

The midbrain reticular formation in flexible visual decision-making

Jordan Robert Shaker

A dissertation

submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:

Nicholas Steinmetz, Chair

Shawn Olsen

Adrienne Fairhall

Program Authorized to Offer Degree:

Neuroscience

©Copyright 2025

Jordan Robert Shaker

Abstract

The midbrain reticular formation in flexible visual decision-making

Jordan Robert Shaker

Chair of the Supervisory Committee:

Nicholas Steinmetz

Department of Neurobiology and Biophysics

A hallmark of mammalian behavior is the ability to rapidly remap actions in response to sensory stimuli depending on internal representations of the environment. Flexible visual decisions are thought to be computed within recurrent interactions across diverse brain circuitry in the frontal cortex, basal ganglia, and midbrain. However, the precise regions involved and their computations remain unclear. Previous work has found that a large and poorly understood structure of the midbrain, the midbrain reticular formation (MRF), contains similar task activity patterns to well-established decision regions and connects extensively with them. Open questions include whether the task dynamics in MRF reflect motor processing versus abstract task rules as well as how structure and function organize within the MRF. In this dissertation, we dissect the organization of MRF and demonstrate a role for MRF in abstract context computation using a novel flexible decision-making task, large-scale electrophysiological recordings, modeling of task behavior and neural population dynamics, and single neuron

morphological reconstructions. In **Chapter 1**, we first review the approaches to studying flexible visual decision-making in the lab and the current understanding of the regions and computations involved. We then discuss what is currently known about MRF's functional and anatomical characteristics and examine historical perspectives on its role in contextual processing. Next, we introduce our novel flexible decision-making task and show that mice achieved high levels of performance by integrating an abstract context belief variable with visual stimuli to remap stimulus-action associations (**Chapter 2**). We performed dense electrophysiological recordings within MRF and established nodes of the flexible-decision making circuitry and find that MRF, in a network with the superior colliculus (SC), secondary motor cortex (MOs), and caudoputamen (CP), maintained a baseline representation of context which putatively enabled flexible remapping by shifting action attractor dynamics between contexts (**Chapter 3**). In **Chapter 4**, we record from both trained and task-naive mice while passively presenting stimuli and find that task-specific visual stimulus representations appear throughout MRF as a result of task learning. Then, we examine the spatial distribution of context-coding neurons and cortical inputs across MRF, revealing that context-coding neurons and cortical axon terminals both contain non-uniform distribution patterns and are spatially aligned (**Chapter 5**). Altogether, our results establish MRF as a key node in the circuitry underlying flexible visual decisions and provide fundamental insights into how the brain processes contextual information to flexibly update responses to environmental stimuli. In **Chapter 6**, we integrate prior literature with the findings in this dissertation to update current theories of MRF and discuss future work to build upon and test these theories.

Table of Contents

Table of Contents.....	1
List of Figures.....	4
Acknowledgements.....	5
Author contributions.....	7
Chapter 1: Introduction.....	8
1.1 Flexible decision-making.....	9
1.1.1 Behavioral flexibility.....	9
1.1.2 Decision-making paradigms.....	12
1.1.3 Canonical circuitry underlying perceptual and value-based decision-making.	18
1.1.4 Flexible decision-making tasks and circuitry.....	22
1.1.5 Distributed computation in flexible decision-making.....	27
1.2 Midbrain reticular formation.....	32
1.2.1 Anatomy and Connectivity.....	32
1.2.2 Historical perspective.....	37
1.2.3 Established motor functions in MRF.....	39
1.3 Main objectives: Re-evaluating MRF function.....	41
1.4 General methods.....	43
1.4.1 Subjects.....	43
1.4.2 Surgical procedures.....	44
1.4.3 Surgeries for electrophysiology.....	45
1.4.4 Electrophysiological recordings.....	46
1.4.5 Histology and imaging for probe tracking.....	47
1.4.6 Neural data spike sorting and preprocessing.....	48
1.4.7 Statistics.....	49
Chapter 2: Mice integrate abstract contextual rules with visual stimuli to remap	

actions in a flexible decision-making task.....	51
2.1 Introduction.....	51
2.2 Methods.....	55
2.2.1 Task structure.....	55
2.2.2 Task training protocol.....	58
2.2.3 Session inclusion criteria.....	60
2.2.4 Disentangled RNN behavioral model.....	61
2.2.5 Context-belief model.....	63
2.2.6 Center stimulus-specific Q-Learning and win-stay-lose-switch.....	64
2.3 Results.....	65
2.4 Discussion.....	71
Chapter 3: The MRF represents abstract context associated with stimulus-action remapping in a network with SC, CP, and MOs.....	76
3.1 Introduction.....	76
3.2 Methods.....	78
3.2.1 Recording and quantification of face and pupil movements.....	79
3.2.2 Neural data preprocessing.....	80
3.2.3 State space trajectory analyses.....	81
3.2.4 Population decoding of baseline context.....	85
3.2.5 Task alignment of spontaneous movements and pupil dynamics.....	86
3.2.6 Kernel regression.....	88
3.2.7 Drift-diffusion modeling of stimulus-action remapping mechanisms.....	91
3.2.8 Pseudosessions.....	92
3.2.9 Viral injections and brain clearing.....	93
3.3 Results.....	95
3.3.1 Stimulus, choice, and context-dependent dynamics in MRF.....	95
3.3.2 Stimulus-action remapping is mediated by baseline context representations in MRF, SCm, CP, and MOs.....	100

3.4 Discussion.....	110
Chapter 4: The MRF develops behaviorally relevant visual stimulus representations as a result of task learning.....	117
4.1 Introduction.....	117
4.2 Methods.....	117
4.2.1 Passive stimulus presentation protocol.....	118
4.2.2 Quantification of stimulus encoding during passive task presentation.....	118
4.3 Results.....	119
4.4 Discussion.....	122
Chapter 5: Context coding and cortical axon terminals spatially align in MRF..	125
5.1 Introduction.....	125
5.2 Methods.....	126
5.2.1 Preprocessing of single neuron morphology datasets.....	126
5.2.2 Non-uniformity testing of cortical terminals.....	127
5.2.3 Non-uniformity testing of baseline context coding.....	128
5.2.4 Voxelization and spatial alignment of cortical terminals and baseline context coding.....	129
5.3 Results.....	130
5.4 Discussion.....	133
Chapter 6: General discussion.....	136
6.1 Updated theory of MRF function and organization.....	137
6.2 Future directions and implications.....	141
Supplementary Figures.....	147
References.....	154
Appendix A. Ultra-high density electrodes improve detection, yield, and cell type identification in neuronal recordings.....	170

List of Figures

Figure 2.1.....	67
Figure 3.1.....	97
Figure 3.2.....	102
Figure 3.3.....	107
Figure 4.1.....	120
Figure 5.1.....	131
Supplementary Figure 1.....	147
Supplementary Figure 2.....	149
Supplementary Figure 3.....	150
Supplementary Figure 4.....	151
Supplementary Figure 5.....	152
Supplementary Figure 6.....	153

Acknowledgements

As I reflect on the course of my graduate school experience, I am reminded by the highly nonlinear path I've traversed, with many equally challenging and rewarding moments throughout. There are a great number of individuals to whom I owe tremendous gratitude for supporting me throughout these years. I would not have been able to get to this point without them.

I would first like to thank my advisor, Nicholas Steinmetz, for his invaluable mentorship and guidance throughout my PhD. I still recall the moment of finding out about his relatively new lab at UW—the research topics were incredibly exciting to me, but I was worried that my lack of experience in programming and quantitative approaches would make me unfit to join the field. Nick fully reassured me upon our first meeting to discuss a potential rotation, and the rest is history. I am amazed at how much I have learned over the past five years, and that is due in no small part to Nick's kind, encouraging, and rigorous mentorship style.

Thank you to the members of my Supervisory Committee—Michele Basso, Adrienne Fairhall, Shawn Olsen, and John Neumaier—for your guidance and consistently helpful feedback on the work throughout this dissertation.

I am grateful for the entirety of the Steinmetz lab, whose members over the years have been amazing and supportive friends and colleagues. I owe a great debt of gratitude to Ljuvica Kolich for the countless hours she spent helping train mice, as well as Jeesoo Yoon, Julia Hopkins, and Rachel Yin for their additional help with training. I would also like to thank Zhiwen Ye who provided a lot of help to me early on in learning how to perform recordings, get code pipelines up and running, and generally troubleshoot things in the lab. Similarly, I would like to thank Daniel Birman for laying a lot of the initial groundwork to get behavioral mouse training up and running in the lab, helping with early designs of the flexible decision-making task we used here, and building many of the awesome tools for 3D visualization of neural data which we used here. Finally, I am very grateful to have worked with Jeremy Schroeter, an incredible colleague and friend, who joined the lab as an undergraduate to work with me on the project and ended up performing the behavioral modeling we show in Chapter 2, which is one of my favorite aspects of the entirety of the work.

On a personal level, there are so many friends and family to thank, and I surely won't be able to do justice to all of them here. First and foremost, I would like to thank my partner, Ariella Halevi, for her endless love and support over the course of this PhD. I have made a great many friends over my time here in Seattle, including my housemates, MD/PhD cohort, and the various friends I've met in between. I deeply appreciate all of them for making Seattle feel like home.

I am thankful for my parents, Orly and Mike Shaker, for their love and fostering of my scientific curiosity from a young age. It is through their support and encouragement that I was able to find my passions and focus on doing what I love. I am also so grateful for my brother Roan, who has been both a great brother and a best friend as we have grown older. Thank you for reminding me to always look at the bigger picture.

And finally, I would not be who I am today without my additional family—Brandon Boss, a best friend and brother, and Caren Ex, his mother and a mother figure to me. Both of them have taught me the values of being open, loving, and finding balance through multiple definitions of success in life. I know that Caren would have been really proud of me at this moment, as she always was.

Author contributions

All chapters were written by J. Shaker, and describe research designed and executed by him, with the following exceptions:

Chapters 2-5 describe research designed and written in collaboration with N. Steinmetz.

Behavioral modeling analyses in Chapter 2 were conducted by Jeremy N. Schroeter, designed in collaboration with J. Shaker and N. Steinmetz. Some panels in Figure 2.1 and Figure S1 were made by J. Schroeter (where noted in text).

A subset of the task-naïve recordings in Chapter 5 were conducted by J. Schroeter.

Early versions of the task design and hardware described in Chapter 2 were built by Daniel Birman.

Chapter 1.2-1.3 and Chapter 6 describe ideas, theories, and future directions formulated in collaboration with N. Steinmetz.

Appendix A was a highly collaborative effort in which J. Shaker contributed specific experiments, analyses, and writing. It is pulled directly from the published work (Ye et al., 2024), wherein details about specific author contributions are laid out.

Chapter 1: Introduction

How do we flexibly adapt our behavior in response to an ever-changing environment and internal state? For example, consider how you may react to a yellow light while driving a car. If the roads are dry, you are more likely to come to a slow stop, but if they are wet, you may drive through the intersection to avoid potential skidding. If you are in a rush to get to a meeting, you are more likely to drive through the intersection than if you are not pressed for time. It is clear that in each of these cases, our brains can take in similar or even identical sensory stimuli but decide to respond in opposing ways depending on internal state. The brain regions and processing involved in the complex and dynamic process of flexible decision-making are being actively explored. In particular, there is mounting evidence that decision-related computations emerge from interactions amongst many interconnected brain regions, shifting away from the traditional focus on cortex alone (Basso and May, 2017; Chen et al., 2024; Ding and Gold, 2010; International Brain Laboratory et al., 2023a; Steinmetz et al., 2019).

In this dissertation, we demonstrated a role for a previously unappreciated node in the flexible visual decision-making circuitry: the midbrain reticular formation (MRF). Despite its large, heterogeneous structure and substantial connectivity with decision-making circuitry (Benavidez et al., 2021; McElvain et al., 2021; Oh et al., 2014), the MRF has been relatively underexplored in the context of decision-making, with early studies primarily focusing on its roles in motor control (Ferreira-Pinto et al., 2021; Waitzman et al., 2002). Recent advances in large-scale electrophysiology have allowed for brainwide recordings during decision-making tasks, revealing that MRF has

similar task activity patterns as canonical decision-making regions (Chen et al., 2024; International Brain Laboratory et al., 2023a; Khilkevich et al., 2024; Steinmetz et al., 2019). However, whether the task dynamics in MRF are related to abstract task structure or the underlying motor processing required to report the decision has remained unknown. In the experiments presented here, we leverage a newly designed flexible visual decision-making task which unambiguously allows for the disentanglement of abstract contextual and motor-related signals. We demonstrate that the MRF represents abstract contextual signals involved in context-dependent visual stimulus-action remapping. We additionally find that visual stimulus responses in MRF neurons evolve with task training to specifically represent the behaviorally relevant task stimuli. Throughout our investigations, we pay particular attention to MRF structure, connectivity, and their relation to function. To put these findings into a broader context, we first review the historical research and current understanding of flexible visual decision-making and the MRF:

1.1 Flexible decision-making

1.1.1 Behavioral flexibility

Behavioral flexibility—the capacity to dynamically adjust responses to environmental stimuli based on internal contextual states—is a defining feature of nervous systems and intelligence (Uddin, 2021). Indeed, moving up the evolutionary tree, we find that behavioral flexibility is a preserved feature which increases in complexity, with higher organisms being able to integrate many stimuli and internal variables to produce a wide array of adaptive behaviors (Cisek, 2021a). Perhaps one of

the simplest examples is the nematode *C. elegans*, which with only 302 neurons can alternate between different foraging strategies based on its internal hunger state and prior feeding experiences (Milward et al., 2011), or the fruit fly *Drosophila*, which can change its response to carbon dioxide depending on whether it is emitted from a fermenting fruit or as a distress signal from other flies (Devineni and Scaplen, 2022). Scaling up, consider the myriad variables and potential behavioral responses displayed in foraging behavior by rodents. A mouse may repeatedly visit the same junction that leads to several food sources and make different decisions of where to go on each visit. These decisions depend on internal variables including satiety, arousal, and previous associations of the opposing directions with reward, as well as integration with external stimuli including light levels informing the time of day, odors revealing the general direction of the food, and auditory signals alerting to surrounding predators (Gire et al., 2016; Morris and Davidson, 2000). Ascending further to primates, behavioral flexibility incorporates additional layers of complexity, as elaborate social structures and abstract reasoning capacities factor into decision-making processes. The mechanisms enabling increasingly complex and flexible behaviors—which integrate more inputs, generate more potential outputs, and require more sophisticated processing of possible solutions—necessitate more elaborate nervous systems with greater numbers of neurons organized into specialized modules. Importantly, these advanced neural architectures are not created *de novo* but rather are built upon and extend the fundamental circuitry established in phylogenetically older systems. Thus, while it is clear that behavioral flexibility scales with the complexity of nervous systems across species, the fundamental mechanisms enabling flexibility are likely well-conserved, as

complexity builds upon and co-opts already existing neural architecture (Cisek, 2021a; Sherman and Usrey, 2021). As this dissertation will demonstrate, flexible decision-making in mammals involves computations within recurrent loops (Benavidez et al., 2021; Foster et al., 2021a) between phylogenetically old and more recently developed brain regions.

The neural circuitry underlying flexible behavior is of particular relevance to human health, as rigid behaviors and impaired decision-making are a prominent feature of numerous neurological and psychiatric disorders. For example, obsessive-compulsive disorder involves repetitive behaviors that persist despite lacking positive reinforcement and often causing physical and emotional harm to the individual (Gruner and Pittenger, 2017). Individuals with autism spectrum disorder (D'Cruz et al., 2013), major depressive disorder (Cella et al., 2010), and schizophrenia (Floresco et al., 2009; Sterzer et al., 2019) have all been shown to display increased perseverative behaviors compared to controls in laboratory flexible decision-making tasks. A clear sign of the progressive deterioration in cognitive abilities in neurodegenerative conditions such as Alzheimer's disease and Parkinson's disease is often an onset of behavioral rigidity and lack of inhibitory control (Gleichgerricht et al., 2010). Each of these disorders affects different components of behavioral flexibility. The prevalence of flexibility impairments across neuropsychiatric disorders, each of which has distinct etiologies, suggests that the circuitry underlying behavioral flexibility is vast, with disruptions to specific regions and neuromodulatory systems likely producing unique behavioral deficits (Cáceda et al., 2014).

