. For each gain curve we collected the time points of the firing rates that were within each quartile range for eye and head position, averaged the firing rates and then compared them with the predicted firing rates from the visual-only model. Each curve therefore represents how much each unit's actual firing rate changed on average when the mouse's eye or head was in the corresponding position.

3.8.4.8 Position-only model fits

Eye and head position signals were used as input into a single-layer network where the input dimension was four and the output dimension was the number of neurons. No regularization was applied during training due to the small number of parameters needed for the fitting. The learning rate for the weights and biases was 1e-3.

3.8.4.9 Visual-only model fits

Eye corrected world camera videos were used as input into the GLM network. The weights from the shifter training for each neuron were used as the initialization condition for the weights, while the mean firing rates of the neurons were used as the initialization for the biases. Parameters for the model were fit over 10,000 epochs with a learning rate of 1e-3. To prevent overfitting, a regularization sweep of 20 values log-base 10 distributed between 0.001 to 100 was performed. The model weights with the lowest test error were selected for each neuron.

3.8.4.10 Joint visual-position model fits

After the visual-only fits, the spatiotemporal weights and biases were frozen. A position module was then added to the model for which the input was the eye and head position signals (see Figure 4G). The output of the visual module was then combined with output of the position module in either an additive or multiplicative manner, then sent through a ReLu nonlinearity to approximate firing rates. The parameters for the position module were then updated with the Adam optimizer with learning rate 1e-3.

3.8.4.11 Speed and pupil diameter fits

To test the contribution of the speed and pupil diameter, the data were first z-scored and GLM fits were conducted with only speed and pupil, with eye/head position only and with speed, pupil and eye/head position. All models were fit with cross-validation with the same train/test split parameters as above. The explained variance (r^2) of the predicted and actual firing rate was calculated between these models to show how these parameters contribute uniquely and sublinearly to the GLM fits. Additionally, we trained the joint fits with eye/head position and speed and pupil and calculated the total contribution of eye/head position versus speed and pupil (Figure 4L-N).

3.8.4.12 Post-training analysis

To better assess the quality of fits, the actual and output firing rates were smoothed with a boxcar filter with a 2 s window. The correlation coefficient (cc) was then calculated between smoothed actual and predicted firing rates of the test dataset. The modulation index of neural activity by position was calculated as the (max-min)/(max+min) of each signal. In order to distinguish between additive and multiplicative models (Figure 4J,K), a unit needs to have a good positional and visual fit. As a result, units which had an cc value below 0.22, or did not improve with incorporating position information were thresholded out for the final comparison.

3.8.4.13 Simulated RF reconstruction

We tested the ability of our GLM approach to recover accurate receptive fields using simulated data. Simulated RFs were created based on Gabor functions and applied to the eye movement-corrected world camera video as a linear filter to generate simulated neural activity, scaled to empirically match the firing rates of real neurons with an average firing rate of 14 Hz. The output was then passed through a Poisson process to generate binned spike counts. Using these simulated data, we then followed the same analysis as for real data to fit a visual GLM model and estimate RFs, using spatiotemporal weights set to zero for the initial conditions.

3.8.4.14 Test-retest analysis receptive fields

To assess how reliable the receptive fields were, we trained the GLM separately on the first and second half of each recording session. We then took the receptive fields that were mapped for each half and calculated the pixel-wise correlation coefficient (Figure S3). A threshold of 0.5 cc was then used as a metric for stable RFs within the same condition. The units that had a stable RF in both head-fixed and freely moving conditions were then used for the analysis in Figure 3.

3.8.4.15 Shifter controls and change in visual scene

Similar to the test-retest for receptive fields, we trained the shifter network on the first and second half of the data. Shifter matrices were created using a grid of eye and head angles after training to see how the network responds to different angles. The coefficient of determination (R^2) was then calculated between the shifter matrices of the first and second half (Figure S2A-C). To further quantify the effect of the shifter network we used frame to frame image registration to measure the visual stability of the world camera video. Displacement between consecutive images was based on image registration performed with findTransformECC function in OpenCV.

We computed the cumulative sum of shifts to get total displacement, then calculated standard deviation in the fixation intervals following analysis in (Michaiel, Abe, and Niell 2020).

3.8.4.16 Dark experiments and analysis

To eliminate all possible light within the arena, the entire behavioral enclosure was sealed in light-blocking material (Thorlabs BK5), all potential light sources within the enclosure were removed, and the room lights 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. To counteract this, prior to the experiment, one drop of 2% Pilocarpine HCl Ophthalmic Solution was applied to the animal's right eye to restrict 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. Tuning curves for eye and head position were generated using the same method as in the light by binning the firing rates into quartiles so the density of each point is equal and then taking the average.

3.8.5 Quantification and statistical analysis

For shuffle distributions, we randomly shuffled spike times within the cross-validated train and test sets and then performed the same GLM training procedure. We defined significant values as two standard deviations away from the mean of the shuffle distribution. For paired t-tests, we first averaged across units within a session, then performed the test across sessions.

3.8.6 Additional Resources

3.8.6.1 Figures

Some figure panels were generated using Biorender.com.

3.8.7 Supplemental Figures and Videos

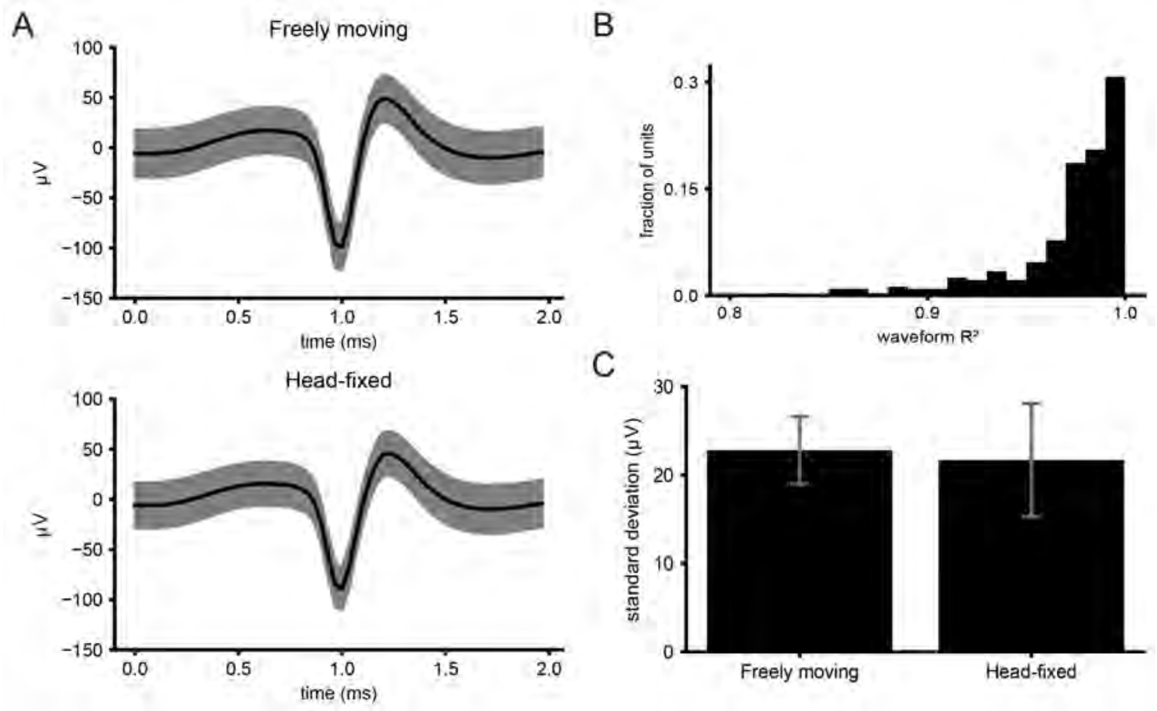


Figure 3.5. **A)** Top: Average spike waveform for one example unit in freely moving recording. Shaded region is one standard deviation. Bottom: Same unit as top but for head-fixed recording of the same unit in the same session. **B)** Histogram of coefficient of determination (R^2) between units of freely moving and head-fixed recordings. **C)** Average standard deviation across 2 ms around spikes for freely moving (FM) and head-fixed (HF) recordings.