1.1.2 Decision-making paradigms

Research into the neural mechanisms of flexible behavior predominantly examine flexibility in decision-making paradigms. Decision-making—broadly defined as the process of evaluating alternatives and selecting a course of action—is a useful paradigm to study behavioral flexibility because decisions are formed at a discrete time, and the inputs and outputs (i.e., available actions used to report the decision) can be precisely controlled. As there are innumerable factors which may play into flexible behavior, decision-making paradigms ensure that a finite number of inputs and outputs can be tracked while controlling for confounding variables around decision formation. In this way, one can confidently ascribe a neural correlate or an effect of a neural perturbation to a specific aspect of the decision-making process, whether that be sensory processing, tracking of internal contextual variables, or motor planning (Gold and Shadlen, 2007; Gold and Stocker, 2017; Hanks and Summerfield, 2017; Heekeren et al., 2008). To make these concepts more concrete, we will review examples of well-established decision-making tasks across species, highlighting those which require behavioral flexibility in **Chapter 1.1.4** and relating task designs to the variables they are able to measure in the brain. But first, we must describe a distinction often made between two major categories of decision-making: perceptual decision-making and value-based decision-making. While this distinction is useful conceptually and experimentally, in naturalistic settings both types are most often happening simultaneously. Each type can incorporate flexibility in different ways, and the neural circuitry underlying each type is distinct but overlapping.

Perceptual decision making involves selecting responses based on sensory evidence and perceptual categorization (e.g., identifying a stimulus as signal versus noise or differentiating between two noisy stimuli), while value-based decision making involves choosing between options based on their subjective value or utility to the decision maker, often involving trade-offs between different rewards or outcomes (Bari and Cohen, 2021). Tasks designed to study perceptual decision-making typically involve a noisy sensory stimulus which must be evaluated to make a categorical decision, and the decision is reported with a motor action for reward. Reward differences between choice alternatives are typically static and binary (i.e., value representations are constant) such that the distinction between correct and incorrect evaluation of the sensory evidence is certain (Gold and Shadlen, 2007; Hanks and Summerfield, 2017). In contrast, value-based decision making tasks typically make use of equally salient, unambiguous sensory stimuli such that perceptual uncertainty is minimized. Subjects learn the value associated with the different stimuli over long or short timescales; value can be varied in reward size (e.g. 3 vs. 1 sucrose water drops) and/or reward type (e.g. sweet vs. bitter food).

There are a few dimensions along which decision-making tasks, both perceptual and value-based, vary across experiments. First is the sensory modality of stimuli - the predominant modality used depends on the species, with non-human primates predominantly performing visual tasks, while rodents have historically mostly trained to perform tasks using olfactory, tactile, and/or auditory cues. More recently, visual tasks have been adapted for rodents (Burgess et al., 2017; Carandini and Churchland, 2013; Glickfeld et al., 2014; Huberman and Niell, 2011). Visual tasks offer several

experimental advantages over other sensory modalities, including precise temporal control of stimuli, extensive flexibility over parameters controlling stimulus properties, and ease of isolation from confounding environmental variables compared to other modalities (e.g., noise from electrical components, external odors, etc.). Visual tasks in mice also facilitate more direct comparisons with the extensive primate visual decision-making literature, as the neural circuitry that is engaged during decision-making tasks in part depends on the sensory modality of the cues (Burgess et al., 2017; Carandini and Churchland, 2013; Hutmacher, 2019). Contrary to the commonly held view that mice are minimally visual, mice heavily rely on their visual system: They use vision for navigation (Chen et al., 2013) and hunting (Hoy et al., 2016), and can become entirely diurnal depending on food availability and environment (Daan et al., 2011). An advantage of using mice for studying decision-making is genetic tractability and accessibility to advanced recording techniques. Indeed, resources and tools such as Allen Brain Atlas (Jones et al., 2009; Q. Wang et al., 2020), extensive transgenic lines (Zeng and Madisen, 2012), optogenetics (Yizhar et al., 2011), and large-scale electrophysiology with Neuropixels probes (Jun et al., 2017b; Steinmetz et al., 2021b) enable experimenters to test hypotheses about decision-making related circuitry that were previously difficult or impossible to address. In **Chapter 3**, we leverage large-scale electrophysiology in flexible visual decision-making experiments with mice, measuring the activity of thousands of neurons from regions across the forebrain and midbrain.

The motor report method in decision-making tasks is constrained by whether the subject is head-fixed or freely moving. Head-fixed non-human primates are usually

instructed to perform saccades to indicate decisions (Basso and Wurtz, 1998; Gold and Shadlen, 2000; Horwitz and Newsome, 1999), though some non-visual tasks have also required motor reports with hand or arm movements (Romo and de Lafuente, 2013; Tsunada et al., 2016). Head-fixed rodent paradigms often use licking (Guo et al., 2014b) or forepaw actions (e.g., wheel turning, lever pressing) (Burgess et al., 2017; Carandini and Churchland, 2013). Other rodent tasks use virtual reality to present stimuli, wherein the subject typically runs on a freely moving ball or wheel and uses locomotion to interact with stimuli and report decisions. Virtual reality paradigms usually introduce a component of spatial navigation to the task, which may engage separate circuitry (discussed further below) (Flossmann and Rochefort, 2021; Kira et al., 2023; Tseng et al., 2022). If rodents are not headfixed, a common way to indicate choices is by nose-poking in designated ports. We mainly focus on head-fixed paradigms here, as they comprise the majority of decision-making tasks wherein neural activity is being recorded due to the experimental advantages of the setup. An additional important distinction in motor reports for a perceptual decision-making task is Go/No-Go vs. two-alternative choice (or forced choice) designs. In the former, the perceptual decision is reported through the execution or withholding of a single action, and the decision is typically based on the presence or absence of a single stimulus. A confound of this paradigm is that the subject's internal motivational state greatly influences propensity to take *any* action, thereby making it difficult to disentangle errors in perceptual judgment vs. disengaged states. Two-alternative choice tasks alleviate this confound, as the reports for presence or absence of the stimulus both require actions, but to opposing directions. These paradigms can also be extended to be two-alternative *forced* choice

(2AFC), wherein two stimuli must be perceived and judged for a particular attribute, and the subject's motor report indicates which of the two stimuli contains this attribute. 2AFC tasks have the advantage of clearer tracking of the decision variable in the subject's behavior: if the subject develops a motor bias toward a particular direction, this is easier to distinguish from changes in perception when there are two stimuli, as matched trials (e.g., 100/50 vs. 50/100 left vs. right contrast discrimination) across the two stimuli can be compared (Carandini and Churchland, 2013; Gold and Shadlen, 2007). Finally, a major way tasks differ is in the time between stimulus presentation and motor report: Tasks may impose a delay period, wherein the subject must perceive the stimulus and withhold their action for a period of time until a go cue. Tasks without delay periods are referred to as reaction-time tasks, as the subject is free to respond as soon as the stimulus appears and thus the response is dependent on reaction time. There are benefits and drawbacks to each of these approaches, which we discuss further in **Chapter 2.1**.

We now review two classical laboratory decision-making tasks in non-human primates and rodents to illustrate the concepts presented above. We focus on head-fixed perceptual decision-making tasks as these are most relevant to the task we designed and used in this dissertation, described in **Chapter 2**. We will start with the random dot motion (RDM) task, which is perhaps the most classical perceptual decision-making paradigm in non-human primates (Newsome and Paré, 1988). The RDM task is a sensory evidence accumulation task. It requires subjects to determine the net direction of motion in a display where a subset of dots moves coherently in one direction while the remaining dots move randomly. The subject fixates on a central point

while viewing a circular aperture containing the moving dots. The difficulty is controlled by adjusting the percentage of coherently moving dots (motion strength), with lower strength creating greater perceptual uncertainty and requiring longer evidence accumulation. After viewing the stimulus for a fixed duration, there is a delay period, after which subjects report their perceived motion direction with a saccade to one of two peripheral targets positioned to correspond with the motion directions (e.g., right or left). Reaction-time based versions of the RDM task (i.e., no fixed duration for stimulus presentation and no delay period) have been used as well, with both accuracy and speed scaling with motion strength (Roitman and Shadlen, 2002). The decision variable primarily measured in the RDM task is evidence accumulation, which reflects the integration of motion signals over time until a decision threshold is reached (Horwitz and Newsome, 2001; Shadlen and Newsome, 2001). As we will discuss below, clear neural signatures of the evidence accumulation process have been found in several brain regions recorded in the RDM task.

In rodents, a prominent perceptual decision-making paradigm uses tactile discrimination and directional licking motor reports (Guo et al., 2017, 2014a, 2014a). In this 2AFC design, head-fixed mice discriminate the relative position of a tactile stimulus delivered to their whiskers. Specifically, a small pole is presented either to the anterior or posterior aspect of a single whisker. Mice are trained to report which side received the stimulus by licking one of two spouts positioned laterally, corresponding to the stimulus location (i.e., lick left if the stimulus was posterior, lick right if it was anterior). Importantly, the task incorporates a delay period (typically 1-2 seconds) between stimulus presentation and an auditory response cue, requiring the animal to maintain a

representation of the stimulus in working memory and prepare the appropriate motor response during this delay. This task paradigm has been adapted to incorporate different types of stimuli, including varying pole textures rather than locations and using auditory rather than tactile stimuli for discrimination. In **Chapter 2**, we discuss a more recently developed visual wheel-turning paradigm designed for head-fixed mice, which we adapt into a flexible decision-making task.

1.1.3 Canonical circuitry underlying perceptual and value-based decision-making

We will next give a broad overview of the brain regions that have been implicated in perceptual and/or value-based decision-making in both non-human primates and rodents. The overview here is not comprehensive of all regions having been studied, but rather, focuses on the regions for which there is the most literature and strongest evidence for roles in decision-related processing. The neural circuitry for perceptual and value-based decision-making comprises distinct yet overlapping networks. We begin by examining non-human primate studies. Starting with cortical regions, two of the most-studied regions involved in perceptual and value-based decisions, respectively, are the lateral intraparietal cortex (LIP) and orbitofrontal cortex (OFC). Historically, the aforementioned RDM tasks heavily focused on LIP, in part due to its connectivity with visual and oculomotor cortical circuitry. LIP neurons exhibit a ramping activity during presentation of a random dot motion discrimination task: This activity is predictive of the subject's upcoming saccade choice, and the steepness of the ramping corresponds to the strength of sensory evidence (i.e. motion coherence) (Shadlen and Newsome, 2001; Steinemann et al., 2023). Similar ramping activity has been seen in the frontal eye fields

(FEF) (Gold and Shadlen, 2000; Hanes and Schall, 1996) and dorsolateral prefrontal cortex (dlPFC) (Kim and Shadlen, 1999). Aside from visual tasks, classical studies using a vibrotactile detection task have found correlates of the decision variable in secondary somatosensory cortex (S2) and medial/ventral premotor cortices (Romo et al., 2002; Romo and de Lafuente, 2013). In contrast, studies on the encoding of value in the cortex have traditionally focused on orbitofrontal (OFC)¹ and anterior cingulate cortex (ACC). OFC neurons have been found to encode the value of presented stimuli independent of the visuospatial presentation of the stimuli or the ensuing motor response (Padoa-Schioppa and Assad, 2006). A wealth of studies have implicated the ACC in determining the value of alternative choices, integrating information about perceived reward, difficulty of acquiring the reward, etc, to enable behavioral flexibility in a dynamic environment (Kolling et al., 2016). The above studies have led to a loose association of perceptual decisions to the parietal cortex and value-based decisions to the frontal cortex (Smith and Krajbich, 2021). Of course, the reality is more complicated, as there is substantial overlap between these processes: representations of reward size and probability have been found in the same LIP and dlPFC neurons involved in perceptual decision-making (Gold and Shadlen, 2007), and both ACC and OFC play roles in action selection (Cruz et al., 2022).

With respect to subcortical involvement, both perceptual and value-based decision-making crucially depend on the striatum. As in area LIP and dlPFC, the caudate has been shown to contain evidence accumulation signals (Ding and Gold,

¹ There is a wealth of literature on the role of OFC (and its interactions with subregions of the striatum) in the field of *flexible* value-based decision-making, in particular reversal learning and model-free vs. model-based reinforcement learning (Bradfield and Hart, 2020; Miller et al., 2022; Wilson et al., 2014). While not our main focus in this dissertation, we will touch on these concepts in **Chapter 2** as they are useful in thinking about mechanisms underlying flexible decision-making more broadly.

2013, 2010). Neuronal correlates of reward expectation have also been found in the caudate and putamen (Haruno and Kawato, 2006). Cortico-striatal loops have been proposed to control heterogeneous decision processes involved in both flexible value-dependent actions and stimulus-bound perception-dependent actions (Balleine et al., 2007). One overarching view is that the cortex provides signals to bias action selection to the striatum, which is responsible for action selection itself. This view is partly based on the recurrent circuitry in the cortex vs. weak/lateral inhibitory circuitry in the striatum, which facilitate information integration and a winner-take-all operation respectively (Bari and Cohen, 2021; Daw et al., 2005). A subcortical region more specifically involved in perceptual decision-making is the superior colliculus (SC). Classical studies have unequivocally shown that SC (the deep layers in particular) plays a key role in perceptual decision-making, with a wealth of both correlational and causal evidence to support this notion (Basso and May, 2017; Basso et al., 2021). For instance, ramping signals are found in the SC during evidence accumulation tasks, as in LIP, FEF, caudate, etc. (Horwitz and Newsome, 1999). More recent work causally tested the role of the SC in evidence accumulation using unilateral inactivation with muscimol: Results using signal detection theory and other modeling showed that the resulting changes in behavior from this manipulation are consistent with the SC regulating the gain of evidence accumulation (Jun et al., 2021). Similar results were found in recordings and manipulations of SC in an abstract visual categorization task (Peysakhovich et al., 2024). A separate study used SC inactivations to propose a causal role for SC in terminating visual decisions by applying a threshold mechanism to signals from LIP (Stine et al., 2023) .

Decision-making circuitry in mice in many ways parallels the non-human primate circuitry, despite obvious differences in cortical organization. In both species, it has been shown that frontal and parietal cortices, basal ganglia, and SC form key nodes in decision networks (Chen et al., 2017; Essig et al., 2021; Felsen and Mainen, 2012, 2008; Goard et al., 2016; Guo et al., 2014b; Lee et al., 2020; L. Wang et al., 2020; Zhong et al., 2019). In mice, the anterior lateral motor cortex (ALM), a subregion of secondary motor cortex (MOs), has been extensively studied in delayed-response directional licking paradigms and shows some functional properties analogous to primate frontal eye fields with saccades in terms of motor planning and choice representation (Chen et al., 2017; Guo et al., 2017, 2014b; Inagaki et al., 2019). A subset of the field examines visually-guided navigational decisions with virtual reality paradigms, focusing on posterior cortical regions—specifically primary visual cortex (V1), retrosplenial cortex (RSP), and posterior parietal cortex (PPC)—and their overlapping roles in processing visual stimuli, spatial positioning, choice dynamics, and locomotion (Driscoll et al., 2017; Kira et al., 2023; Tseng et al., 2022). For value-based decisions, the orbitofrontal cortex (OFC), anterior cingulate cortex (ACC), medial prefrontal cortex (mPFC), and striatum in mice serve roles in comparing expected values of actions and evaluating outcomes to drive learning, as has been found in primates (Bari et al., 2019; Bari and Cohen, 2021; Bradfield and Hart, 2020; Cole et al., 2024; Miller et al., 2022; Ottenheimer et al., 2023). Finally, as in primates, several studies have demonstrated causal roles for the SC in perceptual decision-making (Duan et al., 2021; Essig et al., 2021; Felsen and Mainen, 2012, 2008; Lee and Sabatini, 2021; Thomas et al., 2023; L. Wang et al., 2020).

1.1.4 Flexible decision-making tasks and circuitry

In our review of decision-making circuitry thus far, we have not yet highlighted tasks which specifically include components of behavioral flexibility. We next review four illustrative examples of flexible *perceptual* decision-making paradigms² as well as the brain regions implicated in their performance. Tasks designed to investigate behavioral flexibility in decision-making introduce context-dependence in some form, requiring flexible switching of task rules to remap stimulus-action contingencies (Okazawa and Kiani, 2023). If the stimuli in the task are noisy or require discrimination in some way, then an additional variable of perceptual evidence can be tracked, though the task must be carefully designed such that lapses in perception can be distinguished from failure to track the task context. Flexible decision-making tasks are ideally designed for the experimenter to unambiguously distinguish signals related to abstract task contextual variables from those related to motor planning. In other words, if the action associated with a particular perceptual decision is fixed throughout a task, it is difficult to conclude whether the neural signals correlated with this decision are related to the decision itself or to the motor planning required for the specific action used to report the decision. We further discuss this important concept and its implication in interpreting neural activity in decision-making tasks in **Chapter 2.1**.