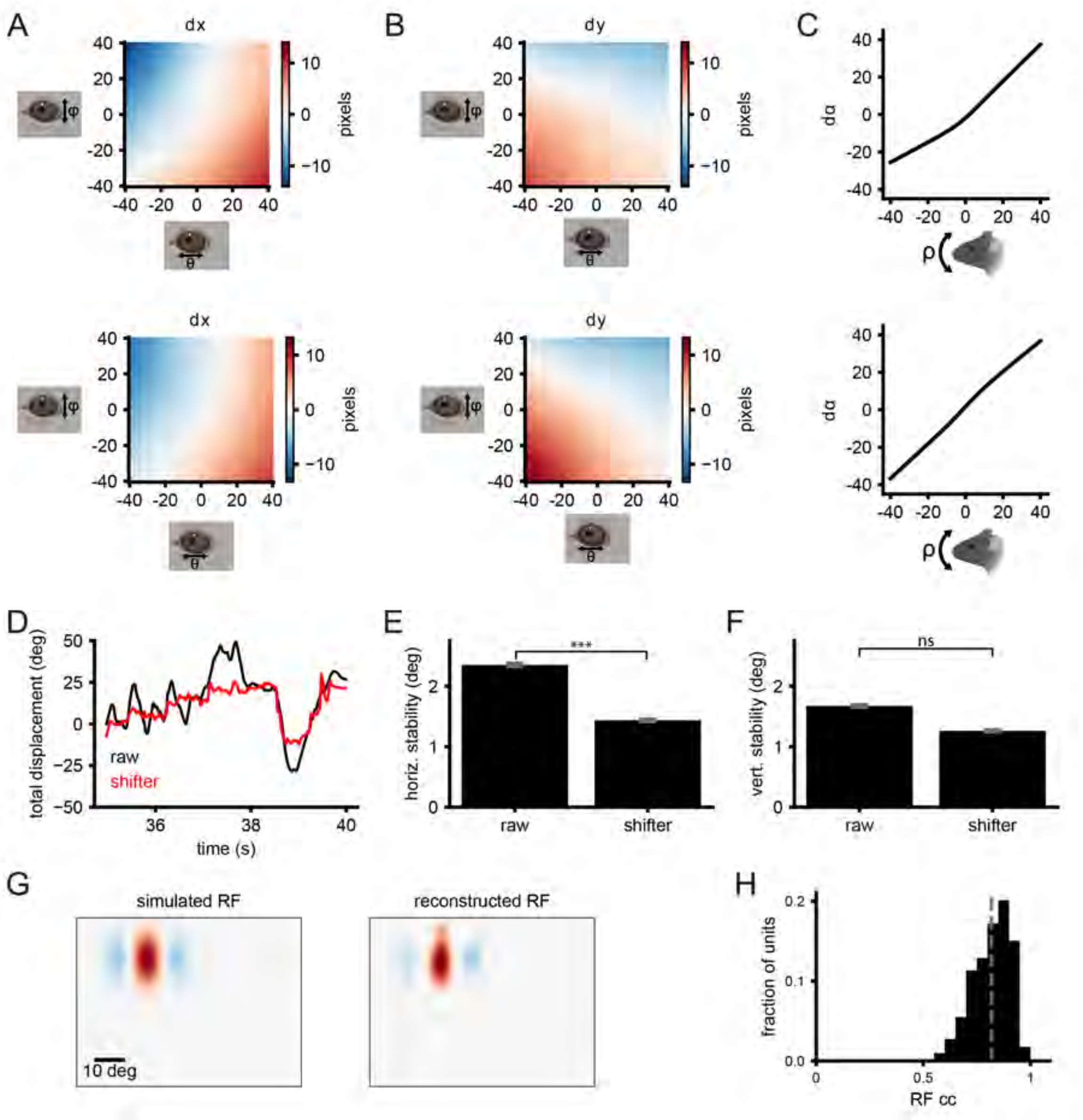


Figure 3.6. **A)** 2-d heat map of horizontal shift for values of theta and phi for first half (top) and second half (bottom) of example recording. **B)** Same as A but for the vertical shift of the image. **C)** Rotation of the image as a function of head pitch for the first half (top) and second half (bottom) of the recording. **D)** Image registration horizontal displacements for shifted and raw world camera video. **E)** Bar plot showing the average horizontal stability of visual angle for compensatory eye movements. **F)** Same as E but for vertical shifts. (***: p-value$<$0.0013) **G)** Simulated (left) and reconstructed (right) receptive fields with three sub-regions. Same training procedure as Figure 2A. **H)** Histogram of correlation coefficients between simulated and reconstructed RFs. The gray dashed line represents the mean of the distribution.

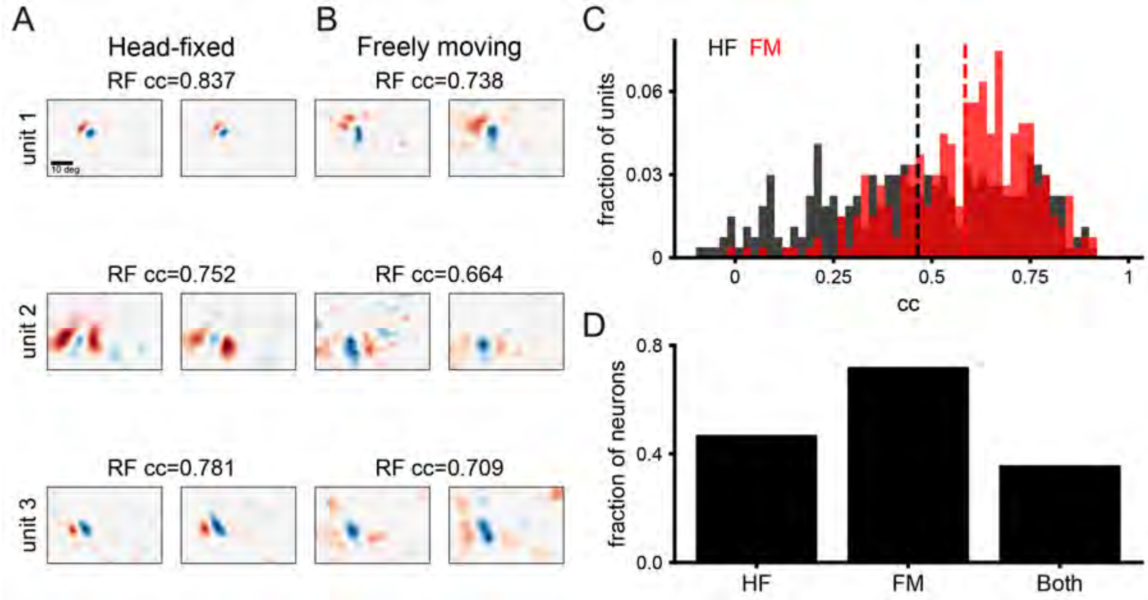


Figure 3.7. **A)** Three example receptive fields mapped in the first (left) and second (right) half of a head fixed recording. Correlation coefficient (cc) given is the pixel-wise cc of the receptive fields. **B)** Same as A but for freely moving recording. **C)** Histogram of cc of receptive fields for first versus second half of recording for head-fixed (gray) and freely moving (red) conditions. Dashed lines indicate the mean of the distribution. **D)** Bar plot showing the fraction of units that have a significant cc between the first and second half of the recordings (cc>0.5).