The majority of flexible decision-making studies have been conducted in non-human primates, as training of such tasks has traditionally been difficult in rodents,

² There is also a wealth of literature on flexible *value-based* decision-making. This process, often studied using reversal learning and dynamic reward paradigms and heavily linked to OFC, striatum, and dopamine-mediated reinforcement of action-outcome values (Bari et al., 2019; Bari and Cohen, 2021; Izquierdo et al., 2017; Schultz et al., 1997), differs from our current focus on how abstract task rules guide flexible perceptual decisions, which likely engages a distinct but overlapping set of underlying circuitry. We will explore the distinctions between these forms of flexibility, particularly in relation to reversal learning tasks, in more detail in **Chapter 2**.

especially using head-fixed preparations. More recently, however, rodents have successfully been trained to achieve high performance on flexible decision-making paradigms (Duan et al., 2021; Rodgers and DeWeese, 2014; Tseng et al., 2022; Wu et al., 2020). A well-established type of flexible decision-making task is delayed match-to-sample (DMS) (Kira et al., 2023; Liu et al., 2014; Lu et al., 1993; Paule et al., 1998; Wallis et al., 2001; Wu et al., 2020). Subjects are briefly presented with a sample stimulus, followed by a delay period and then presentation of multiple test stimuli. The subject must select the test stimulus that matches the sample (or in some variations, the non-matching test stimulus), and the location of the test stimulus can change between trials. Thus, the task requires flexible mapping between the test stimulus and action, and it additionally requires subjects to hold a representation of the sample stimulus in working memory. Regions implicated in mediating the flexible stimulus-action associations of the DMS task include PFC and parietal cortex in primates (Lara and Wallis, 2015; Wallis et al., 2001), ALM and mPFC in mice (Liu et al., 2014; Wu et al., 2020), and PPC/RSP in navigational DMS tasks for mice (Kira et al., 2023). A task that similarly requires flexible stimulus-action associations is the “loose stimulus-response association” (LSRA) task from Horwitz, Batista, and Newsome (Horwitz et al., 2004). The LSRA task is a variant of the RDM task that requires a DMS component: subjects first view the RDM to determine a general direction of coherent motion (i.e., left vs. right), but rather than there being the same two possible saccade targets for the motor report on each trial, there are 14 different locations where the paired saccade targets may appear, and the subject must then choose the relative left or relative right target based on the sample stimulus motion direction. The SC was found to encode abstract

categorical information related to perception of the sample stimulus in this task: The choice was represented in a subset of SC neurons before targets appeared, thus making these signals less concretely related to motor planning (Horwitz et al., 2004). Another flexible variant of the RDM task is the motion/color context task from Mante & Sussillo et al. (Mante et al., 2013). Monkeys view random dot stimuli that simultaneously contain both motion and color information. Depending on a contextual cue indicating the relevant dimension (motion or color), the subject must selectively attend to and report one feature while ignoring the other. Population dynamics in PFC were found to both select the relevant input stimuli and integrate the sensory evidence towards a choice. We explore this landmark study further in **Chapter 3**, wherein we employ similar analyses of dynamics in neural population activity to understand the mechanisms by which context-dependent computation occurs in our task. Finally, in a similar vein as DMS tasks, Pro/Anti tasks dissociate perceptual decisions from motor planning in that they require the subject to sample a task cue and hold the cue in working memory to perform the correct action when the target stimuli appear (Duan et al., 2021, 2015; Munoz and Everling, 2004). In these tasks, rather than matching the sample stimulus to the test stimulus, the first cue indicates a task rule - pro or anti - which indicates whether the subject should make a choice toward or away from the target stimulus (e.g. saccading left when a left target stimulus appears in pro vs. saccading right for a left stimulus in anti). This task has been extensively investigated in non-human primates, wherein dIPFC, LIP, FEF, and regions of the basal ganglia have been proposed to impose top-down inhibition of saccade motor neurons to enable the anti behavior (Munoz and Everling, 2004). A variant of the task has been trained in rats, wherein both

the PFC and SC have been found to play causal roles in the response inhibition and context-dependent stimulus-action remapping required by the anti rule (Duan et al., 2021, 2015).

Studies on the neural circuitry underlying flexible decision-making have focused on cortical regions, and in particular PFC (Aoi et al., 2020; Asaad et al., 2000; Duan et al., 2015; Durstewitz et al., 2010; Mante et al., 2013; Miller and Cohen, 2001; Pasupathy and Miller, 2005; Reinert et al., 2021; Wallis et al., 2001), to a perhaps even greater extent than studies of traditional perceptual decision-making tasks. This focus may in part stem from the notion that flexible decision-making requires higher order cognitive processing. Since PFC evolved relatively recently, expanded dramatically in primates, extensively connects with other sensory, associative, and motor-planning areas, and possesses recurrent neuronal architecture capable of supporting complex computations, it is reasonable to conclude that it is likely the primary substrate for the computational processes that enable behavioral flexibility during decision-making (Miller and Cohen, 2001; Okazawa and Kiani, 2023). However, as highlighted in **Chapter 1.1.1**, flexible behavior is displayed by a wide range of organisms, many of which have less developed cortical structures or even lack a neocortex altogether. Indeed, in terms of the computational complexity involved, consider the demands posed by naturalistic behaviors as in our previous example of dynamic foraging in rodents - these greatly exceed the demands of most laboratory flexible decision-making tasks, e.g. remapping stimulus-action contingencies in the DMS task. As we have seen in the studies utilizing LSRA and Pro/Anti tasks (Duan et al., 2021; Horwitz et al., 2004), as well as others (Peysakhovich et al., 2024), there is substantial evidence that SC (i.e. the optic tectum,

a structure dating back to early vertebrates) computes abstract categorical information required for the flexible behaviors in these tasks. It is important to recall that the cortex was added onto an already functional nervous system, and thus its computations are intertwined with the subcortical circuitry upon which it was built (Cisek, 2021a; Sherman and Usrey, 2021). And while a unique and important aspect of the cerebral cortex is its recurrent architecture, which is indeed necessary to perform the relative computations for flexible-decision making, there is substantial recurrence within the looping connections between cortex, basal ganglia, midbrain, cerebellum, and thalamus (Foster et al., 2021a; Hintiryan et al., 2016a; Lee et al., 2020; Oh et al., 2014; Peters et al., 2021).

How can we reconcile the diversity of studies showing that PFC (and other frontal, parietal, and even motor and sensory cortical regions), SC, and basal ganglia all perform computations, often with functional overlap, that contribute to flexible decision-making? The evidence strongly suggests that each of these regions participate as nodes within a distributed network supporting flexible decision-making (Okazawa and Kiani, 2023), though precisely which region serves which function, and how this network architecture may differ across different task definitions, remains unclear. In particular, there are two key challenges in making general conclusions. First, most studies have examined only one or two brain regions in isolation, making comparisons between regions in a given study difficult, and leaving many brain areas completely unexplored for their potential role in decision processes. Second, methodological differences between studies, which utilize diverse behavioral paradigms with varying definitions of decision variables, can complicate direct comparisons of their findings. We next

highlight a set of studies, enabled by advances in large-scale electrophysiology, which begin to address these challenges.

1.1.5 Distributed computation in flexible decision-making

The scale of neural recordings with electrophysiology has steadily increased since the first recordings in the 1950s, progressing from single recording sites to hundreds per shank. Neuropixels probes—microelectrode arrays based on complementary metal-oxide semiconductor (CMOS) technology containing nearly 1000 recording sites per shank—represent a significant advancement in large-scale electrophysiology capabilities (Jun et al., 2017a; Steinmetz et al., 2021a, 2018). One Neuropixels probe can record from hundreds of neurons spanning multiple brain regions, and their small size allows for multiple (up to 8 in one study) probes to be inserted in a mouse brain simultaneously. The scale and efficiency of Neuropixels probes has made it feasible, for the first time, to record from nearly all mouse brain regions across different sessions within the same study. We next review four key studies leveraging brain-wide Neuropixels recordings during different mouse perceptual decision-making tasks (Chen et al., 2024; International Brain Laboratory et al., 2023a; Khilkevich et al., 2024; Steinmetz et al., 2019). We integrate their findings, highlighting key features of each task, how they reveal distributed coding of specific task variables (including sensory processing, motor planning, and sensorimotor transformation), and common brain regions that encoding similar variables across studies. We also discuss the limitations of these studies, in particular the task designs, in making conclusions about the coding of abstract variables related to flexible decision-making.

The first brain-wide survey of perceptual decision-making task-related activity was performed by Steinmetz et al (Steinmetz et al., 2019). The study recorded 30,000 neurons in 42 brain regions across the forebrain and midbrain (but not the cerebellum, hypothalamus, or hindbrain) while mice performed a visual 2AFC/Go-NoGo task. This head-fixed task utilizes a visual wheel turning paradigm, wherein mice sit in front of three panoramic screens with their forepaws on a wheel³ (Burgess et al., 2017). Mice had to discriminate between the contrasts of gabor stimuli to their left and right, choosing the left or right stimulus by turning the wheel clockwise or counter-clockwise (the horizontal position of the stimulus tracked wheel movement). If no stimuli were present on a trial, the mice had to withhold turning the wheel for reward. Neurons across virtually all brain regions recorded, to varying degrees, exhibited non-selective responses during action initiation. In contrast, neurons encoding visual stimuli and upcoming choices were confined to specific areas within the neocortex, basal ganglia and midbrain. Neurons encoding visual stimuli were present in regions that are part of the visual pathway, e.g. primary visual cortex, lateral geniculate nucleus, and superficial SC, but they surprisingly also appeared, with a longer latency response, in motor and somatosensory cortex and several midbrain regions. Choice-encoding neurons (i.e., those with premotor activity significantly higher for one action over the other) were sparse and emerged simultaneously across a limited set of regions. The regions with the strongest choice-encoding included canonical decision-related regions—secondary motor cortex (MOs), caudoputamen (CP) and substantia nigra pars reticulata (SNr) of the basal ganglia, and deep layers of SC (motor-related SC; SCm), as well as potentially unexpected regions including primary motor cortex (MOp), somatosensory

³ We adapt this paradigm for the novel flexible decision-making task presented in **Chapter 2**.

cortex (SSp), the zona incerta (ZI), and the midbrain reticular formation (MRF).

The International Brain Laboratory (IBL), a consortium of 11 labs, increased the scale of electrophysiology even further, recording from about 300,000 neurons across the entire mouse brain using standardized and repeatable protocols (International Brain Laboratory et al., 2023b). The IBL utilized a similar visual wheel turning task paradigm as Steinmetz et al. 2019, with two key differences: First, only one stimulus appeared on the screen at a time (at varying contrasts), second, the task was block-structured such that the probability of receiving a left vs. right stimulus changed between 80% and 20% between blocks. Mice were shown to incorporate this prior probability into their choices, allowing for the measurement of prior expectation in neural activity in addition to stimulus, action, and choice. The regions representing action, visual stimulus, and choice were broadly similar as in Steinmetz et al. 2019. Choice was additionally found to be encoded in parts of the cerebellum and hindbrain (which weren't sampled in Steinmetz et al 2019). The subjective prior variable was also coded brain-wide, in regions across the cortex, basal ganglia, midbrain, and hindbrain (Findling et al., 2023).

Two additional brain-wide surveys from Chen et al. and Khilkevich et al. use different task designs to provide further insight into circuitry involved in memory-guided movement and sensorimotor transformation (Chen et al., 2024; Khilkevich et al., 2024). Chen et al. utilized the well-established 2AFC directional licking paradigm described in **Chapter 1.1.2**, which emphasizes the interplay between working memory and motor planning through its use of a delay period. They find broadly similar distributions of stimulus, action, and choice coding as Steinmetz et al. 2019 and the IBL. Choice coding in this study was particularly enhanced in ALM and the regions to which it connects,

likely due to its specific role in licking. Khilkevich et al. used a Go/NoGo visual change detection task requiring evidence integration. The stimulus was a gabor whose temporal frequency noisily fluctuated around 1 Hz; when there was a sustained increase frequency imposed by the experimenter, mice had to lick to report the change. Regions containing lick preparatory activity and stimulus temporal frequency responsive neurons were again broadly similar to the distributions of choice and visual neurons in the aforementioned brain-wide studies. However, this study provides additional insight into why coding of sensory stimuli is widely distributed outside the visual pathway: Comparing their recordings in trained vs. task-naïve subjects, Khilkevich et al. found that stimulus representations outside the visual pathway (motor and frontal cortex, cerebellum, caudoputamen, and especially midbrain) emerged with task learning. Areas which integrated visual evidence as a result of task learning were found to contain movement preparatory subspaces whose activity was driven by neurons integrating evidence, thereby supporting a distributed computation of sensorimotor transformation across the brain.

Integrating the results from the four brain-wide surveys of perceptual decision-making tasks reviewed here, we draw three general conclusions. First, action correlates, i.e. non-selective signals prior to the motor report of the decision, are distributed across nearly all regions, possibly reflecting brain-wide forms of corollary discharge. Second, visual responses appear in many non-visual forebrain, midbrain, basal ganglia, and cerebellar regions at longer latencies than visual regions, and the development of these responses with task learning enables a distributed mechanism of transforming sensory signals to action among motor and associative areas. Finally,

while much rarer than action coding, choice coding is surprisingly distributed with particularly strong representations in both canonical decision-making regions and unexpected regions in the midbrain, hindbrain, and somatosensory cortex.

How do the results from these brain-wide studies inform our understanding of the circuitry underlying flexible decision-making? It is difficult to draw direct conclusions as none of the studies employed tasks requiring the context-dependent stimulus-action remapping that is characteristic of flexible decision-making tasks. The IBL task did implement context-dependence in the form of prior expectation, but the neural encoding of subjective prior may just as well be described as a shifting bias in motor preparatory activity. One possibility is that the widespread coding of choice may vary in functional significance across the regions in which it is coded, with choice in some regions representing a low-level motor preparatory signal and in others representing more abstract signals related to categorical representations of task variables⁴. Indeed, most of the regions which encoded choice signals are canonical decision-making regions in the sense that they have been found to code for abstract contextual variables across numerous studies (as described in the preceding section; i.e., frontal cortical regions, SC, and regions within the basal ganglia). This raises the interesting question of whether the strong choice coding in unexpected regions, most notably the MRF, ZI, and hindbrain regions (mainly aspects of the reticular formation in pons and medulla), also represents abstract task-related variables or low-level motor preparatory signals.

⁴ We should note here that the distinction between cognition and premotor processing more broadly is an ongoing matter of debate, with some viewing cognition as an extension of action selection. In this framework, all neural signals nominally related to cognition are instead part of an action abstraction hierarchy. In this dissertation, we do not strongly adopt one view over another: Our conclusions about the computation of abstract contextual signals in MRF described in **Chapter 3**, which we distinguish from low-level motor planning, do not change if one views them through a fundamentally premotor lens, wherein they would instead be described as representing higher level processing along an action abstraction hierarchy. A separate but related concept is embodiment, which we address in **Chapter 3**.

In the remainder of this dissertation, we direct our focus to the MRF and its role in flexible visual decision-making. As we will discuss next, the MRF contains several properties which support its potential role as a key node in the circuitry responsible for the contextual processing enabling flexible behavior.

1.2 Midbrain reticular formation

1.2.1 Anatomy and Connectivity

The midbrain reticular formation (MRF)⁵, named “reticular” (derived from *rete*, Latin for net, i.e. “net-like”) for its undifferentiated structure of loosely packed neurons in dense neuropil, comprises a substantial portion of the mammalian midbrain—comparable in volume to the SC and exceeding that of the periaqueductal gray (PAG) (Q. Wang et al., 2020). It resides ventral to the SC, ventrolateral to the PAG, and dorsal to the substantia nigra pars reticulata (SNr), often delineated as the territory remaining after accounting for embedded or adjacent structures like the red nucleus (RN) and fiber tracts. The MRF contains roughly equal proportions of glutamatergic and GABAergic neurons (Yao et al., 2023), and these neurons are “interneurons” relative to the periphery in that they lack direct connections with primary sensory organs or spinal motor neurons (Hara et al., 1961; Lindsley and Adey, 1961; Oh et al., 2014).

A defining feature of the MRF is its remarkably convergent pattern of afferent connections, positioning it to integrate information across disparate brain systems. It

⁵ In most non-human primate literature, the MRF is referred to as the “mesencephalic reticular formation”, and in the Allen CCF atlas it is referred to as the “midbrain reticular nucleus” or MRN. We opt for the term “formation” here to better reflect the lack of obvious nuclear organization. MRF is distinct from some other regions containing the term ‘reticular’, such as the reticular nucleus of the thalamus (an inhibitory region connecting bidirectionally to thalamic relay nuclei) and the pontine and medullary (aka “bulbar”) reticular formations, which are the main source of the reticulospinal tracts and involved in motor control (Arber and Costa, 2022a). By contrast, MRF has very little spinal projection (Liang et al., 2011).

receives direct projections from a broad swath of the cortex, including prefrontal, motor, and somatosensory areas (further explored in **Chapter 5**) (Gao et al., 2022; Oh et al., 2014; Peng et al., 2021; Winnubst et al., 2019), the basal ganglia through the SNr (McElvain et al., 2021), the cerebellum (Fujita et al., 2020), the SC (Benavidez et al., 2021; Perkins et al., 2014; Sébille et al., 2017) and the ZI (Kolmac et al., 1998; Oh et al., 2014). This extensive convergence led early anatomists to suggest the MRF is situated "to monitor activity in almost every portion of the neuraxis and in virtually every passing fiber system" (Scheibel, 1984). Reflecting its structural heterogeneity, recent unbiased spatial transcriptomic sequencing efforts have identified potentially over 700 distinct neuronal cell types within the MRF, about which almost nothing is known (Yao et al., 2023).

The inputs MRF receives from SNr, the largest output nucleus of the basal ganglia, are particularly notable as they suggest a potential integrative function from MRF. The established architecture of cortico-basal ganglia circuits involves parallel processing streams, where distinct cortical areas project through specific subregions of the striatum (Haber, 2003; Hintiryan et al., 2016b; Hunnicutt et al., 2016), globus pallidus, and the SNr (Foster et al., 2021b). This organization largely maintains segregated functional channels. Work by McElvain et al. (McElvain et al., 2021) has significantly elaborated on the downstream targets of these channels, revealing that the SNr comprises distinct output modules primarily defined by their specific, largely segregated projection targets within the brainstem. Notably, however, this study demonstrated a striking exception to this segregated pattern: the MRF, along with the pedunculopontine nucleus (PPN), receives convergent axon collaterals,

non-topographically, arising from all distinct SNr output modules. This pattern of convergence onto the MRF contrasts sharply with both the specific projections these SNr modules send to other individual midbrain targets and the topographic projections maintained by the very same SNr neurons as they ascend to the thalamus, preserving the parallel loop structure within the forebrain.

Efferent projections from the MRF are similarly widespread and diverse. We will start with MRF's midbrain neighbor, SC. Just as MRF receives dense input from the intermediate and deep layers of SC, it reciprocally connects back to SC via dense innervation of the same layers. These reciprocal connections are bilateral. Within the SC, MRF specifically connects to the centrolateral and lateral zones: Amongst all SC zones, these zones receive the sole input from primary motor cortex, secondary motor cortex, and primary somatosensory cortex, whereas the centromedial and medial zones receive the sole input from visual and auditory cortices. Amongst prefrontal and associative cortical regions, the centrolateral and lateral zones of SC preferentially receive input from orbitofrontal, prelimbic, and dorsal anterior cingulate regions, whereas the centromedial and medial zones preferentially receive input from retrosplenial and ventral anterior cingulate regions (Benavidez et al., 2021). The topography of connectivity between cortex, SC, and MRF raises interesting questions about MRF's function, as MRF is connected to SC zones which preferentially receive input from cortical regions most studied in flexible decision-making—MOs, ORB, ACA, PL; subsections of each of these regions encompass the mouse PFC broadly (Merre et al., 2021).

The MRF has ascending projections targeting key thalamic nuclei, most prominently the intralaminar nuclei (and amongst these, the parafascicular (PF) and centrolateral (CL) nuclei seem to be the most targeted), ventromedial (VM) nucleus, and mediodorsal (MD) nucleus (Kolmac and Mitrofanis, 1998; Oh et al., 2014). Notably, PF and VM thalamus are two of the primary thalamic targets of the SNr and comprise the SNr's densest outputs (Foster et al., 2021a; McElvain et al., 2021). MD thalamus extensively connects with PFC and has been implicated in flexible behavior and executive control (Lam et al., 2025; Mitchell and Chakraborty, 2013; Phillips et al., 2024; Rikhye et al., 2018). Two additional strong ascending projection targets of MRF are the GABAergic reticular nucleus of the thalamus (RT) and adjacent ZI (as ZI also projects to MRF, these regions are reciprocally connected similar to MRF and SC) (Kolmac et al., 1998; Kolmac and Mitrofanis, 1998; Oh et al., 2014). RT and ZI are unique amongst diencephalic nuclei in that they are almost entirely GABAergic and densely innervate all thalamic regions. They have thus been thought of as routing controllers, "gatekeepers", or global behavioral modulators (Crick, 1984; McAlonan et al., 2008; Mitrofanis, 2005; X. Wang et al., 2020).

The primary descending projections of MRF are the medullary and pontine aspects of the reticular formation (in Allen CCF nomenclature, these include the pontine and tegmental reticular nuclei in the pons, and the gigantocellular, magnocellular, parvicellular, paragigantocellular, and medullary reticular nuclei in the medulla) (Hayashi and Yoshii, 1966; Oh et al., 2014). The pontine and medullary reticular formations are the main sources of the reticulo-spinal tract and are broadly responsible for controlling body movements including head-eye coordination, orofacial movements, and forelimb

reaching (Arber and Costa, 2022a; Grantyn and Berthoz, 1987; Schepens and Drew, 2004, 2006).

To summarize, the MRF receives strikingly broad input from multiple nodes in the circuitry underlying flexible decision-making, and it connects back to this circuitry in recurrent loops via its outputs to thalamic nuclei and SC. Walking through these loops at a high level: The MRF receives direct inputs from prefrontal, frontal, and motor cortical areas. These same cortical areas topographically project to the basal ganglia, connecting to striatum, then SNr in parallel output modules. The parallel output modules in SNr have distinct brainstem targets including the SC, and all of these modules project topographically to thalamus, especially VM and PF thalamus. The axons going from SNr to thalamus broadly collateralize to MRF, such that the parallel output modules all converge in a non-topographic manner. In addition, the MRF reciprocally connects with SC, in particular the more lateral zones which receive prefrontal and motor cortical input. The MRF also sends ascending projections to the thalamic nuclei connected to the relevant cortical and basal ganglia circuitry—VM, PF, and MD. Additional ascending targets of MRF are the GABAergic thalamic control centers RT and ZI, the latter of which sends reciprocal connections back to MRF. A subset of these connections is summarized in Figure 3.1B.

Many interesting questions arise when considering MRF's connectivity. How is the connectivity organized within MRF? The aforementioned studies generally did not track topography or specific subregions within MRF. Are MRF's connections with the forebrain loop anatomically segregated from the descending motor-related projections to the pontine and medullary reticular formations? What may be the functional

significance of SNr's parallel modules converging non-topographically in MRF via axon collaterals targeting thalamus? While the focus of this dissertation was to explore MRF's functional role in flexible visual decision-making, we speculate on how MRF's connectivity may inform its function and discuss future avenues of investigation in **Chapter 6**.

1.2.2 Historical perspective

In the 1950's and 60's, the MRF was a central focus of neuroscience research. An international symposium with a book of over 500 pages of studies was dedicated to it, featuring some of the world's most prominent neuroscientists at the time ("International Symposium: Reticular Formation of The Brain," 1957). Quotes such as "the vast dumb millions of cells in the cerebral cortex are, it appears, the poor relations of the reticular system, pensioners living on its bounty" (Jefferson, 1958) and "it may be only a slight overstatement that any satisfactory model of a nervous system is a model of the reticular formation" (Scheibel and Scheibel, 1968) illustrate views at the time which stand in stark contrast to its MRF's status over the last twenty years as one of the most understudied regions in neuroscience. What caused the authors at the time to view MRF as central to whole-brain dynamics, and why the precipitous decline in interest?

The first studies to ignite the field's interest in MRF were performed by Moruzzi, Magoun, and colleagues in 1949, wherein electrical stimulation of MRF in anesthetized cats immediately induced widespread desynchronization of the cortical electroencephalogram (EEG), mirroring the classical patterns observed during

wakefulness and arousal (Moruzzi and Magoun, 1949). Subsequent work demonstrated that MRF stimulation could enhance perceptual performance in awake animals (Fuster and Uyeda, 1962; Lindsley, 1958), while lesions encompassing the MRF resulted in subjects that remained alive for weeks but had profound and persistent states of unresponsiveness, resembling coma (Peterson et al., 1949). These findings led to the influential concept of the ascending reticular activating system (ARAS), proposing the MRF as a critical center for maintaining consciousness and regulating cortical tone. The central positioning of MRF, receiving a convergence of diverse inputs, coupled with its apparent lack of direct connections to the periphery and its profound effects on global brain state, fueled speculation about MRF subserving higher-order integration and even consciousness. Many studies on MRF's integrative function followed, revealing responses to a diversity of sensory stimuli, motor patterns, and behavioral states (Bell et al., 1964; Büttner-Ennever and Büttner, 1978; Scheibel et al., 1955; Steriade and Glenn, 1982).

The decline in MRF research began in the 1960's and 70's, when the discovery and characterization of widely-projecting neuromodulatory systems originating in the brainstem was made. In particular, the identification of the locus coeruleus (LC) (Dahlstrom, 1964), located just caudal to the MRF, and its neurons releasing norepinephrine (NE) to effect widespread changes in arousal and cortical activity (Sara, 2009), provided an alternative explanation for the ARAS phenomena. Indeed, NE-releasing axons from LC course directly through the center of MRF (Oh et al., 2014). The observed effects of MRF stimulation in desynchronizing cortical EEG and MRF lesions leading to loss of consciousness may therefore have been attributable to

the stimulation and lesion of these NE axons. It has also been shown, however, that cortical desynchronization as a result of MRF stimulation is dependent on cholinergic transmission in the substantia innominata, to which the MRF has been shown to directly project (Müller et al., 1993). Nonetheless, by the 1980s, discussions of the ARAS focused almost exclusively on monoaminergic and cholinergic neuromodulatory systems, no longer even mentioning potential roles for the glutamatergic and GABAergic MRF neurons (McCormick, 1989). The scant research on neurons in MRF since this period have focused on motor functions, which we briefly review next. Despite the waning enthusiasm of the field, the early results about diverse coding properties of MRF and strikingly integrative connectivity still stand, and they form part of the basis for the new hypothesis of MRF function we lay out in **Chapter 1.3**.

1.2.3 Established motor functions in MRF

A line of research in the non-human primate saccade literature has found a role for a subregion of the MRF, the “central” MRF or cMRF, in eye movements and gaze control. As it is difficult to define boundaries based on the diffuse and poorly defined cytoarchitecture of MRF, cMRF was defined physiologically as the region of MRF in which elicited horizontal saccades (Bohlen et al., 2022; Cohen et al., 1985; Warren et al., 2008). cMRF comprises a small part of the total volume of MRF, and while exact anatomical boundaries are not described, it roughly sits dorsolateral to the red nucleus and ventrolateral to the interstitial nucleus of Cajal (Chen and May, 2000). cMRF has reciprocal connections with SC, and together cMRF and SC are proposed to work together to control saccades, with SC selecting saccade targets and cMRF generating

the saccade metrics necessary to drive oculomotor muscles appropriately to the target (Cromer and Waitzman, 2006; Graf and Ugolini, 2006; Waitzman et al., 2002, 2002). Relatedly, there were a small set of studies using electrophysiological recordings to measure activity in SC related to proximal limb reaching movements which additionally recorded from MRF neurons below, finding similar patterns of activity in SC, MRF, and the muscles (Stuphorn et al., 1999; Werner et al., 1997).

A broader field of research has characterized a region of the midbrain responsible for locomotion, the mesencephalic locomotor region (MLR), a small part of which overlaps with MRF. Similar to cMRF, the MLR is a physiologically defined region, first identified in cats when researchers observed that continuous electrical stimulation of this part of the midbrain evoked persistent locomotion (Shik et al., 1969). The MLR is primarily composed of glutamatergic neurons in the pedunculopontine and cuneiform nuclei, with a small part of it overlapping with the adjacent caudal-most aspect of MRF. Recent work has separated MLR into distinct subpopulations segregated by axonal target, correlated with movements beyond just locomotion (Ferreira-Pinto et al., 2021; Roseberry et al., 2016).

Interestingly, a recent 2022 study identified a population of neurons mostly confined to the MLR, within the MRF and PPN specifically, that play a specific role in release cue-triggered actions. The MRF/PPN population was defined by its projections to the parts of thalamus that project to ALM. Using the delayed-response directional licking paradigm described in **Chapter 1.1.2**, stimulation of these MRF/PPN neurons was found to mimic presentation of a go cue in its behavioral effect and in changes to frontal cortical dynamics; that is, transforming planning-related cortical activity to a

motor command, which in turn drove the task-appropriate licking movement in the response period.

In sum, there are specific subregions encompassing relatively small areas of the caudal MRF that play roles in saccades, locomotion, and release cue-triggered movements. The precise anatomical boundaries and overlap of these subregions remains unclear. Outside of these restricted subregions, the vast majority of the space within MRF has not been functionally characterized. Given the vast progress in recent decades in characterizing the roles of numerous brain regions in a wide variety of behaviors, we believe that the reason this large swath of MRF remains uncharacterized is not due to a lack of investigation. Rather, it is consistent with a more general role in computing contextual state related to flexible behavior, a role more difficult to elucidate with traditional tools as opposed to the highly specific innate functions observed in well-defined areas such as hypothalamic nuclei.

1.3 Main objectives: Re-evaluating MRF function

In this dissertation, we test the hypothesis of a novel function for MRF in setting abstract contextual state to subserve flexible behavior in visual decision-making. The hypothesis is based on two key MRF attributes:

1. The MRF's connectivity positions it as an integrative node within the recurrent loops subserving flexible decision-making (reviewed in **Chapter 1.2.1**, summarized in Fig. 3.1B). Parallel streams of the cortex and basal ganglia all converge in MRF via SNr collaterals, thereby allowing MRF to receive information about all of the relevant motor, sensory, and internal state variables

that contribute to defining context. The MRF additionally receives direct projections from prefrontal and motor cortical areas and projects back to them through the matching parts of thalamus. Similarly, the MRF reciprocally connects to the zones of SC which receive input from prefrontal and motor cortical areas.

2. The MRF contains remarkably similar activity patterns as canonical decision-related regions in perceptual decision-making tasks: sensory stimuli, action, choice, and modulation by prior expectation (reviewed in **Chapter 1.1.5**). Evolving views on the circuitry underlying flexible behavior suggest that this circuitry is much more distributed than once thought. Having traditionally been considered purely the domain of cortex, understanding shifted to recognize similar involvement of the SC and basal ganglia, and has now arrived at the current view of even broader involvement spurred by the finding of brain-wide correlates of task variables.

In the remainder of this dissertation, we present a set of experiments to test the hypothesis that MRF computes abstract contextual information during flexible visual decision-making. We first designed a novel context-dependent stimulus-action remapping task, and using behavioral analyses and modeling, we show that mice use abstract task rules to update an internal context belief variable that subserves the required flexible behavior (**Chapter 2**). We then densely record from MRF and other established nodes of the flexible visual decision-making circuitry. MRF neurons, in a network with MOs, CP, and SC, maintained a stable baseline representation of context. This baseline context representation formed a dimension in state space orthogonal to the choice dimensions, putatively setting action attractor dynamics to enable

stimulus-action remapping between contexts (**Chapter 3**). Recordings from task-naive and trained mice during passive stimulus presentation showed that MRF also undergoes long-term contextual adaptations, with MRF neurons in trained mice containing task-specific visual stimulus responses (**Chapter 4**). In **Chapter 5**, we begin to explore structure-function relationships in MRF and find a significant alignment between baseline context coding and cortical inputs which is particularly enriched in the anterior aspect of MRF. Finally, in **Chapter 6**, we discuss updated theories of MRF function and organization based on our findings as well as the multitude of future directions to elucidate the MRF's roles in the brain-wide circuits driving flexible mammalian behavior.

1.4 General methods

The methods described below apply to the experiments presented in **Chapters 2-5**. Methodology specific to individual chapters will be presented within the respective chapter.