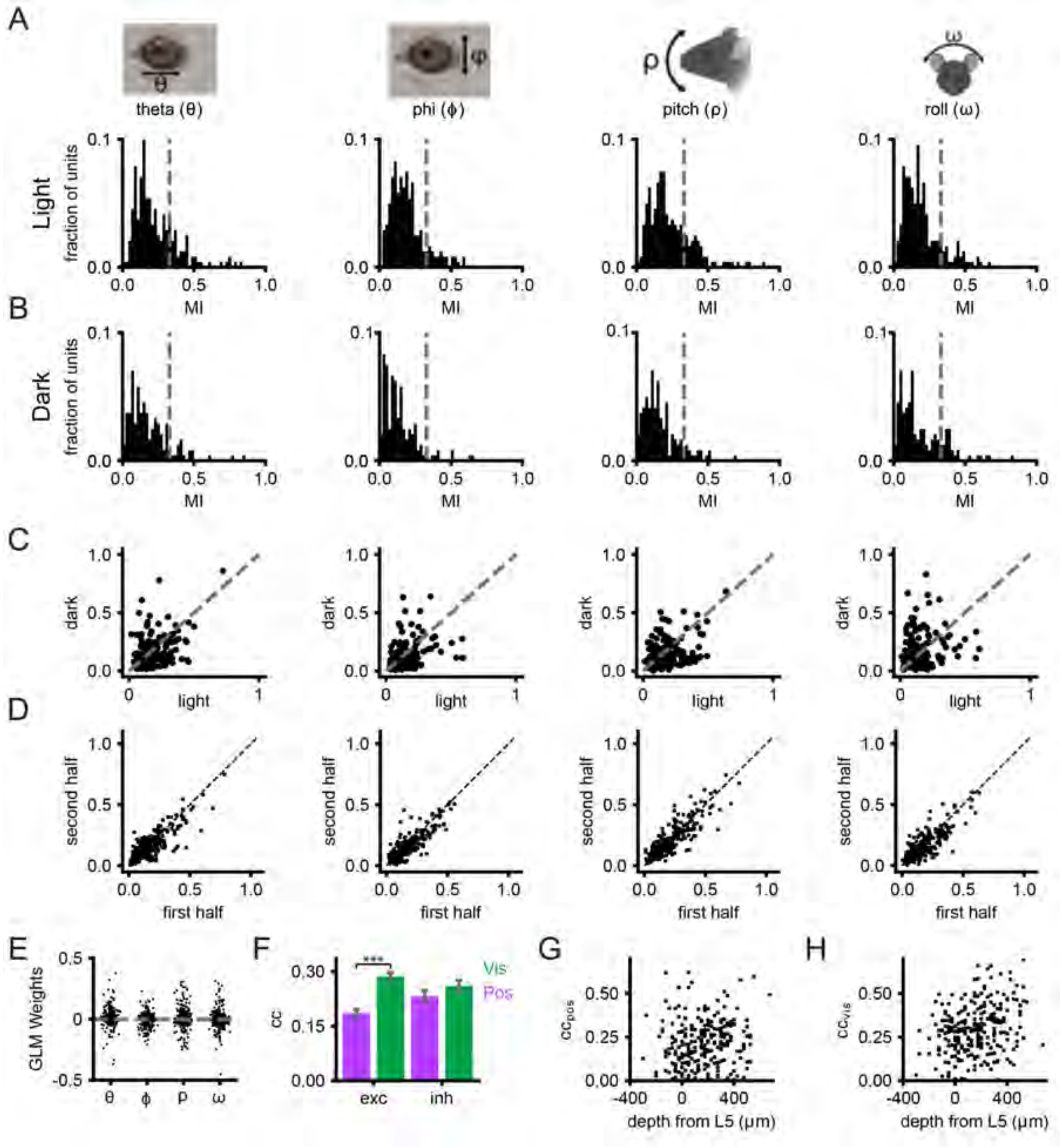


Figure 3.8. **A-D)** Columns correspond to analyses for theta, phi, pitch and roll respectively. **A)** Histograms of modulation index for single units recorded during free movement in the light. **B)** Same as A but recorded during free movement in darkness. C) Scatter plot comparison of light and dark modulation index for each unit. **D)** Modulation index calculated for first half and second half of the freely moving experiments in the light. **E)** Distribution of weights for position only GLM fit for eye/head position. **F)** Correlation coefficient (cc) of predicted versus actual firing rate for visual and position fits split by putative excitatory and inhibitory units. Error bars indicate standard error (***: p-value<0.001, between excitatory visual and position fits). **G)** Correlation coefficient as a function of depth from layer 5 for position fits (>0 deeper, <0 shallower). **H)** Same as G but for visual fits.

Supplemental Video 1: Sample experimental data from a fifteen second period during free movement. Relates to Figure 1.

Supplemental Video 2: Example of raw (left) and shifter network-corrected (right) world camera video from a single experiment. Relates to Figure 2.

3.9 BRIDGE TO CHAPTER IV

In this chapter, we investigated how V1 encodes information in the primary visual cortex. We mapped the first freely moving visual receptive fields using a novel experimental paradigm that capture the visual scene, neural activity, and the eye/head position. Our analysis and model show single neurons mainly encode visual information, but can be modulated by eye and head position and the integration of these signals mostly occurs via nonlinear gain modulation. In Chapter IV, I investigated how the mechanisms of higher-order visual representations can develop without explicit training in a deep neural network. By tasking a network to predict future visual information, representations of distance were shown to naturally form by transforming the two-dimensional inputs to form a three-dimensional representation of the environment. This theoretical modeling creates the foundation to bridge new interactions between experimental and theoretical research in natural visual behavior.

CHAPTER IV

EMERGENCE OF DEPTH REPRESENTATIONS IN PREDICTIVE NEURAL NETWORKS

4.1 AUTHOR CONTRIBUTIONS

The following chapter highlights research done with Yashar Ahmadian during his tenure at the University of Oregon and before he accepted a new position at Cambridge University. Unpublished material with Philip Parker, Yashar Ahmadian, and Cristopher Niell. E.T.T.A., Y.A., and C.M.N. contributed to the conception of this study. P.R.L.P. helped with the discussion about biological plausibility; E.T.T.A. designed, created, and implemented training of the model.

4.2 INTRODUCTION

To enable complex behaviors, the brain must extract useful representations of the environment from sensory inputs. For example, higher stages of the visual system encode variables relevant for guiding behavior that are not explicitly available in sensory input. Traditionally, such “actionable latent variables” were thought to be encoded strictly through visual processing, with neural representations becoming progressively more abstract as information ascends a hierarchy of cortical areas (Felleman and Van Essen, 1991; Hubel and Wiesel, 1959). However, during natural behavior, self-motion is strategically used to obtain new sensory input. In this process motor and positional information is sent to visual cortex where they are combined with visual input. Recent findings show how motor activity strongly and intricately modulates visual cortical responses (Guitchounts et al., 2020b; Musall et al., 2019; Niell and Stryker, 2010; Parker et al., 2022a; Stringer et al., 2019). However, understanding the function of these modulations remains a key area for future research in systems neuroscience.

In this study, we leveraged theoretical modeling inspired by an ethologically relevant distance estimation task in mice, where mice utilize self-motion to estimate the distance to objects in the environment (Parker et al., 2022b). Almost all animals with a visual system rely on motion parallax as a depth cue (de la Malla et al., 2016; Ferris, 1972; Kral, 2003), where points in the visual field closer to the observer move more than points at a further distance. Other depth cues such as binocular disparity can also be used for distance estimation, however, the focus of this study is on self-motion with motion parallax. Classical studies have shown rodents require their visual cortex to use motion parallax by performing characteristic head 'bobbing' to judge and jump the distance to a platform (Carey et al., 1990; Legg and Lambert, 1990). However, there has been a lack of theoretical modeling on the potential role of cortical visual-motor integration in depth estimation from motion parallax.

To model a visual scene in a controlled experiment, we built a simulated environment, initially using DeepMind Lab (Beattie et al., 2016), then with Unity to simulate a mouse locomoting around an environment. Within these simulations, a camera agent records the first-person point-of-view of the visual scene while moving in an arena with obstacles. The visual information is then used as input into a convolutional recurrent neural network (RNN) trained to predict future visual input (Lotter et al., 2018; Straka et al., 2020). Post-training, we were able to linearly decode the distance information from the neural activity. Interestingly, deeper layers where the representation was more abstract showed a stronger correlation with the ground-truth depth information. Although this research was initially inspired by the integration of visual-motor information, the bulk of the following work will focus on building complex representations with predictive processing, with the integration of motor information left for future study.

4.3 RESULTS

4.3.1 Simulations of freely moving mice in a virtual environment

Within the virtual environment (square arenas with diverse visual features) first in Deepmind Lab (Beattie et al., 2016) and then in Unity, the camera recorded first-person visual scene as the agent explored the arena by randomly sampling a location within the arena and then navigated to that location. For simplicity, the agent was constrained to only move forward while smoothly rotating in the horizontal plane (Figure 1). From these simulations, three data sequences were obtained: the 2D visual scenes captured by the camera agent, the corresponding 2D depth maps, and the motor commands generating the agent's trajectory. The visual scene was then used as input to the RNN, while the ground-truth depth maps were only used post-training to assess depth representations in the trained network.

4.3.2 A predictive neural network naturally creates a representation of distance

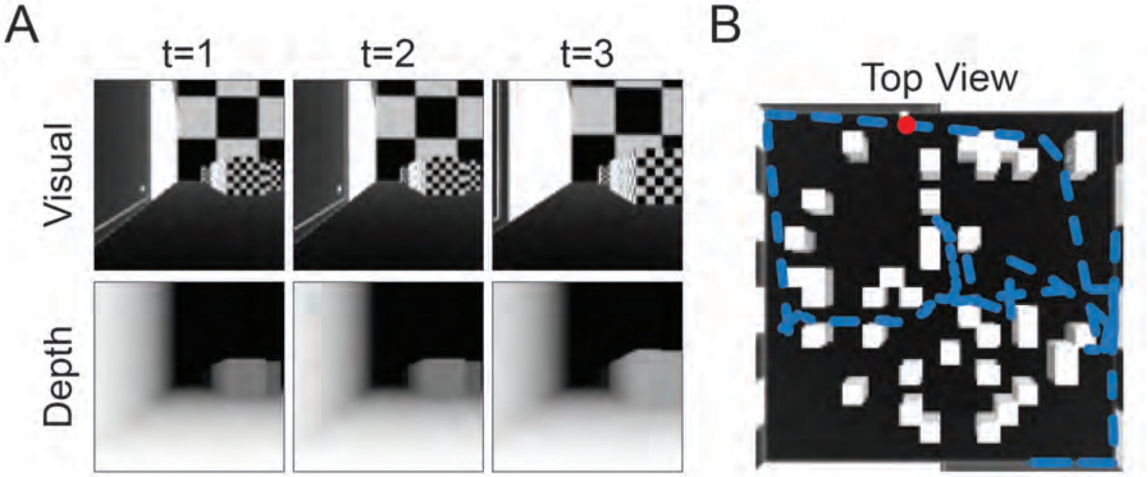


Figure 4.1. **A)** Example sequence of frames of the first-person visual scene (top) and depth maps (bottom) recorded by the camera agent. **B)** Top view of the virtual arena overlayed with the trajectory of movement taken by the camera agent. The current position shown in A is indicated by the red dot.

In recent years, predictive processing has been proposed as a useful framework for understanding the dynamic interactions between the motor and sensory systems (Keller Mrsic-Flogel, 2018). Inspired by this approach, we utilized a deep learning network to build a normative model of primary visual cortex (V1), based on the assumption of predictive processing. Predicting how projections of environmental objects on the retina move due to self-motion requires knowledge of an object's distance. Thus, tasking downstream networks simply to predict future visual inputs, without tasking them explicitly to estimate distance, can potentially shape those networks so they encode an explicit representation of depth (in addition to other latent variables).