1.4.1 Subjects

Subjects were 15 male and female C57BL/6 mice (n = 6 male, n = 9 female) single-housed on a 12 hr light/dark cycle and aged 22-68 weeks at the time of recordings. 7 of the mice were trained on the task (n = 3 male, n = 4 female), and behavioral data from each of these mice was used for behavioral modeling. 5 of these trained mice (n = 2 male, n = 3 female) were used for electrophysiological recordings during the task. 6 mice (n = 3 male, n = 3 female) were not trained on the behavioral

task and were used for electrophysiological recordings during passive stimulus mapping. 4 of the trained mice (n = 2 male, n = 2 female) and 2 additional mice (n = 2 female) that were trained on an early version of the behavioral task were used for passive stimulus mapping neural recording data only. All mice were the wild-type genotype (C57Bl6/J). Electrophysiology experiments were performed during the light cycle. Mice were given free access to food in their home cages for the duration of the experiment. Mice trained on the behavioral task were water restricted for the duration of the experiments and maintained at above 80% of their baseline weight (Guo et al., 2014a). Task-naive mice were water restricted for one week prior to and during the electrophysiology experiment period. All experimental procedures were performed in strict accordance with protocol 4461–01 approved by the Animal Care and Use Committee at the University of Washington. The sample sizes for task recordings (n = 5 mice, n = 36 recording sessions; n = 3378 neurons), task-naive passive stimulus mapping (n = 6 mice, n = 21 sessions, n = 1681 neurons), and trained passive stimulus mapping (n = 6 mice, n = 40 sessions, n = 4988 neurons) were not determined with a power analysis.

1.4.2 Surgical procedures

For all surgical procedures, mice were anesthetized under 3-4% isoflurane in an induction chamber first, then maintained at 1-2% isoflurane for the duration of the 1-2 hour procedure. Carprofen (5 mg/kg) was administered subcutaneously and lidocaine (2 mg/kg) was injected under the scalp for postoperative analgesia. The mouse was kept in a warmed recovery chamber until recovery from anesthesia before being returned to

its home cage. Mice were allowed to recover in the home cage for at least 5 days before habituation to head fixation and task training. Carprofen (0.05 mg/ml) was given for 3 days in water after surgery.

1.4.3 Surgeries for electrophysiology

The headbar implantation and craniotomy surgical preparations were performed as previously described (Ottenheimer et al., 2023; Steinmetz et al., 2019; Zatka-Haas et al., 2021). The scalp was shaved and further cleaned with hair removal cream before mice were transferred to a separate stereotaxic frame for surgery. Eye ointment (Alcon, Systane) was applied before surgery to prevent drying and body temperature was kept at 37°C with a far infrared warming pad (Kent Scientific). The skin and periosteum connective tissue was cleared off the dorsal surface of the skull. A layer of cyanoacrylate (Vetbond Tissue Adhesive) was then applied to seal the junction between the exposed skull and cut skin. A 3D-printed plastic recording chamber was implanted on top of the skull with dental cement (C&B Metabond), and a custom-made steel headplate was attached to the skull over the interparietal bone with Metabond for head fixation. To a protective layer over the skull, two separate thin layers of fast curing optical adhesive (NOA81, Thorlabs) were applied and cured until solidification with a 365 nm, 3 W output power UV Flashlight (LIGHTFE, UV301D).

After completion of task training, brief (15–30 min) surgery was conducted to perform 1 to 2 craniotomies for Neuropixels probe insertion. Briefly, following induction of anesthesia, small (2 × 1.5 mm (w × h)) craniotomies were made over areas of interest with a handheld dental drill. The craniotomies were covered with a soft silicone

gel (DOWSIL 3–4680) or Kwik-Cast, and the recording chamber was covered with a 3D-printed lid sealed with Kwik-Cast elastomer to protect craniotomies from debris.

1.4.4 Electrophysiological recordings

During recording sessions, mice were headfixed. Recordings were made using Neuropixels 2.4 electrode arrays (4 shanks, 5120 sites) (Steinmetz et al., 2021b), which have 384 selectable recording sites. Probes were mounted to a dovetail and affixed to a steel rod held by a micromanipulator (uMP-4, Sensapex Inc). For later electrode track localization within the brain, probes were coated with a red lipophilic fluorescent dye (CM-Dil, ThermoFisher Vybrant V22888) by holding 3 μ l in a droplet on the end of a micropipette and painting the probe shanks. Probe entry coordinates and angles were selected using Pinpoint software (Birman et al., 2023). Medial/lateral and anterior/posterior coordinates were zeroed at bregma, and the depth axis was zeroed at the surface of the brain at the craniotomy. In each session, one or two probes were advanced through the surface of the brain to their final position at 4-5 μ m/s. Probes were driven 100-150 μ m past the final depth, then withdrawn the same distance, to relieve tension on the brain and reduce motion of the brain relative to the probe. Electrodes were allowed to settle for around 10 min before starting recording. Recordings were made in internal reference mode using the ‘tip’ reference site on shank 0, with a 30 kHz sampling rate. Recordings were repeated at different locations on each of multiple subsequent days, with at least 100 microns of distance and/or 5 degrees of probe pitch between probe insertions to allow for reliable distinguishing of probe tracks post hoc. Additional craniotomies were performed as needed.

1.4.5 Histology and imaging for probe tracking

Animals were anesthetized with isoflurane. Mice were perfused intracardially with 0.9% phosphate buffer saline followed by 4% paraformaldehyde (PFA).

Post-hoc tracking of electrophysiological recording trajectories was performed as previously described (Ottenheimer et al., 2023; Ye et al., 2023). Brains were extracted immediately following perfusion and post-fixed in 4% paraformaldehyde for 24 h. In preparation for light sheet imaging, brains were cleared using organic solvents following the 3DISCO protocol (Ertürk et al., 2012) (<https://idisco.info/>), with some modification. Briefly, on day 1 brains were washed 3X in PBS and dehydrated in a series of increasing MeOH concentrations (20%, 40%, 60%, 80%, 100%, 100%; 1 hr each) then incubated overnight for lipid extraction in 66% dichloromethane (DCM) in MeOH. On day 2 brains were washed 2X twice in 100% MeOH for 1 hr each, then bleached overnight in 5% H₂O₂ in MeOH at 4°C. On day 3 brains were washed 2X in 100% MeOH, then final lipid extraction was accomplished in a series of DCM incubations (3 hr in 66% DCM in MeOH, 2X 100% DCM for 15 min each) before immersion in dibenzyl ether (DBE) for refractive index matching.

Brains were imaged on a light sheet microscope (LifeCanvas Technologies SmartSPIM) at least 3 days after completion of the iDISCO procedure. Brains were immersed in DBE in the imaging well secured in the horizontal position, and illuminated by one or two light sheets (100% width) imaging with the 1.6x objective (4x4 um resolution). Images were gathered from the dorsal surface of the brain to the ventral surface in 10 µm steps in 488 nm (autofluorescence, ~30% power) and 594 nm (Dil, ~10% power) excitation channels. The raw TIF images were compiled into a single

multi-image file with 10 μm voxels, then spatially downsampled to 25 μm voxels for transformation to the Allen CCF volume (Q. Wang et al., 2020) using the Elastix algorithm (Shamonin et al., 2014). CCF-transformed volumes were used to generate CCF fluorescent probe tract locations (pixel coordinates along the probe tract) using Lasagna (<https://github.com/SainsburyWellcomeCentre/lasagna>). Probe track CCF pixel coordinates (origin front, top, left) were transformed to bregma coordinates (origin bregma, x==ML, y==AP, and z==DV) in preparation for final integration with electrophysiology recordings using the International Brain Lab electrophysiology atlas GUI (<https://github.com/int-brain-lab/iblapps/tree/master/atlas electrophysiology>). For recording alignment, which was done per shank, sorted spike firing rates and amplitudes, a correlation matrix between channels, single neuron waveforms on each channel were displayed spatially in relation to the estimated channel locations in Atlas space from the tracked probe. The recording sites were then aligned to the Atlas by manually identifying a warping such that recording sites were best fit to the electrophysiological characteristics of the brain regions (e.g. matching location of ventricles or white matter tracts with low firing activity bands). This procedure has been estimated to have a 70 μm error (Liu et al., 2021; Steinmetz et al., 2019). Individual neuron locations were determined using the recording channel brain coordinates of each unit's maximum-amplitude waveform.

1.4.6 Neural data spike sorting and preprocessing

Neural recording data were automatically spike sorted with Kilosort2.5 (Steinmetz et al., 2021b) (<https://github.com/MouseLand/Kilosort>). Sorted units were curated using

automated quality control (International Brain Laboratory et al., 2024): exclusions were based on spike floor violations (the estimated proportion of spikes that were missed because they fell below the noise level of the recording), and sliding refractory period violations (the estimated proportion of spikes arising from the non-primary neuron). Quality control accuracy was assessed by manually reviewing a subset of the data using the phy GUI (<https://github.com/kwikteam/phy>). For population decoding analyses (Fig. 3.2, Fig. S2), we included multi-unit activity in population firing rates (only the spike floor metric and a 0.1 hz decoding period firing rate cutoff were applied).

1.4.7 Statistics

In all figures, * $p < 0.05$; ** $p < 0.01$, *** $p < 0.001$. Non-parametric tests were used unless otherwise stated (Wilcoxon-rank sum or Wilcoxon signed-rank for paired data). All statistical tests were two-sided. To calculate confidence intervals for proportions of encoding neurons (Fig. S4), we performed bootstrap resampling 1,000 times with an alpha of 0.05 using the 'bootci' function in MATLAB (bias-corrected and accelerated percentile method). An “above chance” effect is defined here as having its confidence interval bounds at a given alpha lie outside of the false discovery rate (FDR) for a given statistical test. For comparing regions in single neuron-based analyses, each neuron was considered a replicate (Fig. 3.3, 4.1, Fig. S4-6). For comparing regions in neural population-level analyses, each included session was considered a replicate (Fig. 3.2, Fig. S2).

For all analyses in which empirical null distributions were generated (Kernel regression, Fig. 3.3, Fig. S4; Population decoding, Fig. 3.2, Fig. S2-3; Spatial

distributions of features across cell type clusters, Fig. 5.1), p-values were defined as the percentile of the actual test statistic $T(X)$ within the null ensemble $\{T(X')\}$ such that if $T(X)$ is larger than M of N random shuffles, then $p=(1+N-M)/(1+N)$ (Harris and Miller, 2023). Effect sizes were reported as the difference between the test statistic $T(X)$ and the median of the null ensemble $\{T(X')\}$.

Chapter 2: Mice integrate abstract contextual rules with visual stimuli to remap actions in a flexible decision-making task

2.1 Introduction

To test the hypothesis that MRF processes abstract contextual information during flexible decision-making, we first sought to design a task which both required flexible behavior and allowed for the measurement of disentangled task variables in neural activity. In the past decade, several flexible perceptual decision-making tasks have been designed for rodents, examples of which we describe in depth in **Chapter 1.1.4** (Duan et al., 2021; Okazawa and Kiani, 2023; Rodgers and DeWeese, 2014; Tseng et al., 2022; Wu et al., 2020). Below, we will highlight key features which are shared and which differ amongst these tasks, followed by which features we chose for our task design and why we chose them.

1. Sensory stimuli. Tasks have used odor (Liu et al., 2014; Wu et al., 2020), auditory (Rodgers and DeWeese, 2014), visual (specifically in virtual reality navigational tasks) (Tseng et al., 2022), and a combination of auditory and visual (Duan et al., 2021, 2015) stimuli as both contextual cues and/or the perceptual cues upon which the choice is made. In the task presented here, we solely used visual stimuli (in the form of gabors) as perceptual cues. We did not use sensory stimuli as contextual cues (elaborated on next). We opted to use visual stimuli for two main reasons. First, it facilitates more direct comparison to the flexible decision-making circuitry which has been studied in the non-human primate literature, the vast majority of which utilized visual task paradigms (**Chapter 1.1.2 - 1.1.3**). Second, we sought to use a task design most analogous to the visual

decision-making tasks used in recent brainwide surveys (International Brain Laboratory et al., 2023b; Steinmetz et al., 2019), which were the first studies to find diverse correlates of task activity in MRF.

2. Context-dependence. Flexible perceptual decision-making tasks always require some form of context-dependence to remap actions. Context can either be explicitly cued by a sensory stimulus on each trial, as in DMS and Pro/Anti tasks (Duan et al., 2021, 2015; Liu et al., 2014; Wu et al., 2020), or can be uncued on each trial, instead requiring subjects to infer context on the basis of trial history (e.g. action-reward outcomes) (Rodgers and DeWeese, 2014; Tseng et al., 2022). Uncued context typically switches several times over the course of a session in blocks of trials to allow subjects to integrate over a long enough trial history to update their context belief. In contrast, an advantage of tasks with cued context is that they can rapidly switch contexts from trial-to-trial. Our task uses uncued context, and is distinct from the cited tasks in that context can be inferred from aspects of the trial history other than action-reward outcomes (described below in **Chapter 2.3**). An advantage of *not* explicitly cueing context with a sensory stimulus on each trial is that is much easier to unambiguously measure the coding of contextual variables in neural activity; i.e., delineating whether neural responses following contextual cues are related to coding of the abstract context is confounded by processing of the sensory stimulus itself.
3. Head-fixed vs. freely moving. While freely moving tasks are easier to train due to being closer to naturalistic behavior, head-fixed tasks contain many experimental advantages including precise control of sensory stimuli and the ability to perform

many acute recordings in the same subject (Carandini and Churchland, 2013). Several of the aforementioned rodent flexible-decision making tasks used head-fixed preparations (Liu et al., 2014; Tseng et al., 2022; Wu et al., 2020), and advancements in head-fixed visual psychophysics for mice have been made in recent years (Burgess et al., 2017; International Brain Laboratory et al., 2024). We thus used a head-fixed preparation for our task.

4. Reaction time vs. delayed response. Many rodent perceptual decision-making tasks utilize delay periods—an imposed period of time in between stimulus presentation and response epochs—as a way to assess working memory and its interactions with motor planning (Chen et al., 2024; Guo et al., 2014b; Inagaki et al., 2019; Li et al., 2016; Svoboda and Li, 2018). Delay period tasks can be more difficult to train across species than reaction-time based tasks, all else equal, because they impose an additional demand of inhibitory control (Luhmann, 2009). Here, we used a reaction-time based task (i.e., subjects were free to respond upon presentation of the stimulus), as our primary focus was not on working memory or motor planning. As we will show in **Chapter 3**, the contribution of stimulus vs. premotor processing in neuronal responses can be assessed by leveraging variation in reaction times between trials.

Briefly, the task we developed, described in much greater detail in the following two sections, was a flexible visual stimulus-action remapping task. The task required mice to respond to visual stimuli via wheel turns (Burgess et al., 2017; Steinmetz et al., 2019), but crucially, the rewarded action for an *identical* center stimulus (i.e. a stimulus directly in front of the mouse, an azimuth of 0°) depended entirely on an uncued,

block-wise context. This context could be inferred, in part, from the identity of unambiguous “outside” stimuli specific to each block (to the left or right of the center stimulus, i.e. azimuths of -80° or $+80^\circ$), which required the opposite action as the center stimulus, as well as the rewarded action on center stimulus trials. Thus, mice had to integrate an abstract rule-dependent trial history to flexibly update their actions on identical visual stimuli.

There is a wealth of literature on reversal learning paradigms, particularly in the field of value-based decision-making, often implicating dopamine-dependent circuits most prominently involving mPFC, OFC and striatum (Izquierdo et al., 2017; Wilson et al., 2014). While our task shares some similarities with these paradigms, it differs in both the cognitive processes it engages and thus the specific variables it measures in three key ways. First, reversal learning often investigates the adaptation to shifts in stimulus-outcome or action-outcome values, frequently unfolding over longer timescales that may span multiple sessions, and engaging mechanisms of associative relearning and synaptic plasticity linked to value updating (Ghosh and Zador, 2021; Hervig et al., 2020; Huda et al., 2020; Izquierdo et al., 2017). Our task, in contrast, features rapid, uncued reversals of context occurring multiple times within a single session (40-70 trials per block). This design promotes the dynamic tracking of a hidden state (the current task rule or context) rather than focusing primarily on slower associative relearning processes based on accumulated reward history. Second, in reversal learning paradigms where reversals do dynamically occur within a session, paradigms most often assign a higher value to one action over another across blocks by either probabilistically (e.g., dynamic two-armed bandit task) or deterministically switching

reward outcomes associated with each action (Bari et al., 2019; Kim et al., 2009; Le et al., 2023). This structure is designed to evaluate value learning, and it allows subjects to potentially adopt simpler strategies like developing a block-dependent action bias (e.g., favoring left turns in one block, right turns in another), rather than needing to perform a context-dependent mapping of specific stimuli to actions. Our task specifically precludes such motor bias strategies by balancing the overall reward probability for both clockwise and counterclockwise wheel turn actions across blocks, necessitating a true remapping for the central stimulus. Third, unlike pure reversal learning tasks, our task includes implicit contextual cues unique to each block—the “outside” stimuli. As the outside stimuli require actions opposite to that of the center stimulus, use of them to update the inferred context would indicate the subject’s use of an abstract task rule, rather than merely updating values based on direct action reinforcement. The above concepts are related to theoretical work about model-free vs. model-based systems in reinforcement learning (Daw et al., 2005), which we touch on in the discussion section for this chapter.