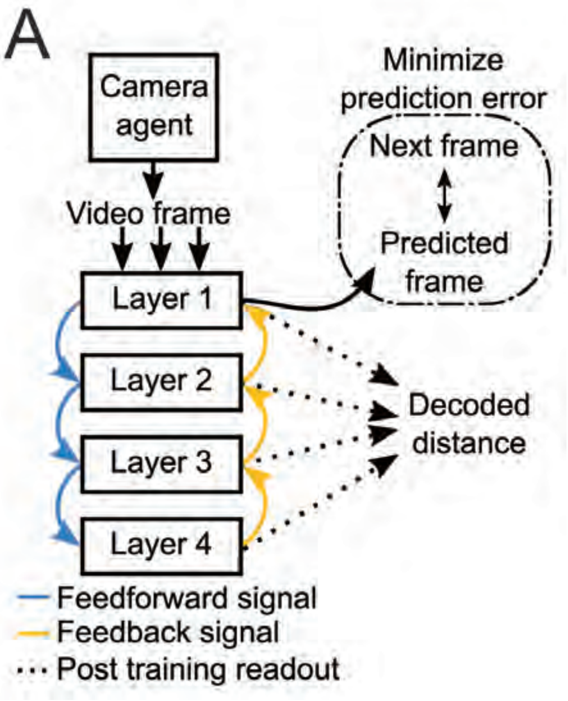


Figure 4.2. **A)** The first-person camera agent records video frames which then become input to the predictive network. Each layer consists of convolutional recurrent units and receives feedforward and feedback signals at every timestep. The network is trained with backpropagation through time by minimizing pixel-wise prediction and the actual next frame. Post-training the activations of units are collected and used to decode the distance.

We built a multi-layer convolutional RNN as a normative model of visual processing. Unsupervised predictive learning (UPL) was used to train the RNN, i.e., we optimized the RNN's connection weights to minimize its error in predicting future visual inputs (Lotter et al., 2018; Straka et al., 2020). The full network consists of stacked recurrent convolutional layers, with feedforward connections sending prediction errors to higher layers and feedback connections relaying top-down modulation (Figure 2). Each recurrent layer provides an output prediction for the inputs from the previous layer, which is then used to determine the accuracy of the prediction during training. The loss function is defined as the mean absolute error between the predicted input and the actual input, minimized by stochastic gradient descent. At the pixel layer, this loss measures the mismatch between the next incoming video frame as predicted by the first layer of the RNN.

First, we visually verified the network is predicting by time aligning the predicted video frames and the real video frames (Figure 3A). Intuitively, if the network can accurately predict the dynamics of the visual scene such as objects moving relative to each other, then there must be a representation of distances. To quantify the accuracy of the predictions, the pixel-wise mean square error (MSE) between the predicted and actual video frame separated by time Δt was calculated. The minimum error occurred at Δt=1, signifying the network does not simply use the previous frame as a prediction for the next frame (Figure 3B).

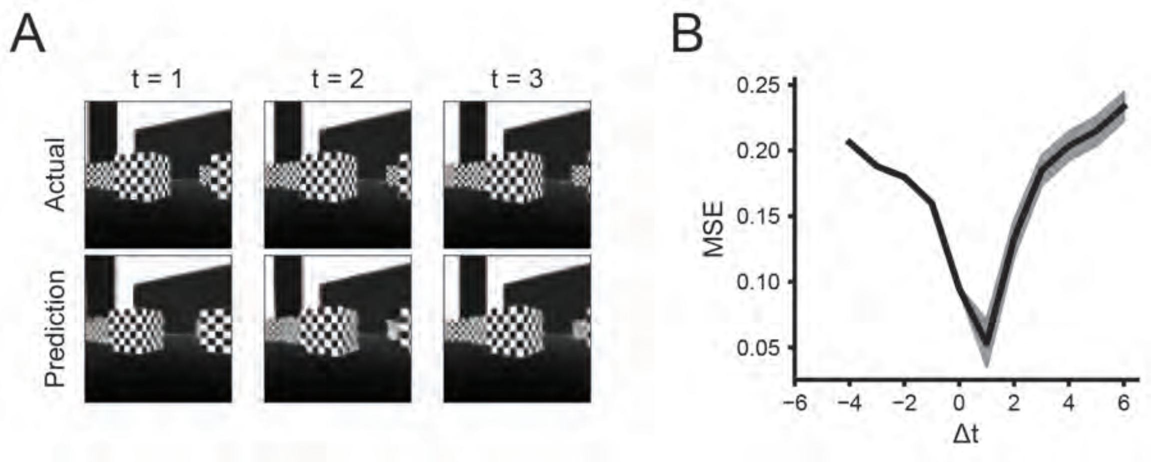


Figure 4.3. **A)** Three example frames of the actual (top) and predicted (bottom) visual scene time aligned. Predictions visually match the actual incoming frame. **B)** Plot of the mean square error (MSE) between each frame and the prediction separated by Δt. The minimum occurs at Δt=1 meaning the network is making a prediction into the future and not just using the previous frame as a prediction.

To assess how readily and explicitly depth is represented in different layers of the trained RNN, the distance was decoded from each layer's neural activations using linear readouts (which are decoders with minimal complexity). Readouts were trained (post RNN training), by cross-validated convolutional ridge regression given ground-truth depth maps and RNN activations on a test dataset. Each layer of neurons was used to reconstruct the corresponding part of visual space in a pixel-wise manner. Reconstructions of the depth maps can be seen as increasing in accuracy with higher layers, up until layer 3 (Figure 4A). When taking the center pixel reconstruction across time and comparing this with the actual distance, the model shows a strong correspondence with the ground-truth distance. The r^2 (coefficient of determination) of this regression was used to assess the accuracy of depth representation. Our

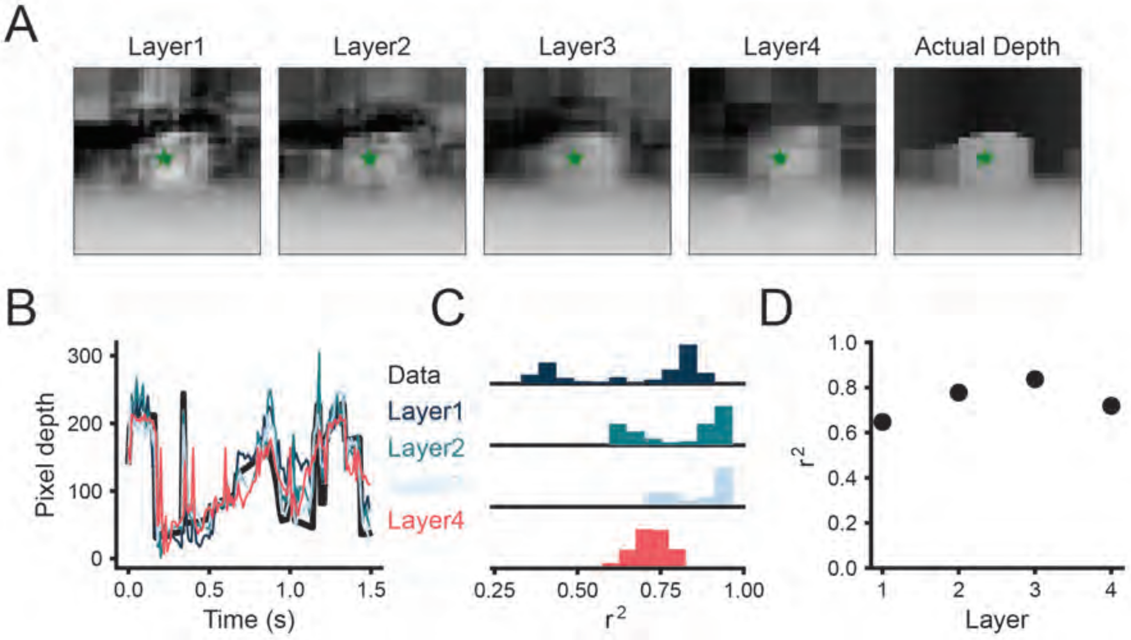


Figure 4.4. **A)** Comparison of depth maps reconstructed from layer activations with actual depth frame. **B)** Trace of center pixel (green star) across time for each layer. **C)** Histograms of the r^2 values between each layer's pixel reconstruction and the ground truth distance. Y-axis depict count of pixel r^2 values. **D)** Scatter plot of the average r^2 values for each layer.

analysis shows the RNN does indeed form depth representations, without being explicitly tasked to do so (Figure 4). Finally, using the linear decoder, we find that the representation of distance strengthens in deeper layers (Figure 4D). The non-monotonic increase in the representation of distance is likely due to the max pooling operation between the layers of the predictive network. The representation of visual space is compressed in higher layers, and this combined with the convolutional design of the linear decoder results in low resolution reconstructions with the same neurons reconstructing larger parts of the pixel space.

In summary, we developed a virtual environment for simulating freely moving exploration with a first-person point-of-view camera and a multi-layer RNN trained by UPL which can successfully extract explicit representations of higher-level dynamical variables. These variables are only implicitly represented in the visual input, supporting the hypothesis predictive processing may be an effective and efficient method to build representations of the environment during natural behavior. Unfortunately, the incorporation of motor information is left to future work.