2.2 Methods

The methods below consist of technical details about the task structure, training protocol, and behavioral models fit to subjects’ choices in the task. Context and rationale for use of these methods are extensively described in the results section (Chapter 2.3).

2.2.1 Task structure

The behavioral task uses the same rig hardware and software as described

previously (Steinmetz et al., 2019). Mice sat on a plastic apparatus with forepaws on a rotating wheel surrounded by three computer screens (Adafruit, LP097QX1) at right angles covering 270×70 degrees of visual angle (d.v.a.). Each screen was roughly 11 cm from the mouse's eyes at its nearest point and refreshed at 60 Hz. The screens were fitted with Fresnel lenses (Wuxi Bohai Optics, BHPA220-2-5) to ameliorate reductions in luminance and contrast at larger viewing angles near their edges, and these lenses were coated with scattering window film ('frostbite', The Window Film Company) to reduce reflections. The wheel was a ridged rubber Lego wheel affixed to a rotary encoder (Kubler 05.2400.1122.0360). A plastic tube for delivery of water rewards was placed near the subject's mouth. Experiments were run with Rigbox software for MATLAB (Mathworks, Inc.).

To reduce baseline period movement and discourage the subject from choosing actions before stimulus onset, trials were only initiated after the subject had held the wheel still for a short interval (duration exponentially distributed between 0.2–0.5 s on each trial; Fig. 2.1E). At trial initiation, a visual stimulus was present on either the left screen (-80° azimuth), the center screen directly in front of the mouse (0°), or on the right screen (80°). All stimuli were present at 0° altitude. As this task was designed to assess context-dependent abstract rule belief rather than perceptual discrimination, stimuli were designed to be highly salient and unambiguous: stimuli were full contrast square wave Gabor patches with orientation 45° , sigma 7 degrees of visual angle (d.v.a.), and spatial frequency 0.1 cycles per degree. The patches were flipping gratings with a randomized initial phase and flipping frequency of 2 Hz. Coincident with stimulus onset was a go cue (10 kHz pure tone, 0.1s duration), after which movement of the

wheel was coupled to movement of the visual stimulus on the screen, and mice were free to move the stimulus to the target location for a reward. Wheel turns in which the top surface of the wheel was moved to the subject's right (clockwise turns) led to rightward movements of stimuli on the screen, that is, a stimulus on the subject's left moved towards the central screen. A CW or CCW turn was registered when the wheel was turned by an amount sufficient to move the visual stimuli by 40 d.v.a. in either direction (i.e., the distance to the intermediate position between the two possible stimuli of a given Block). If the wheel turn was in the correct direction (CW for left stimulus/center stimulus right Block; CCW for right stimulus/center stimulus left Block), mice received a sucrose reward (with 25 ms delay) which was coincident with a reward tone (15 kHz, 0.1s duration). If the wheel turn was in the incorrect direction (miss trial), a white noise burst was played (1.5 s duration). If the mouse did not turn the wheel sufficiently in either direction within the response window (10 s), the trial counted as a miss and a white noise burst was played. To get the precise times of wheel movement onsets post-hoc, an algorithm was used to identify the earliest moment at which the wheel began detectably moving, described previously (Steinmetz et al., 2019).

The Block structure of the task was predetermined by random draws from a uniform distribution between 40 and 70. Block lengths were only considered on non-repeat trials, i.e., the mouse had to complete the predetermined number of non-repeat trials (between 40 and 70) for the Block to switch. To encourage updating the center stimulus rule, miss trials were repeated until the correct action was chosen. Repeat trials were not included in the analysis of neural data, as subjects could potentially use a different rule without integrating context to choose the correct action on

these trials (i.e., choose the opposite action from the previous action). Mice performed the task until disengaged ($\geq 4\times$ reaction time difference from baseline) anywhere between 350 to 1000 trials. In each Block, there is an equal probability of each of the two stimuli appearing (i.e., 50% left and 50% center stimuli in the left Block, 50% right and 50% center stimuli in the right Block), thereby preventing subjects from successfully performing the task using a Block-dependent motor bias strategy.

2.2.2 Task training protocol

Mice were trained on this task in 3 stages. The first stage was intended to have mice learn to turn the wheel for reward and associate stimulus positions with wheel turn directions (left, CW, right, CCW. Stimuli were presented at either -40° or 40° and had to be turned CW or CCW for reward respectively. Once mice achieved high accuracy and initiated movements rapidly ($\geq 80\%$ performance on non-repeat trials, and with average reaction times at 500 ms or lower, the mice progressed to stage two (at experiment's discretion). In stage two, both stimulus locations were gradually shifted to the right or left by the same amount as a function of the mouse's performance. Stimuli were shifted in 6.67 d.v.a increments from the starting positions of -40° and 40° until they reached either 0° and 80° (full rightward shift) or -80° and 0° (full leftward shift). Stimuli were shifted based on windowed performance of the mouse on non-repeat trials: 75% accuracy had to be achieved on both stimulus positions for the shift to occur. The performance window started at 200 trials was decreased by the experimenter as mice more quickly improved their performance on different stimulus positions, down to a 50 trial window at the lowest. The performance window at the right and left extremes was

always 3 times larger than the window for the intermediate positions (i.e., the window was 150 trials long for right and left extremes when the parameter was set to 50 trials). Once mice consistently shifted between the left and right stimulus position extremes in single training sessions, they progressed to the third stage. As performance in stage two was typically very high in the intermediate positions (>90%), but lower for the center (0°) stimuli in the left and right extremes, the third stage quickly progressed through the intermediate positions, starting at a fixed length of 10 trials, while staying at the extreme positions for 200 or fewer trials. The intermediate stimulus position window was gradually decreased to 0 over the span of about 2 weeks such that only the left and right extremes remained (i.e. the left and right Blocks of the task). At this point, the number of trials it took for mice to improve their performance on center stimuli was monitored, and the length of the Blocks was decreased accordingly by the experimenter, until mice were able to successfully switch between left and right Blocks with >70% performance on center stimuli with Block lengths of 50-60 trials. Mice were then ready for the final stage of the task used in neural recordings, where Block lengths were randomized between 40-70 trials.

Throughout the course of training, the following additional parameters were tuned at the experimenter's discretion to aid in learning of the task: wheel gain, balance of instructed CW or CCW trials (changed only temporarily when mice had a strong CW or CCW bias; never present during recording sessions), and/or inter trial interval (defaults were 1.5 s after hit, 3 s after miss; changed anywhere from 1 to 4 s). The total length of training from day one to neural recordings took between eight and twenty weeks. Roughly 40% of mice trained on the task made it to the final stage, with most mice

failing during stage two. Two of the mice included in behavioral plots (Fig. 2.1F,H-J, Fig. S1C-F) who were proficient in the task did not have any electrophysiology recordings performed on them.

2.2.3 Session inclusion criteria

For both neural and behavioral modeling analyses, we sought to only analyze sessions wherein the subject was sufficiently tracking context to successfully remap center stimulus actions. Sessions also needed to have enough trials for adequately powered statistical analysis. Thus, session inclusion criteria were as follows: ≥ 200 non-repeat trials, 50-trial sliding window of performance at some point crossing 60% correct on non-repeat center stimulus trials and 80% correct on non-repeat outside stimulus trials for both left and right Blocks. Sessions included for behavioral modeling include those during the final stage of task training prior to neuropixels recording as well as recording sessions. Trials with no wheel turn response were excluded from analysis. Behavioral models were trained and evaluated on both repeat and non-repeat trials. Raw behavioral performance was always evaluated on non-repeat trials. For analyses of neural data during performance of the task, we sought to only analyze non-repeat trials wherein the mouse didn't have flinch nor overly delayed wheel turn responses. Thus, trials with a reaction time between 75 and 400 ms were included for further analysis.

2.2.4 Disentangled RNN behavioral model

The following behavioral models (Chapter 2.2.4-2.2.6) were designed and fully implemented by Jeremy N. Schroeter, with J. Shaker working in collaboration to contribute to the broader ideas and design.

We implemented disentangled recurrent neural networks (DisRNN) as in (Miller et al., 2023), with minor modifications (Fig. S1A-F). Briefly, the DisRNN is a recurrent neural network architecture designed to discover sparse and interpretable cognitive models directly from data. The DisRNN achieves this in two ways: (1) the model is trained to minimize both a supervised loss, and a complexity penalty which encourages parsimony in the learned model; (2) each element of the models latent state is updated through a separate feed-forward network. At each timestep the model receives as input the current latent state $\mathbf{z}_t$ as well as the observation of the previous trial $\mathbf{o}_t = [s_t, a_t, r_t]$ where s_t , a_t , and r_t are the stimulus (-1, 0, 1), choice (-1, 1), and reward (0, 1) received on trial t respectively. The elements of these vectors are passed in parallel through noisy channels referred to as information bottlenecks. Given input x , the outputs of these bottlenecks are samples $\hat{x} \sim \text{Normal}(mx, \sigma)$ where m and σ are learned parameters. Each bottleneck b receives a penalty for deviating from the unit Normal:

$$L_b = \sum_t D_{KL}(\mathcal{N}(m_b x_{b,t}, \sigma_b) \parallel \mathcal{N}(0, 1)) = \frac{1}{2}((m_b x_{b,t})^2 + \sigma_b - \log \sigma_b - 1)$$

where D_{KL} is the Kullback-Leiber divergence (KLD). For a given bottleneck, this penalty is exactly 0 when $m = 0$ and $\sigma = 1$. In that case the bottleneck's output will be samples from the unit Gaussian, independent of the input, and no information will pass through the bottleneck. Samples from these bottlenecks are then passed through parallel

multilayer perceptrons (MLP). Each MLP is responsible for updating a single element of the latent state by outputting an update value w_t^i and candidate element u_t^i . These are combined to update the latent state like so:

$$\begin{aligned} \begin{bmatrix} w_t^i \\ u_t^i \end{bmatrix} &= \text{MLP}^i \left(\begin{bmatrix} \hat{\mathbf{z}}_t \\ \hat{\mathbf{o}}_t \end{bmatrix} \right) \\ z_{t+1}^i &= (1 - w_t^i) \hat{z}_t^i + w_t^i u_t^i \\ \mathbf{z}_{t+1} &= (1 - \mathbf{w}_t) \odot \hat{\mathbf{z}}_t + \mathbf{w}_t \odot \mathbf{u}_t \end{aligned}$$

The elements of this latent state are then passed through a second set of “latent” bottlenecks to encourage sparsity in the latents themselves. The state is then concatenated with the stimulus $\mathbf{s}_{t+1}$, and passed through a final “choice” MLP which outputs the log-odds of a counter-clockwise choice on the next trial given the stimulus.

$$\log \frac{P(a_{t+1} = CCW)}{P(a_{t+1} = CW)} = \text{MLP}_{\text{choice}} \left(\begin{bmatrix} \tilde{\mathbf{z}}_{t+1} \\ \mathbf{s}_{t+1} \end{bmatrix} \right)$$

Where $\sim$ denotes that the state has gone through the latent bottleneck.

During fitting the model weights are adjusted to minimize the negative log-likelihood given the predicted choice probabilities and true behavior, as well as the complexity cost associated with the bottlenecks. After fitting, bottlenecks corresponding to elements of the latent state and observations which are necessary to predict future actions have parameters $m \approx 1$ and $\sigma < 1$, meaning that information is allowed to pass through. During evaluation, all bottlenecks are deterministic and output $m_b \mathbf{x}_t$.

We found that a 5 dimensional latent state was large enough to accommodate the solutions found across these values. The update MLPs all had 3 layers of width 5, and the choice MLP had 2 layers of width 3. The relative importance of accuracy or simplicity in the fitted model can be controlled through changing the weight β of the

bottleneck term.

$$L = \text{NLL}(a_t, P(a_t)) + \beta \sum_{b \in \text{bottlenecks}} L_b$$

For our hyperparameter sweep we found solutions for each subject at the following KL weights [0, 5e-4, 1e-3, 3e-3, 1e-2, 1]. We sought to optimize parsimony and performance, therefore, the model for each subject was chosen based on the highest KLD weighting before model performance dropped.

Model solutions were consistent across subjects, using a single latent variable that evolves according to a set of linear, observation-dependent updates. The linearity arises from the bottlenecks associated with $\mathbf{o}_t$ being the only ones left open after fitting (Fig. S1E). For example, if a single latent $\mathbf{z}^i$ was used after fitting with open update bottlenecks $\mathbf{z}^i$, s_t , a_t , r_t then the update rule for $\mathbf{z}^i$ would depend on both its current value and the trial observation, making its set of update rules potentially non-linear and observation dependent. However, after fitting, models consistently only had a single latent variable with open update bottlenecks s_t , a_t , r_t ; therefore, u_t^i and w_t^i are only dependent on the trial observation and each of the update rules are linear in $\mathbf{z}^i$.

2.2.5 Context-belief model

The context belief model was based on the model class discovered by the DisRNN; we describe this in detail in the results section below. We represent the subject's latent context-belief variable $\mathbf{z}_t$ as a linear function of its previous value, with parameters conditioned on the previous trials observation $\mathbf{o}_t$. Each trial observation $\mathbf{o}_t$ consists of the stimulus presented s_t , the action taken by the mouse a_t , and the reward received r_t .

$$z_{t+1} = a(\mathbf{o}_t)z_t + b(\mathbf{o}_t) \quad \mathbf{o}_t = \begin{bmatrix} s_t \\ a_t \\ r_t \end{bmatrix}$$

The predicted probability of taking a CCW wheel turn on trial t is given by the logistic function applied to a linear transformation of the latent variable.

$$P(a_t = CCW) = \sigma(\rho z_t + \delta)$$

Before fitting, model parameters governing the updates of the latent variable were initialized at the identity function. All model parameters were fit to minimize the negative log-likelihood (NLL) on center stimulus trials. In the equation below a_t is 0 for leftward wheel turns and 1 for rightward wheel turns.

$$\mathcal{L}_{\text{NLL}} = - \sum_{\text{sessions}} \sum_{t=1}^T 1(s_t = 0) \left[a_t \log P(a_t = CW) + (1-a_t) \log P(a_t = CCW) \right]$$

2.2.6 Center stimulus-specific Q-Learning and win-stay-lose-switch

The Q-learning agent (Rescorla and Wagner, 1972) consists of two variables $Q(a=CCW)$ and $Q(a=CW)$ which capture recent rewards associated with making counter-clockwise and clockwise wheel turns respectively, on center stimulus trials. Each Q-value is updated according to the following equation, but only when its associated action was selected

$$Q_{t+1}(a_t) = (1 - \alpha)Q_t(a_t) + \alpha r_t \quad Q_{t+1}(a \neq a_t) = Q_t(a \neq a_t)$$

The predicted probability of taking a counterclockwise wheel turn on trial t is given by the scaled difference between the two Q-values, passed through the logistic function.

$$P(a_t = CCW) = \sigma(\beta(Q(CCW) - Q(CW)))$$

Model parameters α and β were fit to minimize the negative log-likelihood on center

stimulus trials.

The win-stay-lose-switch (WSLS) strategy predicts subjects will react to an unrewarded action on a center stimulus trial by selecting the opposite action on the subsequent center stimulus trial. On rewarded center stimulus trials the strategy predicts subjects will repeat their rewarded action.

To cross-validate model performances for all of the above behavioral models, for each mouse we made 5 random partitions of sessions into training, validation, and testing sets, each containing approximately 70%, 15%, and 15% of a subject's sessions respectively. For each subject, the model with the best test set performance was chosen for all downstream analyses.

2.3 Results

To distinguish between the coding of cognitive variables, sensory stimuli, and motor-related processes, we designed a novel context-dependent flexible decision-making task. In this task, mice were required to integrate uncued context, changing in blocks of 40-70 trials, with visual stimulus location to choose the correct action for a reward. The task utilized a visual stimulus wheel-turning paradigm for mice as in previous studies (Burgess et al., 2017; International Brain Laboratory et al., 2023a; Steinmetz et al., 2019) (Fig. 2.1A). One full contrast stimulus appeared on the screens on a given trial, and there were two possible stimulus locations during each Block (6-15 Blocks per session). In a “Left Block”, the stimulus appeared either on the left screen (left stimulus, azimuth = -80°) or on the center screen (center stimulus, azimuth = 0°). In a “Right Block”, the stimulus appeared on the right screen (right stimulus, azimuth =

+80°) or on the center screen (center stimulus, azimuth = 0°). Transitions between Blocks occurred every 40-70 trials and were not cued to the subject. We refer to left and right stimulus (-80° and +80° azimuth) trials collectively as “outside stimulus” trials and the 0° azimuth trials across Blocks as “center stimulus” trials. In each Block, the relative position of the stimulus amongst the two possible locations indicated the correct wheel turn direction (left→CW, right→CCW; i.e, in a left Block, the outside stimulus is on the left and center stimulus is on the right, requiring CW and CCW turns, and vice-versa for the right Block) (Fig. 2.1B-E). The absolute position of center stimuli thus remains the same between Blocks, but the relative position and action contingency flips.