4.4 DISCUSSION

Historically, from anesthesia to head fixation, visual neuroscience has been constrained to require precise control over the movements and visual input a animal is viewing. Recent work in the Niell lab has made strides in studying visual processing under more natural movement conditions (Parker et al., 2022b,a, 2020). However, these studies are still limited to low-level sensory processing. In this work, we developed a virtual environment as a reconstruction of freely moving mice in a naturalistic arena as a way to bridge experiment and theoretical work to investigate how higher-order visual representations are formed. Additionally, we show how a network trained to predict future visual input without explicit training on higher-order representations can naturally develop these higher-order representations, such as with distance.

4.4.1 Using Predictions to Understand the World

Predictive coding has been a well-established normative model for visual processing of low level representations (Jiang and Rao, 2022; Rao et al., 2022; Rao and Ballard, 1999; Rao and Jiang, 2022). However, large-scale testing of higher-order representations, features of the visual scene that are not explicit in the low level details, has not been studied. For example, through either development or experience, animals intuitively build a three-dimensional sense of how to engage with the environment from two-dimensional projections of light on the retina. What mechanisms might be employed to build this intuition? Previous studies connecting machine learning with visual neuroscience have relied on existing datasets of static images in a head-fixed paradigm (Yamins et al., 2014). Other research has shown predictive learning as a powerful mechanism to develop place fields given the correct information (Recanatesi et al., 2021). Additionally, deep learning networks trained to explicitly reconstruct depth maps with either supervised or unsupervised learning (Kuznietsov et al., 2017; Masoumian et al., 2022; Wang et al., 2018; Zhao et al., 2020) have had remarkable success in specific conditions. Interestingly, there is a significant increase in the accuracy of the depth maps when there is a separate component to estimate changes in the pose position of the camera. This is reminiscent of how motor commands can be combined with the visual input, although these networks were all tasked to explicitly represent depth. In our work, we show how simply making predictions about future incoming visual input is sufficient to generate a representation of depth. More generally, active sensing by minimizing the error between predictions about future sensory input may be an efficient mechanism to learn a world model. (Rao et al., 2022).

4.4.2 Future Work

Up until this point, the processing of the predictive network has relied solely on visual input. The next step would be to incorporate the motor commands recorded by the camera agent by performing a systematic search across layers to see where motor commands would be most effective. Intuitively, incorporating motor commands at the lowest layer would give the network computational power to transform information into a coordinate system that can easily be combined with visual representations. With the integration of visual and motor representations, the network would have everything needed to predict the changing scene due to self-motion and thus, be able to generate more accurate predictions. Additionally, this would also yield a more accurate depth representation that is linearly decodable. A final control for the depth representations would also be to train an autoencoder to reconstruct information at the current time step. Then after training test if the latent representation of the autoencoder contains information about the depth map.

An additional expansion of this model would be to extend the temporal prediction beyond one timestep. Currently, the hierarchical nature of the existing network results in each layer's activations being explicitly dependent on the previous timestep during the forward pass of the model. To address this, instead of predicting the pixel-level representations, an alternative approach would be to predict the future latent representations (Han et al., 2019; van den Oord et al., 2018) whereby an encoder would compress the representations for an RNN to make predictions on. With this architecture, various learning mechanisms from contrastive predictive coding (van den Oord et al., 2018) to dynamic predictive coding (Jiang and Rao, 2022) can be implemented.

An exciting future direction would be to combine the experimental work from the Niell lab with theoretical modeling. Recent developments in deep learning have enabled the construction of 3D models by simply using a camera application called PolyCam. With this app, a 3D model of the arena can be created and directly imported into Unity. During experiments, the location of the mouse's head and body are recorded and can be placed in the 3D model. Then, Unity ground-truth depth maps can be calculated and correlated with neural activity as the mouse is freely exploring. Additional latent variables could also easily be extracted from the 3D virtual model such as optic flow, foreground/background, and object identity. Taken together, this system would represent a rich framework for investigating visual processing in freely moving mice.

4.5 MATERIALS AND METHODS

4.5.1 Unity Simulations

The virtual environment was a square arena with classical visual stimuli on the walls, such as orientated gratings. On initialization, cubes with a black and white grid pattern were randomly placed in the arena. A random location within the arena was selected and the NavMeshAgent was used to navigate to the selected location. The agent recorded the visual scene, depth maps, as well as the position and velocity. Trials lasted 100 seconds and were recorded at 100 frames per second. A total of 300 trials with different random seeds were used in the training set, while 100 different trials were used in the test set.

4.5.2 Network Parameters

The predictive network consisted of four convolutional gated recurrent unit (GRU) layers of size (32 channels, 64 x 64 pixels), (64,32,32), (128,16,16), and (256, 8, 8) respectively. Max pooling and upsampling occurred between layers for feedforward and feedback connections. The network was trained using the mean absolute error between the image frame prediction and the actual next frame. Training consisted of 200 epochs (full run-throughs of the training dataset) with a batch size of 32. Weights were optimized using the Adam optimizer, and a learning rate scheduler with an initial learning rate of 0.001 and decreased the learning rate by a factor of 2 every 50 epochs.

4.5.3 Linear Decoding

For linear decoding, the activations of each layer were recorded in response to the test dataset and z scored. For each layer, a convolutional window of size [15, 8, 4, 2] was used to choose which neurons in each layer contributed to predicting the pixel depth. Every two pixels in the depth map were predicted using independent ridge regression models with the corresponding neurons and alpha=0.1. This decoding was cross-validated with a train/test split of 0.75/0.25 with the test dataset. Models were trained using the high-performance computing cluster Talapas at the University of Oregon.

BIBLIOGRAPHY

R A Andersen and V B Mountcastle. The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *Journal of Neuroscience*, 3(3):532–548, March 1983.

Aslı Ayaz, Aman B Saleem, Marieke L Schölvinck, and Matteo Carandini. Locomotion controls spatial integration in mouse visual cortex. *Current Biology*, 23(10):890–894, May 2013.

Vernon Bailey and Charles Sperry. Life History and Habits of Grasshopper Mice, Genus Onychomys. Technical Report 1488-2016-123948, 1929.

Pouya Bashivan, Kohitij Kar, and James J DiCarlo. Neural population control via deep image synthesis. *Science (New York, N.Y.)*, 364(6439), May 2019.

Charles Beattie, Joel Z Leibo, Denis Teplyashin, Tom Ward, Marcus Wainwright, Heinrich Küttler, Andrew Lefrancq, Simon Green, Víctor Valdés, Amir Sadik, Julian Schrittwieser, Keith Anderson, Sarah York, Max Cant, Adam Cain, Adrian Bolton, Stephen Gaffney, Helen King, Demis Hassabis, Shane Legg, and Stig Petersen. DeepMind Lab. *arXiv*, pages 1–11, 2016.

Isaac H Bianco, Adam R Kampff, and Florian Engert. Prey capture behavior evoked by simple visual stimuli in larval zebrafish. *Frontiers in systems neuroscience*, 5: 101, 2011. doi: 10.3389/fnsys.2011.00101.

Andrew Blake and Alan L. Yuille, editors. *Active Vision.* The MIT Press, Cambridge, Mass., November 1992. ISBN 978-0-262-02351-1.

Adam Bleckert, Gregory W. Schwartz, Maxwell H. Turner, Fred Rieke, and Rachel O. L. Wong. Visual space is represented by nonmatching topographies of distinct mouse retinal ganglion cell types. *CURRENT BIOLOGY*, 24(3):310–315, February 2014. ISSN 0960-9822. doi: 10.1016/j.cub.2013.12.020.

Vincent Bonin, Mark H Histed, Sergey Yurgenson, and R Clay Reid. Local diversity and fine-scale organization of receptive fields in mouse visual cortex. *Journal of Neuroscience*, 31(50):18506–18521, December 2011.

Howard C Boone, Jason M Samonds, Emily C Crouse, Carrie Barr, Nicholas J Priebe, and Aaron W McGee. Natural binocular depth discrimination behavior in mice explained by visual cortical activity. *Current Biology*, 31(10):2191–2198.e3, May 2021.

Guy Bouvier, Yuta Senzai, and Massimo Scanziani. Head Movements Control the Activity of Primary Visual Cortex in a Luminance-Dependent Manner. *Neuron*, 108(3):500–511.e5, November 2020.

Peter R Brotchie, Richard A Andersen, Lawrence H Snyder, and Sabrina J Goodman. Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature*, 375(6528):232–235, May 1995.

Laura Busse, Jessica A Cardin, M Eugenia Chiappe, Michael M Halassa, Matthew J McGinley, Takayuki Yamashita, and Aman B Saleem. Sensation during Active Behaviors. *Journal of Neuroscience*, 37(45):10826–10834, November 2017.

Daniel A Butts. Data-Driven Approaches to Understanding Visual Neuron Activity. *Annual Review of Vision Science*, 5(1):451–477, 2019.

Matteo Carandini, Jonathan B Demb, Valerio Mante, David J Tolhurst, Yang Dan, Bruno A Olshausen, Jack L Gallant, and Nicole C Rust. Do we know what the early visual system does? *Journal of Neuroscience*, 25(46):10577–10597, November 2005.