To adequately perform the task, mice were required to use abstract task rules and context. An important aspect of the task design is that outside and center stimuli were set to have an equal probability of appearing in each Block (Fig 2.1B): This means that the global probability of earning a reward from either action (CW vs. CCW) was equal and stable, i.e. not dependent on context. Thus, subjects could not simply employ a value-based action biasing strategy by integrating action-reward history, such as Q-learning (Rescorla and Wagner, 1972). The two requirements to achieve above chance performance on center stimulus trials were to accurately perceive the stimulus location and to integrate trial history dependent on contextual task rules. For example, a subject could leverage the task rule that presence of a preceding left stimulus indicates a CCW contingency on subsequent center stimulus (and vice-versa for the right stimulus) and thus use a strategy wherein center stimulus actions are updated based on previous outside stimulus identity. Similarly, a subject could switch center stimulus actions after hearing the miss noise (Fig. 2.1E) on a previous incorrect center stimulus

trial. Note that an ideal performer of the task would use a combination of these two strategies to deterministically switch their center stimulus actions, though the strategies could just as well involve probabilistic action updates in the non-ideal case. Another strategy, used alone or in combination with the former strategies, could be to update center stimulus actions by only integrating action-reward history on center stimulus trials. Each of these strategies fundamentally depend on encoding abstract task rules and context, whether using the identity of outside stimuli as direct evidence for the current Block rule or using center stimulus trials as a specific context to employ reinforcement learning-based updating of the Block rule.

Figure 2.1. Mice remap stimulus-action associations in a flexible decision-making task by updating an internal context belief variable. A) Schematic of task performance and Neuropixels recordings. Mice sat with forepaws on a wheel in front of three screens where visual stimuli were presented. One or two Neuropixels 2.0 probes were inserted into target brain regions during behavior. **B)** Flexible stimulus-action remapping task schematic. There are 2 Blocks (Left and Right), and within each Block, and 50/50 probability of two gabor stimulus locations: -80° or 0° in the Left Block, and $+80^\circ$ or 0° in the Right Block. Wheel must be turned CW or CCW such that the stimulus is moved to the middle of the two possible stimulus locations (i.e. a $\pm 40^\circ$ turn). Colors of arrows indicate trial type: Blue colors indicate CW wheel turns, and orange colors indicate CCW wheel turns; darker hues indicate the left Block, and

lighter hues indicate the right Block. **C)** Block trace for an example session, representing 500 trials. Block lengths were randomized between 40 and 70 trials. **D)** Trial structure across Blocks, showing randomized order and equal frequency of trials requiring CW and CCW responses. **E)** Timeline of a single trial. Trial start was uncued, and mice held the wheel still for 200-500 ms for the visual stimulus to appear. An auditory go cue played at the same time as the visual stimulus appearance. Mice were free to move the wheel upon presentation of the stimulus, and turns in the correct direction were rewarded with sucrose water drops, whereas turns in the incorrect direction resulted in a white noise burst plus timeout. **F)** Percentage CW turns for each trial type as a function of number of trials after Block switch. Data from 7 mice are plotted; percentage correct is compiled across all included sessions for a given mouse ($n = 37 \pm 6$ behavioral sessions per mouse). Traces are colored by trial type according to the convention in B). **G)** Example session latent context variable (z) from context belief behavioral model fit to center stimulus choices of each subject. Trial outcome and type are displayed above the latent trace. Inset: zoom-in on subset of trials with vectors representing the latent update depending on trial type/outcome and current latent value. Panel created by J. Schroeter. **H)** Context belief model accuracy in fitting each subject's center stimulus choices (% correctly predicted choices). Panel created by J. Schroeter. **I)** Contribution of different trial types to context belief model's performance and comparison with Q-learning and DisRNN. All Δ log-likelihoods are for fits to center stimulus choices. Model was constrained to freeze latent on outside stimulus, center stimulus unrewarded, or center stimulus rewarded trials, and deltas in performance are plotted for each subject. Panel created by J. Schroeter. **J)** Percentage correct on center stimulus trials after a Block switch, conditioned on whether the first trial post-switch was an outside or center stimulus.

Mice performed the task accurately, with consistent behavior across mice (mean $\pm$ standard error of the mean (SEM) $94 \pm 0.9\%$ accuracy outside stimulus trials, $79 \pm 1.8\%$ accuracy across final 5 center stimulus trials per Block, $n = 7$ mice, 257 sessions) (Fig. 2.1F). Performance on outside stimulus trials was stably above 90%. In contrast, performance on center stimulus trials after a Block switch started below chance, indicating use of the previous Block's contingency, and quickly increased to above chance (5 ± 0.5 trials to reach 60%), with saturating performance of $79 \pm 1.8\%$ toward the end of a Block (trials 20-25 after Block switch). This high level of performance precludes a reliance on purely stimulus-driven or value-based strategies and instead indicates that mice flexibly integrated sensory information with an internal representation of task context. To get a clearer understanding of how the mice construct their internal context representation, we next used behavioral modeling to fully characterize the strategies mice used to remap actions to center stimuli.

Behavioral modeling suggested that mice tracked a single latent variable representing belief about the current task context. To quantify the explanatory power of different strategies, we fit various behavioral models to predict subject choices. Given that mice did not have a step-switch in their actions on center stimuli after a Block switch (Fig. 2.1F), even after viewing an outside stimulus trial (Fig. 2.1J), we considered strategies which used integration of center stimulus action-outcome history. The simplest strategy is stimulus-specific win-stay, lose-switch (WSLS)(Evenden and Robbins, 1984) heuristic, wherein mice repeat their previous center stimulus trial choice if it was rewarded, or make the opposite choice if it was unrewarded. Similarly, mice could have used a variant of Q-learning (Rescorla and Wagner, 1972) that is stimulus-specific, wherein the variable is based on center stimulus action-reward associations over several past trials. Other strategies may incorporate additional task rules, such as using outside stimulus identity on previous trials to update center stimulus choices. To disambiguate the myriad of potential strategies in an unbiased manner, we leveraged disentangled recurrent neural networks (DisRNN)(Miller et al., 2023), which are designed to discover parsimonious and interpretable behavioral models directly from data (Fig. S1A-E). DisRNNs were separately fit to each subject's choice data and accurately predicted behavioral performance on held out sessions (NLL: 0.56 ± 0.009 , accuracy: $75.7\% \pm 0.3\%$). Crucially, the DisRNN's performance on each subject's choice data on center stimulus trials was higher than stimulus-specific WSLS and Q-learning models (Fig. S1F, $p=0.016$), indicating that these models do not fully capture the behavioral strategies mice use. Across subjects, the DisRNN consistently discovered a model which uses a single latent variable that evolves

according to a set of linear, observation-dependent updates (Fig. S1B,E; Chapter 2.2.4). The transformation of this variable used to predict future trial choices was also found to be approximately linear (Fig. S1D). This latent variable appeared to coevolve with the task Block (Fig. S1B), and only affected the prediction of center stimulus choices (Fig. S1D). Given these attributes of the latent variable, the model class discovered by DisRNN appeared to capture the mice's context belief.

Mice updated their context belief using a combination of center stimulus action-reward history and outside stimulus identity. To better understand the latent variable discovered by DisRNN, we directly instantiated its model class and found it to have similar performance to DisRNN (Fig. 2.1G-I, $p=0.578$). We modeled the subject's internal context belief z_t at each trial t as a linear function of its value on the previous trial, with function parameters (a and b) indexed by the previous trial observation $\mathbf{o}_t$. Each trial observation consists of the stimulus presented s_t , the action taken by the subject a_t , and the reward received r_t ; thus, $\mathbf{o}_t$ can take on 8 unique configurations. The log-odds of the subject making a CCW choice on a center stimulus trial was modeled as a linear function of z_t whose parameters (ρ and δ) are fixed across trials.

$$z_{t+1} = a(\mathbf{o}_t)z_t + b(\mathbf{o}_t) \quad \mathbf{o}_t = \begin{bmatrix} s_t \\ a_t \\ r_t \end{bmatrix} \quad \log \frac{P(a_t = CCW)}{P(a_t = CW)} = \rho z_t + \delta$$

To determine the relative importance of different trial types for context integration following a Block switch, we evaluated model performance on held out sessions whilst fixing z_t on a subset of trial types. This analysis revealed that for all mice, integration of center stimulus trial history had the greatest impact on accurate prediction of center stimulus choices, with rewarded center stimulus trial history specifically contributing to

most of the effect (Fig. 2.1I, mean $\Delta = -0.112, -0.091, p=0.016, 0.016$ signed-rank test). Integration of unrewarded center stimulus trials also contributed to performance, but to a lesser degree (Fig. 2.1I, mean $\Delta=-0.059, p=0.016$). Importantly, inclusion of outside stimulus trial observations significantly improved prediction of center stimulus trial choices (Fig. 2.1I, mean $\Delta=-0.084, p=0.016$), despite the fact that they require an action opposite to that of center stimulus trials for a reward. The improvement was almost equally as large as that of rewarded center stimulus trial history. Corroborating this finding, mice improved center stimulus performance after a Block switch more quickly if the first trial was an outside stimulus (Fig. 2.1J, $\Delta=14.6\%$ improvement in performance, $p=0.016$). These findings suggest that mice update their context belief with an additional abstract task rule—the presence of previous left or right outside stimuli indicate switching of the center stimulus contingency. Altogether, our behavioral modeling indicates that mice solve the stimulus-action remapping in this task by maintaining an internal context belief variable which updates based on abstract rule-dependent trial history. The latent context variable of the model is used in the rest of the study as a representation of the subject's internal belief about context (simply referred to as “context” hereafter). This variable, rather than the veridical Block identity trace, serves as the regressor or target variable for encoding and decoding of context in neural activity in **Chapters 3 and 5**, unless otherwise stated.

2.4 Discussion

In this chapter, we introduced a novel flexible stimulus-action remapping task designed to probe the neural mechanisms underlying abstract contextual processing in

mice. We demonstrated that mice successfully learn this challenging task, achieving near perfect accuracy on the unambiguous outside stimulus trials and high accuracy via flexible, adaptive rule-switching on center stimulus trials following uncued context reversals (Fig. 2.1F). As the task structure precludes sole reliance on simple stimulus-driven responses or global action biases to achieve above chance performance on center stimulus trials, we conclude that mice learned to use abstract task rules and context to achieve their level of performance in the task. We leveraged recent advances in unbiased cognitive model discovery via DisRNNs (Miller et al., 2023) to disambiguate the many possible strategies mice could have used in this novel task (Fig. S1). The highest-performing and most parsimonious model revealed that mice internally tracked a single latent variable representing their belief about the current task context or rule. This context belief is updated using a combination of stimulus-specific action-outcome history (specifically for the center stimulus) and, crucially, information gleaned from the identity of outside stimuli, indicating an understanding and utilization of the abstract task structure (Fig. 2.1G-J).

The reliance of subjects performing this task on inferring a hidden state (context) and flexibly applying abstract rules aligns with theoretical distinctions between model-based and model-free reinforcement learning systems (Daw et al., 2005). The task's structure, particularly the ambiguity of the center stimulus and the globally balanced action values, strongly discourages simple model-free strategies using cached values and necessitates model-based-like computations that leverage an internal representation of the task rules and current context.

This paradigm is one of the first successful implementations of a head-fixed,

context-dependent visual decision-making task in mice that necessitates the inference and application of abstract task rules for flexible sensorimotor remapping, without relying on explicit context cues (Biró et al., 2019). While dynamic visual navigation-based decision-making tasks requiring context inference exist, the cognitive and sensorimotor demands of navigating a virtual environment differ from the discrete stimulus-action mapping required here, likely engaging distinct underlying circuitry (Arlt et al., 2022; Kira et al., 2023; Tseng et al., 2022). Our task's structure provides a focused means to investigate the neural representation of abstract rules and context variables divorced from sensory or premotor processing.

A potential confound in interpreting neural activity during decision-making tasks is the widespread influence of uninstructed body movements, especially those aligned to task events (Musall et al., 2019a; Stringer et al., 2019). Attempting to mitigate this confound, we designed the task to minimize systematic behavioral differences between Blocks: The visual stimulus attributes, the required motor actions (CW/CCW turns), and the overall probability of executing either action are balanced across Blocks (Fig. 2.1A-C). As we will demonstrate quantitatively in **Chapter 3** using simultaneous video analysis, the task does not appear to elicit systematic, context-dependent differences in spontaneous movements during key task epochs, thereby strengthening the interpretation of context-related neural signals.

The primary limitation of this task is its relative difficulty to train compared to simpler discrimination or value-based paradigms, with approximately 40% of mice reaching full proficiency within a timeframe of eight to twenty weeks. The multi-stage training protocol likely facilitates learning by first establishing basic stimulus-action

associations, then forcing generalization of a relative spatial rule through the incremental stimulus position shifting stage, and finally introducing the block structure incrementally. Gradually shifting stimulus positions might prevent fixation on absolute location-response mappings, while the incremental introduction of block structure and initially longer block lengths may allow mice to learn the contingency switching without excessive frustration. Nevertheless, future applications requiring larger cohorts, such as brain-wide mapping or causal manipulation experiments, would significantly benefit from optimizing the training pipeline, potentially leveraging high-throughput automated training systems which are becoming increasingly prevalent (Hao et al., 2021; International Brain Laboratory et al., 2021).

This study underscores the power of combining sophisticated behavioral design with quantitative behavioral modeling to dissect cognitive and behavioral strategies (Wilson and Collins, 2019). The behavioral modeling—particularly the systematic analysis of how trial histories of distinct task events contribute to the context-belief model discovered via DisRNN—was crucial for uncovering the mechanisms mice used to compute context belief. Mice used both the direct outcomes of their choices on center stimulus trials (stimulus-dependent action-outcome history) and the identity of outside stimulus trials providing evidence about the current rule (use of implicit contextual cues). This demonstrates a capacity to integrate several types of information according to the abstract structure of the task.

Finally, the latent context variable derived from the context belief model represents our most accurate and parsimonious estimate of the mice's internal belief state during the task. Therefore, throughout the subsequent chapters analyzing neural

activity (**Chapters 3 and 5**), this latent variable, rather than the experimenter-defined Block identity, will serve as the primary regressor or target variable when measuring neural representations of the task context.

Chapter 3: The MRF represents abstract context associated with stimulus-action remapping in a network with SC, CP, and MOs

3.1 Introduction

The MRF has emerged as a compelling candidate for contributing to the computations underlying flexible visual decisions. Indeed, it has extensive connectivity with the relevant circuitry—including prefrontal and motor cortical inputs (Gao et al., 2022; Oh et al., 2014; Peng et al., 2021), converging inputs from parallel basal ganglia streams through SNr (McElvain et al., 2021), reciprocal connections with SC (Benavidez et al., 2021), and outputs to integrative and frontal cortex-projecting thalamic nuclei (Kolmac et al., 1998; Kolmac and Mitrofanis, 1998; Oh et al., 2014)—and has consistently been found to have robust task-related correlates in recent brain-wide surveys of perceptual decision-making tasks (Chen et al., 2024; International Brain Laboratory et al., 2023a; Khilkevich et al., 2024; Steinmetz et al., 2019). To understand the role MRF may play within the broader flexible decision-making circuitry, we devised a large-scale electrophysiological recording strategy during performance of the task presented in **Chapter 2** with two main objectives. The first was to densely tile the MRF. As discussed in **Chapter 1.2**, the MRF is a large, poorly differentiated, and diverse region; and most of the space within it has not been functionally characterized. By using 4-shank Neuropixels 2.0 probes (Steinmetz et al., 2021a) to record from the whole span of MRF, we were able to assess functional heterogeneity across the region and ensure robust sampling of potential neural populations of interest. Our second recording objective was to probe established nodes in the flexible decision-making circuitry. We targeted regions known to be involved in visually-guided decisions and flexible behavior,

including MOs, SCm, and CP (Duan et al., 2021; International Brain Laboratory et al., 2023a; Steinmetz et al., 2019; T.-Y. Wang et al., 2020; Wu et al., 2020). Recordings from these areas allowed us to have positive controls—i.e., areas in which we expect to see context coding—and to compare potential differences in the coding properties of each region which may reflect different computational roles in the behavior.