David P Carey, Melvyn A Goodale, and Erin G Sprowl. Blindsight in rodents: The use of a 'high-level' distance cue in gerbils with lesions of primary visual cortex. *Behav. Brain Res.*, 38(3):283–289, 1990.

E J Chichilnisky. A simple white noise analysis of neuronal light responses. *Network (Bristol, England)*, 12(2):199–213, May 2001.

BJ Clark, DA Hamilton, and IQ Whishaw. Motor activity (exploration) and formation of home bases in mice (C57BL/6) influenced by visual and tactile cues: Modification of movement distribution, distance, location, and speed. *PHYSIOLOGY & BEHAVIOR*, 87(4):805–816, April 2006. ISSN 0031-9384. doi: 10.1016/j.physbeh.2006.01.026.

James A Crowell, Martin S Banks, Krishna V Shenoy, and Richard A Andersen. Visual self-motion perception during head turns. *Nature Neuroscience*, 1(8):732–737, December 1998.

Sandeep Robert Datta, David J Anderson, Kristin Branson, Pietro Perona, and Andrew Leifer. Computational Neuroethology: A Call to Action. *Neuron*, 104 (1):11–24, October 2019.

Cristina de la Malla, Stijn Buiteman, Wilmer Otters, Jeroen B J Smeets, and Eli Brenner. How various aspects of motion parallax influence distance judgments, even when we think we are standing still. *J. Vis.*, 16(9):1–14, 2016.

Saskia E J de Vries, Jerome A Lecoq, Michael A Buice, Peter A Groblewski, Gabriel K Ocker, Michael Oliver, David Feng, Nicholas Cain, Peter Ledochowitsch, Daniel Millman, Kate Roll, Marina Garrett, Tom Keenan, Leonard Kuan, Stefan Mihalas, Shawn Olsen, Carol Thompson, Wayne Wakeman, Jack Waters, Derric Williams, Chris Barber, Nathan Berbesque, Brandon Blanchard, Nicholas Bowles, Shiella D Caldejon, Linzy Casal, Andrew Cho, Sissy Cross, Chinh Dang, Tim Dolbeare, Melise Edwards, John Galbraith, Nathalie Gaudreault, Terri L Gilbert, Fiona Griffin, Perry Hargrave, Robert Howard, Lawrence Huang, Sean Jewell, Nika Keller, Ulf Knoblich, Josh D Larkin, Rachael Larsen, Chris Lau, Eric Lee, Felix Lee, Arielle Leon, Lu Li, Fuhui Long, Jennifer Luviano, Kyla Mace, Thuyanh Nguyen, Jed Perkins, Miranda Robertson, Sam Seid, Eric Shea-Brown, Jianghong Shi, Nathan Sjoquist, Cliff Slaughterbeck, David Sullivan, Ryan Valenza, Casey White, Ali Williford, Daniela M Witten, Jun Zhuang, Hongkui Zeng, Colin Farrell, Lydia Ng, Amy Bernard, John W Phillips, R Clay Reid, and Christof Koch. A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nature Neuroscience*, 23(1):138–151, January 2020.

Emily Jane Dennis, Ahmed El Hady, Angie Michaiel, Ann Clemens, Dougal R Gowan Tervo, Jakob Voigts, and Sandeep Robert Datta. Systems Neuroscience of Natural Behaviors in Rodents. *Journal of Neuroscience*, 41(5):911–919, February 2021.

Daniel A. Dombeck, Anton N. Khabbaz, Forrest Collman, Thomas L. Adelman, and David W. Tank. Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *NEURON*, 56(1):43–57, October 2007. ISSN 0896-6273. doi: 10.1016/j.neuron.2007.08.003.

U C Drager. Observations on monocular deprivation in mice. 41(1):28–42, 1978.

Jean-Baptiste Durand, Yves Trotter, and Simona Celebrini. Privileged processing of the straight-ahead direction in primate area V1. *Neuron*, 66(1):126–137, April 2010.

Daniel J Felleman and David C Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex*, 1(1):1–47, 1991.

R C Feord, M E Sumner, S Pusdekar, L Kalra, P T Gonzalez-Bellido, and Trevor J Wardill. Cuttlefish use stereopsis to strike at prey. *Science advances*, 6(2):eaay6036, January 2020.

Steven H Ferris. Motion parallax and absolute distance. *J. Exp. Psychol.*, 95(2): 258–263, 1972.

Emmanouil Froudarakis, Paul G Fahey, Jacob Reimer, Stelios M Smirnakis, Edward J Tehovnik, and Andreas S Tolias. The Visual Cortex in Context. *Annual review of vision science*, 5:317–339, September 2019.

James J Gibson. *The Ecological Approach To Visual Perception.* Lawrence Erlbaum Associates, 1979.

Grigori Guitchounts, William Lotter, Joel Dapello, and David Cox. Stable 3D Head Direction Signals in the Primary Visual Cortex. *bioRxiv : the preprint server for biology*, 2020a.

Grigori Guitchounts, Javier Masís, Steffen B E Wolff, and David Cox. Encoding of 3D Head Orienting Movements in the Primary Visual Cortex. *Neuron*, 108(3): 512–525.e4, November 2020b.

Tengda Han, Weidi Xie, and Andrew Zisserman. Video Representation Learning by Dense Predictive Coding. (arXiv:1909.04656), September 2019.

Wenfei Han, Luis A. Tellez, Miguel J. Rangel, Jr., Simone C. Motta, Xiaobing Zhang, Isaac O. Perez, Newton S. Canteras, Sara J. Shammah-Lagnado, Anthony N. van den Pol, and Ivan E. de Araujo. Integrated control of predatory hunting by the central nucleus of the amygdala. *CELL*, 168(1-2):311+, January 2017. ISSN 0092-8674. doi: 10.1016/j.cell.2016.12.027.

L Harkness and H C Bennet-Clark. The deep fovea as a focus indicator. *Nature*, 272 (5656):814–816, April 1978.

M Hayhoe and D Ballard. Eye movements in natural behavior. *TRENDS IN COGNITIVE SCIENCES*, 9(4):188–194, April 2005. ISSN 1364-6613. doi: 10.1016/j.tics.2005.02.009.

Emily Higgins and Keith Rayner. Transsaccadic processing: Stability, integration, and the potential role of remapping. *ATTENTION PERCEPTION & PSYCHOPHYSICS*, 77(1):3–27, January 2015. ISSN 1943-3921. doi: 10.3758/s13414-014-0751-y.

Jennifer L. Hoy, Iryna Yavorska, Michael Wehr, and Cristopher M. Niell. Vision drives accurate approach behavior during prey capture in laboratory mice. *CURRENT BIOLOGY*, 26(22):3046–3052, November 2016. ISSN 0960-9822. doi: 10.1016/j.cub.2016.09.009.

Jennifer L. Hoy, Hannah I. Bishop, and Cristopher M. Niell. Defined cell types in superior colliculus make distinct contributions to prey capture behavior in the mouse. *CURRENT BIOLOGY*, 29(23):4130+, December 2019. ISSN 0960-9822. doi: 10.1016/j.cub.2019.10.017.

D H Hubel and T N Wiesel. Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.*, 148:574–591, October 1959.

D H Hubel and T N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.*, 160:106–154, January 1962.

Linxing Preston Jiang and Rajesh P. N. Rao. Dynamic Predictive Coding: A New Model of Hierarchical Sequence Learning and Prediction in the Cortex. Preprint, Neuroscience, June 2022.

Ashley L Juavinett, Jeffrey C Erlich, and Anne K Churchland. Decision-making behaviors: Weighing ethology, complexity, and sensorimotor compatibility. *Current Opinion in Neurobiology*, 49:42–50, April 2018.

Ashley L Juavinett, George Bekheet, and Anne K Churchland. Chronically implanted Neuropixels probes enable high-yield recordings in freely moving mice. *Elife*, 8, August 2019.

Hadas Ketter Katz, Avichai Lustig, Tidhar Lev-Ari, Yuval Nov, Ehud Rivlin, and Gadi Katzir. Eye movements in chameleons are not truly independent - evidence from simultaneous monocular tracking of two targets. *Journal of Experimental Biology*, 218(Pt 13):2097–2105, July 2015.

Diederik P Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. December 2014.

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

Yevhen Kuznietsov, Jörg Stückler, and Bastian Leibe. Semi-supervised deep learning for monocular depth map prediction. *arXiv*, 2017-Janua:2215–2223, 2017.

R Lal and M J Friedlander. Effect of passive eye movement on retinogeniculate transmission in the cat. *Journal of Neurophysiology*, 63(3):523–538, March 1990.

M F Land. Motion and vision: Why animals move their eyes. *Journal of Comparative Physiology - A Sensory, Neural, and Behavioral Physiology*, 185(4):341–352, 1999.

Michael F Land. Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research*, 25(3):296–324, May 2006.

Michael F. Land. The evolution of gaze shifting eye movements. *PROCESSES OF VISUOSPATIAL ATTENTION AND WORKING MEMORY*, 41:3–11, 2019. ISSN 1866-3370. doi: 10.1007/7854_2018_60.

C.R. Legg and S. Lambert. Distance estimation in the hooded rat: Experimental evidence for the role of motion cues. *Behavioural Brain Research*, 41(1):11–20, 1990. ISSN 0166-4328. doi: 10.1016/0166-4328(90)90049-K.

David A Leopold and Soo Hyun Park. Studying the visual brain in its natural rhythm. *Neuroimage*, 216:116790, August 2020.

Goncalo Lopes, Niccolo Bonacchi, Joao Frazao, Joana P. Neto, Bassam V. Atallah, Sofia Soares, Luis Moreira, Sara Matias, Pavel M. Itskov, Patricia A. Correia, Roberto E. Medina, Lorenza Calcaterra, Elena Dreosti, Joseph J. Paton, and Adam R. Kampff. Bonsai: An event-based framework for processing and controlling data streams. *FRONTIERS IN NEUROINFORMATICS*, 9, April 2015. ISSN 1662-5196. doi: 10.3389/fninf.2015.00007.

William Lotter, Gabriel Kreiman, and David Cox. A neural network trained to predict future video frames mimics critical properties of biological neuronal responses and perception. *bioRxiv*, pages 1–18, 2018.

Liqun Luo, Edward M. Callaway, and Karel Svoboda. Genetic dissection of neural circuits. *NEURON*, 57(5):634–660, March 2008. ISSN 0896-6273. doi: 10.1016/j.neuron.2008.01.002.

Mauro Manassi, Bilge Sayim, and Michael H. Herzog. When crowding of crowding leads to uncrowding. *Journal of Vision*, 13(13):10, November 2013. ISSN 1534-7362. doi: 10.1167/13.13.10.

Graham R. Martin. What is binocular vision for? A birds' eye view. *JOURNAL OF VISION*, 9(11), 2009. ISSN 1534-7362. doi: 10.1167/9.11.14.

Armin Masoumian, Hatem A. Rashwan, Julián Cristiano, M. Salman Asif, and Domenec Puig. Monocular Depth Estimation Using Deep Learning: A Review. *Sensors*, 22(14):5353, July 2022. ISSN 1424-8220. doi: 10.3390/s22145353.

Alexander Mathis, Pranav Mamidanna, Kevin M. Cury, Taiga Abe, Venkatesh N. Murthy, Mackenzie Weygandt Mathis, and Matthias Bethge. DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *NATURE NEUROSCIENCE*, 21(9):1281+, September 2018. ISSN 1097-6256. doi: 10.1038/s41593-018-0209-y.

Matthew J McGinley, Martin Vinck, Jacob Reimer, Renata Batista-Brito, Edward Zagha, Cathryn R Cadwell, Andreas S Tolias, Jessica A Cardin, and David A McCormick. Waking State: Rapid Variations Modulate Neural and Behavioral Responses. *Neuron*, 87(6):1143–1161, September 2015.

Elisha P Merriam, Justin L Gardner, J Anthony Movshon, and David J Heeger. Modulation of Visual Responses by Gaze Direction in Human Visual Cortex. *Journal of Neuroscience*, 33(24):9879–9889, June 2013.

Arne F Meyer, Jasper Poort, John O'Keefe, Maneesh Sahani, and Jennifer F Linden. A Head-Mounted Camera System Integrates Detailed Behavioral Monitoring with Multichannel Electrophysiology in Freely Moving Mice. *Neuron*, 100(1):46–60.e7, October 2018.

Arne F. Meyer, John O'Keefe, and Jasper Poort. Two distinct types of eye-head coupling in freely moving mice. *CURRENT BIOLOGY*, 30(11):2116+, June 2020. ISSN 0960-9822. doi: 10.1016/j.cub.2020.04.042.

Angie M Michaiel, Elliott TT Abe, and Cristopher M Niell. Dynamics of gaze control during prey capture in freely moving mice. *eLife*, 9:e57458, July 2020. ISSN 2050-084X. doi: 10.7554/eLife.57458.

Cory T Miller, David Gire, Kim Hoke, Alexander C Huk, Darcy Kelley, David A Leopold, Matthew C Smear, Frederic Theunissen, Michael Yartsev, and Cristopher M Niell. Natural behavior is the language of the brain. *Current Biology*, 32(10):R482–R493, May 2022.

Bartul Mimica, Benjamin A Dunn, Tuce Tombaz, V P T N C Srikanth Bojja, and Jonathan R Whitlock. Efficient cortical coding of 3D posture in freely behaving rats. *Science (New York, N.Y.)*, 362(6414):584–589, November 2018.

Adam P Morris and Bart Krekelberg. A Stable Visual World in Primate Primary Visual Cortex. *Current Biology*, 29(9):1471–1480.e6, 2019.

RGM MORRIS. SPATIAL LOCALIZATION DOES NOT REQUIRE THE PRESENCE OF LOCAL CUES. *LEARNING AND MOTIVATION*, 12(2):239–260, 1981. ISSN 0023-9690. doi: 10.1016/0023-9690(81)90020-5.

Simon Musall, Matthew T Kaufman, Ashley L Juavinett, Steven Gluf, and Anne K Churchland. Single-trial neural dynamics are dominated by richly varied movements. *Nature Neuroscience*, 22(10):1677–1686, October 2019.

Cristopher M. Niell. Cell Types, Circuits, and Receptive Fields in the Mouse Visual Cortex. *Annual Review of Neuroscience*, 38(1):413–431, July 2015. ISSN 0147-006X, 1545-4126. doi: 10.1146/annurev-neuro-071714-033807.

Cristopher M. Niell and Massimo Scanziani. How Cortical Circuits Implement Cortical Computations: Mouse Visual Cortex as a Model. *Annual Review of Neuroscience*, 44(1):517–546, July 2021. ISSN 0147-006X, 1545-4126. doi: 10.1146/annurev-neuro-102320-085825.

Cristopher M Niell and Michael P Stryker. Highly selective receptive fields in mouse visual cortex. *Journal of Neuroscience*, 28(30):7520–7536, 2008.

Cristopher M. Niell and Michael P. Stryker. Modulation of visual responses by behavioral state in mouse visual cortex. *NEURON*, 65(4):472–479, February 2010. ISSN 0896-6273. doi: 10.1016/j.neuron.2010.01.033.

Ramon Nogueira, Chris C Rodgers, Randy M Bruno, and Stefano Fusi. The geometry of cortical representations of touch in rodents. September 2021.

Daniel H O'Connor, Daniel Huber, and Karel Svoboda. Reverse engineering the mouse brain. *Nature*, 461(7266):923–929, October 2009.

Philip R L Parker, Morgan A Brown, Matthew C Smear, and Cristopher M Niell. Movement-Related Signals in Sensory Areas : Roles in Natural Behavior. *Trends in Neurosciences*, pages 1–15, 2020.

Philip R. L. Parker, Elliott T. T. Abe, Emmalyn S. P. Leonard, Dylan M. Martins, and Cristopher M. Niell. Joint coding of visual input and Eye/Head position in V1 of freely moving mice. *Neuron*, 2022a. ISSN 0896-6273. doi: 10.1016/j.neuron.2022.08.029.

Philip RL Parker, Elliott TT Abe, Natalie T Beatie, Emmalyn SP Leonard, Dylan M Martins, Shelby L Sharp, David G Wyrick, Luca Mazzucato, and Cristopher M Niell. Distance estimation from monocular cues in an ethological visuomotor task. *eLife*, 11:e74708, September 2022b. ISSN 2050-084X. doi: 10.7554/eLife.74708.

Hannah L. Payne and Jennifer L. Raymond. Magnetic eye tracking in mice. *ELIFE*, 6, September 2017. ISSN 2050-084X. doi: 10.7554/eLife.29222.

Jonathan W Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M Litke, E J Chichilnisky, and Eero P Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, July 2008.

Denise M Piscopo, Rana N El-Danaf, Andrew D Huberman, and Cristopher M Niell. Diverse visual features encoded in mouse lateral geniculate nucleus. *Journal of Neuroscience*, 33(11):4642–4656, March 2013.

Rajesh P. N. Rao and Dana H. Ballard. Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, January 1999. ISSN 1097-6256, 1546-1726. doi: 10.1038/4580.

Rajesh P. N. Rao, Dimitrios C. Gklezakos, and Vishwas Sathish. Active Predictive Coding: A Unified Neural Framework for Learning Hierarchical World Models for Perception and Planning. October 2022.

Rajesh P.N. Rao and Linxing Preston Jiang. Predictive Coding Theories of Cortical Function. *Oxford Research Encyclopedia of Neuroscience*, November 2022. doi: 10.1093/acrefore/9780190264086.013.328.

Stefano Recanatesi, Matthew Farrell, Guillaume Lajoie, Sophie Deneve, Mattia Rigotti, and Eric Shea-Brown. Predictive learning as a network mechanism for extracting low-dimensional latent space representations. *Nature Communications*, 12(1):1417, March 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-21696-1.

Jacob Reimer, Emmanouil Froudarakis, Cathryn R Cadwell, Dimitri Yatsenko, George H Denfield, and Andreas S Tolias. Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *Neuron*, 84(2):355–362, October 2014.

Edgar Riba, Dmytro Mishkin, Daniel Ponsa, Ethan Rublee, and Gary Bradski. Kornia: An Open Source Differentiable Computer Vision Library for PyTorch. October 2019.

Mattia Rigotti, Omri Barak, Melissa R Warden, Xiao Jing Wang, Nathaniel D Daw, Earl K Miller, and Stefano Fusi. The importance of mixed selectivity in complex cognitive tasks SUPPLEMENTARY INFORMATION. *Nature*, 497(7451):585–590, 2013.

David Rosenbluth and John M Allman. The effect of gaze angle and fixation distance on the responses of neurons in V1, V2, and V4. *Neuron*, 33(1):143–149, January 2002.

Tomoya Sakatani and Tadashi Isa. Quantitative analysis of spontaneous saccade-like rapid eye movements in C57BL/6 mice. *NEUROSCIENCE RESEARCH*, 58(3): 324–331, July 2007. ISSN 0168-0102. doi: 10.1016/j.neures.2007.04.003.

E Salinas and L F Abbott. A model of multiplicative neural responses in parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 93(21):11956–11961, October 1996.

E Salinas and L F Abbott. Invariant visual responses from attentional gain fields. *Journal of Neurophysiology*, 77(6):3267–3272, June 1997.

Emilio Salinas and Terrence J Sejnowski. Book Review: Gain Modulation in the Central Nervous System: Where Behavior, Neurophysiology, and Computation Meet. *The Neuroscientist*, 7(5):430–440, 2001.

Jason M. Samonds, Wilson S. Geisler, and Nicholas J. Priebe. Natural image and receptive field statistics predict saccade sizes. *NATURE NEUROSCIENCE*, 21(11): 1591+, November 2018. ISSN 1097-6256. doi: 10.1038/s41593-018-0255-5.

Nicholas J Sattler and Michael Wehr. A Head-Mounted Multi-Camera System for Electrophysiology and Behavior in Freely-Moving Mice. *Serotonin Receptors in Neurobiology*, 0, 2021.

Congping Shang, Aixue Liu, Dapeng Li, Zhiyong Xie, Zijun Chen, Meizhu Huang, Yang Li, Yi Wang, Wei L. Shen, and Peng Cao. A subcortical excitatory circuit for sensory-triggered predatory hunting in mice. *NATURE NEUROSCIENCE*, 22 (6):909+, June 2019. ISSN 1097-6256. doi: 10.1038/s41593-019-0405-4.

R Angus Silver. Neuronal arithmetic. *Nature Reviews Neuroscience*, 11(7):474–489, June 2010.

Anderson Speed, Joseph Del Rosario, Christopher P. Burgess, and Bilal Haider. Cortical state fluctuations across layers of V1 during visual spatial perception. *CELL REPORTS*, 26(11):2868+, March 2019. ISSN 2211-1247. doi: 10.1016/j.celrep.2019.02.045.

JS Stahl. Using eye movements to assess brain function in mice. *VISION RESEARCH*, 44(28):3401–3410, December 2004. ISSN 0042-6989. doi: 10.1016/j.visres.2004.09.011.

Nicholas A Steinmetz, Cagatay Aydin, Anna Lebedeva, Michael Okun, Marius Pachitariu, Marius Bauza, Maxime Beau, Jai Bhagat, Claudia Böhm, Martijn Broux, Susu Chen, Jennifer Colonell, Richard J Gardner, Bill Karsh, Fabian Kloosterman, Dimitar Kostadinov, Carolina Mora-Lopez, John O'Callaghan, Junchol Park, Jan Putzeys, Britton Sauerbrei, Rik J J van Daal, Abraham Z Vollan, Shiwei Wang, Marleen Welkenhuysen, Zhiwen Ye, Joshua T Dudman, Barundeb Dutta, Adam W Hantman, Kenneth D Harris, Albert K Lee, Edvard I Moser, John O'Keefe, Alfonso Renart, Karel Svoboda, Michael Häusser, Sebastian Haesler, Matteo Carandini, and Timothy D Harris. Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science (New York, N.Y.)*, 372(6539), April 2021.

Hans Straka, Andreas Zwergal, and Kathleen E. Cullen. Vestibular animal models: Contributions to understanding physiology and disease. *JOURNAL OF NEUROLOGY*, 263(1):S10–S23, April 2016. ISSN 0340-5354. doi: 10.1007/s00415-015-7909-y.

Zdenek Straka, Tomas Svoboda, and Matej Hoffmann. PreCNet: Next Frame Video Prediction Based on Predictive Coding. December 2020.

Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Charu Bai Reddy, Matteo Carandini, and Kenneth D Harris. Spontaneous behaviors drive multidimensional, brainwide activity. *Science (New York, N.Y.)*, 364(6437):255, April 2019.

Yves Trotter and Simona Celebrini. Gaze direction controls response gain in primary visual-cortex neurons. *Nature*, 398(6724):239–242, March 1999.

Jumpei Ukita, Takashi Yoshida, and Kenichi Ohki. Characterisation of nonlinear receptive fields of visual neurons by convolutional neural network. *Scientific Reports*, 9(1):1–17, March 2019.

Gert Van den Bergh, Bin Zhang, Lutgarde Arckens, and Yuzo M Chino. Receptive-field properties of V1 and V2 neurons in mice and macaque monkeys. *Journal of Comparative Neurology*, 518(11):2051–2070, June 2010.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation Learning with Contrastive Predictive Coding. *arXiv*, 2018.

Mateo Vélez-Fort, Edward F Bracey, Sepiedeh Keshavarzi, Charly V Rousseau, Lee Cossell, Stephen C Lenzi, Molly Strom, and Troy W Margrie. A Circuit for Integration of Head- and Visual-Motion Signals in Layer 6 of Mouse Primary Visual Cortex. *Neuron*, 98(1):179–191.e6, 2018.

Martin Vinck, Renata Batista-Brito, Ulf Knoblich, and Jessica A Cardin. Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. *Neuron*, 86(3):740–754, May 2015.

Edgar Y Walker, Fabian H Sinz, Erick Cobos, Taliah Muhammad, Emmanouil Froudarakis, Paul G Fahey, Alexander S Ecker, Jacob Reimer, Xaq Pitkow, and Andreas S Tolias. Inception loops discover what excites neurons most using deep predictive models. *Nature Neuroscience*, 22(12):2060–2065, 2019.

Damian J. Wallace, David S. Greenberg, Juergen Sawinski, Stefanie Rulla, Giuseppe Notaro, and Jason N. D. Kerr. Rats maintain an overhead binocular field at the expense of constant fusion. *NATURE*, 498(7452):65–69, June 2013. ISSN 0028-0836. doi: 10.1038/nature12153.

Gordon L Walls. The vertebrate eye and its adaptive radiation [by] Gordon Lynn Walls. 1942.

Chaoyang Wang, Jose Miguel Buenaposada, Rui Zhu, and Simon Lucey. Learning Depth from Monocular Videos Using Direct Methods. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2022–2030, 2018.

Joseph B. Wekselblatt, Erik D. Flister, Denise M. Piscopo, and Cristopher M. Niell. Large-scale imaging of cortical dynamics during sensory perception and behavior. *Journal of Neurophysiology*, 115(6):2852–2866, June 2016. ISSN 0022-3077. doi: 10.1152/jn.01056.2015.

T G Weyand and J G Malpeli. Responses of neurons in primary visual cortex are modulated by eye position. *Journal of Neurophysiology*, 69(6):2258–2260, June 1993.

Robert H. Wurtz. Neuronal mechanisms of visual stability. *VISION RESEARCH*, 48(20):2070–2089, September 2008. ISSN 0042-6989. doi: 10.1016/j.visres.2008.03.021.

Daniel L. K. Yamins, Ha Hong, Charles F. Cadieu, Ethan A. Solomon, Darren Seibert, and James J. DiCarlo. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, 111(23):8619–8624, June 2014. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1403112111.

Alfred L Yarbus. Saccadic Eye Movements. pages 129–146, 1967.

Jacob L Yates, Shanna H Coop, Gabriel H Sarch, Ruei-Jr Wu, Daniel A Butts, Michele Rucci, and Jude F Mitchell. Beyond Fixation: Detailed characterization of neural selectivity in free-viewing primates. November 2021.

ChaoQiang Zhao, QiYu Sun, ChongZhen Zhang, Yang Tang, and Feng Qian. Monocular depth estimation based on deep learning: An overview. *Science China Technological Sciences*, 63(9):1612–1627, September 2020. ISSN 1674-7321, 1869-1900. doi: 10.1007/s11431-020-1582-8.

Zheng-dong Zhao, Zongming Chen, Xinkuan Xiang, Mengna Hu, Hengchang Xie, Xiaoning Jia, Fang Cai, Yuting Cui, Zijun Chen, Lechen Qian, Jiashu Liu, Congping Shang, Yiqing Yang, Xinyan Ni, Wenzhi Sun, Ji Hu, Peng Cao, Haohong Li, and Wei L. Shen. Zona incerta GABAergic neurons integrate prey-related sensory signals and induce an appetitive drive to promote hunting. *NATURE NEUROSCIENCE*, 22(6):921+, June 2019. ISSN 1097-6256. doi: 10.1038/s41593-019-0404-5.