We use two main approaches to analyze the neural data from the flexible stimulus-action remapping task recordings. The first quantifies the contributions of task variables to single neuron firing rates in response to task events, using a widespread approach in systems neuroscience based on kernel fitting with generalized linear models (GLM) (Fig. 3.3) (Musall et al., 2023; Park et al., 2014; Steinmetz et al., 2019). A key advantage of this approach is the ability to statistically disentangle the influence of multiple, potentially correlated variables on neural activity, providing insights into mixed selectivity at the single-neuron level. However, it has been increasingly recognized that sensory, motor, and cognitive computations in the brain arise not from isolated neurons but from the coordinated activity of large populations (Ebitz and Hayden, 2021; Saxena and Cunningham, 2019). Dimensionality-reduction based approaches in particular have been shown to provide tremendous insight into how neural systems compute during active sensation (Nogueira et al., 2023), motor planning (Elsayed et al., 2016; Kaufman et al., 2014), and context-dependent selection and integration (Aoi et al., 2020; Mante et al., 2013). We thus use target dimensionality reduction to project the population activity from MRF and other recorded regions into a low-dimensional state space consisting of choice and context axes, revealing a mechanism of context-dependent shifting of action attractor dynamics (Fig. 3.2).

Finally, a critical consideration when interpreting neural activity during cognitive tasks is the potential confound of uninstructed movements. Mice have been shown to engage in spontaneous movements, which are often task-aligned when performing perceptual decision-making tasks, such as whisking, sniffing, or twitching of muscles. A strikingly large amount of the variance in neural activity can be explained by these movements (Musall et al., 2019a; Stringer et al., 2019). If these movements systematically co-vary with task variables, as was found in Musall et al. 2019, it can confound the attribution of neural coding to the task variable of interest (this phenomenon is related to a broader theory about the embodiment of cognition (Cisek, 2021b)). Our task design incorporates a feature intended to mitigate this issue, particularly for analyzing baseline activity: mice were required to hold the wheel still for a variable period before trial initiation (Fig. 2.1E). This of course does not limit the mice from performing uninstructed movements elsewhere on their body or face, however. In the methods and results below, we describe several ways in which address the potential confounds of uninstructed movements in our neural analyses, using high-speed videography of the face and pupil to track these movements.

3.2 Methods

The methods below consist of technical details about the models and statistical approaches used to analyze the neural activity recorded during task performance. Context and rationale for use of these methods are extensively described in the results section (**Chapter 3.3**).

3.2.1 Recording and quantification of face and pupil movements

For electrophysiology experiments, the subject was illuminated with infrared light (850 nm, CMVision IR30) and eye and face movements were monitored. The right eye was monitored with a camera (FLIR CM3-U3-13Y3M-CS) fitted with a zoom lens (Thorlabs MVL7000) and long-pass filter (Thorlabs FEL0750), recording at 70 fps. Face movements were monitored with another camera (FLIR CM3-U3-13Y3M-CS, zoom lens Thorlabs MVL16M23, long-pass filter Thorlabs FEL0750) directed at a 2×3 cm mirror reflecting the left side of the face, recording at 70 fps. Licks were detected from the face video by thresholding the average intensity of an ROI centered between the lips and the lick spout, calculated for every frame.

We measured pupil area and position (Fig. 2.1A, 3.3A, Fig. S3C) from the high-zoom videos of the subject's eye, using DeepLabCut (Mathis et al., 2018). In approximately 50 training frames randomly sampled from a subset of sessions, 4 points spaced at 90° around the pupil were manually identified, and the network was trained with default parameters. Frames were then manually annotated focusing on frames with errors, and the network re-trained. The pupil area was taken to be the area of an ellipse with major and minor axis lengths given by the distances between opposite pairs of detected points. Pupil position was defined as the center point between the dorsal/ventral and medial/lateral annotated points. The traces were passed through a median filter with a filter order of 5. Two recordings were excluded from these analyses owing to video quality that was unusable ($n = 2$ out of 36). We measured video motion energy SVDs (Fig. 2.1A, 3.3A, Fig. S3C) from the low-zoom videos of the frontal aspect of the subjects, which included the face, arms, and part of the torso of the mice. Motion

energy SVDs were calculated using Facemap (www.github.com/MouseLand/FaceMap), with the top 16 principal components (PCs) being used for analyses (Stringer et al., 2019). Both pupil area and face video SVDs were normalized using robust z-scoring before kernel regression fitting or decoding.

3.2.2 Neural data preprocessing

For visualizing neuron firing rates in peri-event time histograms (PETHs) aligned to task events (Fig. 3.1E-H), neural activity was binned at 0.015 s and smoothed with a causal half-Gaussian filter with standard deviation 0.05s. For the kernel regression analysis (Fig. 3.3, Fig. S4), each neuron's activity was binned at 0.01 s, smoothed with a causal half-Gaussian filter with standard deviation 0.025 s, and baseline soft normalized by dividing the trial spike rates by the baseline firing rate (-300 to 0 ms before stimulus onset) + 1 spikes/s. PETHs constructed for visualization included correct trials only, but both correct and error trials were included in the kernel regression analysis.

For the kernel regression analysis (Fig. 3.3, Fig. S4), to be discussed below, neurons were included in the analysis if their firing rate was significantly modulated during the task. Units were first excluded if their trial firing rate (analysis window between stimulus and movement onsets) was less than 0.1 Hz. Then, a set of six statistical tests was used to detect changes in activity during various task epochs and conditions: (1) Wilcoxon signed-rank test between trial firing rate and baseline rate (defined in period -0.3 to 0 s relative to stimulus onset on each trial); (2) signed-rank test between stimulus-driven rate (firing rate between 0 and 0.125 s after stimulus

onset) and baseline rate; (3) signed-rank test between pre-movement rates (–0.1 to 0 s relative to movement onset) and baseline rate; (4) Wilcoxon rank-sum test between pre-movement rates on CW choice trials and those on CCW choice trials; (5) rank-sum test between trial firing rates on left Blocks vs. right Blocks; (6) Kruskal-Wallis test between left, center, and right stimulus-driven firing rates. A neuron was considered active during the task, or to have detectable modulation during some part of the task, if any of the p values on these tests were below a Bonferroni-corrected alpha value ($0.15/6 = 0.025$). The lenient $\alpha = 0.15$ threshold was chosen due to the coarse nature of the tests which make them relatively insensitive to transient activity and thus increasing the likelihood of false negatives.

3.2.3 State space trajectory analyses

To establish how neural populations may solve the stimulus-action remapping problem posed by the task, we projected neural population activity onto dimensions that encode CW choices, CCW choices, and context. This approach allowed us to visualize the movement of neural population activity across different trial types in a low-dimensional state space to understand how populations transition between actions in different contexts in response to identical sensory stimuli.

We defined neural populations based on neuron locations within anatomical regions aligned to the Allen CCF atlas, including the eight regions we specifically targeted in our electrophysiological recordings: MRF, SCm, APN, RN, MOs, MOp, CP, and VPM/VPL. For each region, to ensure representations of population trajectories were adequately powered, we only included sessions with at least 10 neurons meeting

our quality control criteria. The criteria selected for neurons that had sufficient firing rates at baseline and/or within a trial as well as filtering out neurons with excess missing spikes (baseline or trial firing rate exceed 0.05 Hz; spike floor metric passed).

The three coding dimensions were defined using binned population firing rates as predictors and logistic regression to classify CW actions from baseline, CCW actions from baseline, and baseline activity between left and right Blocks. Specifically, the CW and CCW dimension decoders used pre-stimulus neural activity (-0.3 to -0.05 s relative to stimulus onset) and activity prior to a CW or CCW turn (-0.075 to 0 s relative to movement onset) to decode whether trials belonged to the baseline or pre-movement periods. The context dimension decoder only used pre-stimulus activity (-0.3 to -0.05 s relative stimulus onset) to decode whether trials belonged to the left or right Block.

For defining the context dimension, we used logistic regression to classify trials by Block identity (left vs. right) rather than using linear regression to fit the continuous context belief traces from our behavioral model. This approach was chosen because our analysis relies on comparing population trajectories across four discrete trial types defined by the task Block (left stimulus, center stimulus in left Block, center stimulus in right Block, and right stimulus), making the binary Block classification more appropriate than the continuous context variable. To ensure the neural trajectories represented successful implementation of task rules, we included only trials where the mouse made the correct choice in the decoders and subsequent projection onto decoding dimensions.

For each dimension, we applied L2-regularized logistic regression using a nested cross-validation approach. The outer loop performed 25-fold cross-validation across

trials to define dimensions on each training set and project only the held-out test data. Within each of these folds, we used a 5-fold inner cross-validation loop to select the optimal regularization parameter (λ). To address any class imbalances, we used inverse class frequency weighting on each outer loop iteration. The three dimensions (CW, CCW, and context) derived from each training set were orthogonalized using QR decomposition to ensure they represented independent components of the neural state space when projecting the corresponding test data. Since regression coefficients can have arbitrary sign flips between sessions, we standardized the orientation of dimensions by applying sign corrections when needed. Specifically, based on train set projections, we used the convention that pre-movement activity was greater than baseline activity along CW and CCW dimensions, and that right Block activity exceeded left Block activity along the Block dimension.

For trajectory visualizations and subsequent trajectory-based analyses, we implemented additional inclusion criteria based on decoding performance. Sessions were only included if they met minimum decoding accuracy thresholds: action decoding (classifying baseline from pre-movement trial periods irrespective of choice) cross-validated accuracy $\geq 70\%$ and choice decoding (classifying CW from CCW choices based on pre-movement activity) accuracy $\geq 60\%$. These thresholds helped ensure that included sessions were properly representative of how choice dynamics unfolded in a given region. We did not analyze APN, RN, or VPM/VPL for these analyses as they contained fewer than 5 sessions passing minimum neuron and decoding accuracy thresholds, precluding reliable estimates of trajectories.

With CW, CCW, and context dimensions defined for each session, we calculated

population trajectories by binning spike data using 75 ms sliding windows with 10 ms step sizes aligned to both stimulus onset (0 to 250 ms) and movement onset (-250 to 0 ms) and projecting the binned activity onto each dimension. To avoid overfitting, we only projected activity on held-out trials from our cross-validation procedure, ensuring all trials were represented through the k-fold partitioning described above.

To combine trajectories across sessions for each brain region, we first applied robust z-scoring (across all trials and timepoints) to each session's trajectories independently, bringing them into comparable units. We then calculated the median value of each of the four trial types for each time bin across all trials and sessions to obtain representative trajectories for each brain region. To standardize trajectory comparisons across brain regions, we calculated a rotation matrix for each region that aligned the main axis of motion (i.e., the mean trajectory between the four trial types from the first to final time bin) in the CCW-CW plane to a consistent 45-degree angle. This transformation maintained the context dimension values while rotating only the CW and CCW coordinates, enabling direct comparison of trajectory angles and stimulus input vectors.

To test different hypothesized mechanisms for stimulus-action remapping, we quantified two key features of these trajectories:

1. Baseline positions: We calculated the difference between CCW and CW coordinates at baseline for left and right Blocks separately. This allowed us to test whether there was a context-dependent motor bias in the population state prior to stimulus presentation. Differences between Blocks were calculated per session, and signed-rank tests were used to test for significance in each region.

2. Stimulus input angles: We measured the angle formed between trajectory positions at baseline (median position amongst 0-20 ms bins relative to stimulus onset) and post-stimulus response (median position amongst 100-150 ms bins relative to stimulus onset) in the CCW-CW plane using the four-quadrant inverse tangent (“atan2” in Matlab). Angles were calculated for individual trials, and then the median angle per trial type was computed across trials. We used rank-sum tests for each region to determine whether the initial trajectories of center stimuli followed different angles depending on the context.

Together, these analyses enabled us to distinguish between different possible mechanisms for how neural populations across the targeted regions implement context-specific stimulus-action mappings: through baseline motor bias shifts, through shifting stimulus input angles, or through context-dependent changes in attractor dynamics.

3.2.4 Population decoding of baseline context

We used cross-validated linear regression to decode context from baseline (pre-stimulus; 300 to 50 ms before stimulus onset) firing rates of neural populations. Context was decoded as the continuous latent context variable from the context belief model (Fig 2.1G, Chapter 2.2.5 & 2.3). Units were included as predictors if they passed the spike floor quality metric (described above) and had an average baseline firing rate ≥ 0.1 Hz. The sliding refractory period metric was not used as multi-unit activity is acceptable for population decoding. Sessions were included in the decoding analysis if ≥ 3 units fell in the region of interest and passed the above metrics. Elastic net

regularized ($\alpha = 0.5$) five-fold nested cross-validated logistic regression was used, wherein trials were split into train and test sets, and within the train set, the optimal lambda value was picked using five-fold cross validation (using the package cvglmnet for Matlab https://hastie.su.domains/glmnet_matlab/index.html). Performance of the decoding was measured as the Pearson's correlation between the predicted and actual context variable for a given session and region. To account for slight instability in the decoding performance due to random cross-validation partitions, the entire decoding procedure for each variable was run 10 times, and mean balanced accuracies were taken across the 10 decodings at each time bin. Baseline context decoding performance was not dependent on the number of neurons included in the decoder (Fig. S3A), behavioral performance on the decoded session (Fig. S3B), or uninstructed movements/pupil position during the baseline period (Fig. S3C-E, described below).

To obtain null distributions for decoding performance, the decoders were refit to 500 shuffles of the context variable. Shuffles were performed as described in the “Pseudosessions” (Chapter 3.2.8) section below.

3.2.5 Task alignment of spontaneous movements and pupil dynamics

A potential confound in the neural coding of baseline context is context-dependent uninstructed movements or pupil position changes. These could be driving neural activity that is correlated with context rather than the activity representing the context variable itself. In particular, if mice systematically shifted their pupil position before stimulus onset based on the current Block, this could alter the retinal input for identical stimulus positions across Blocks, potentially providing an alternative

mechanism to solve the task.

We used a sensitive approach to detect any context-dependent uninstructed movements or pupil position changes by decoding context from face video motion energy SVDs and pupil position traces (Fig. S3C). The predictor matrix was composed of the mean values during the baseline period (-300 to -50 ms relative to stimulus onset) on each trial of the top 16 principal components of face motion SVDs (capturing facial, whisker, front of body, and forelimb/wheel movements) and pupil x and y positions. This predictor matrix was regressed against the context variable using ordinary least squares (OLS) regression.

Performance was calculated on held-out trials using 5-fold cross validation. To account for variability in random data partitions on the results, the entire decoding procedure was run 25 separate times per session, and the final performance was calculated as the mean score across all 25 runs.

To assess statistical significance, null distributions were formed by running the same decoding procedure 1000 times using context variables from pseudosessions. Context could be significantly decoded ($p < 0.05$) from uninstructed movements/pupil positions in 8 out of 35 sessions. To compare decoding performances from uninstructed movements with those from neural activity, if multiple neural populations were decoded in a session, the population with the lowest p value and highest score were used to represent a given session. Only 2 out of the 16 sessions with significant neural decoding of context showed significant uninstructed behavioral decoding (Fig. S3D). Furthermore, there was no significant correlation between the decoding score (measured as ΔR^2 over median of null) for context predicted from uninstructed

behaviors versus context predicted from neural activity (Fig. S3E, $r = -0.13$, $p = 0.46$).

3.2.6 Kernel regression

To identify and disentangle the encoding of choice, stimulus, and contextual modulation of trial responses in single neuron firing rates, (Fig. 3.3, Fig. S4), we fit a kernel regression model (Park et al., 2014; Steinmetz et al., 2019). The firing rate of each neuron in this analysis is described as a linear sum of temporal filters aligned to task events. Only stimulus onset and movement onset kernels were required as this analysis was restricted to the time period between the two on each trial. The predicted firing rate $f_n(t)$ for neuron n on timepoint t is given as:

$$f_n(t) = \sum_c \sum_{t_s \in S_c} K_{c,n}(t - t_s) + \sum_{t_s \in S_{c=0}} Z_{c=0} K_{Z_v,n}(t - t_s) + \sum_{t_m \in M} [K_{m,n}(t - t_m) + D_m K_{D,n}(t - t_m) + Z_m K_{Z_a,n}(t - t_m) + Z_m K_{Z_a,n}(t - t_m) \cdot D_m K_{D,n}(t - t_m)] + \sum_p P_{p,n} + \sum_f F_{f,n}$$

Here, c represents the three visual stimulus types (-1, 0, 1 = left, center, right). S_c represents the set of times for which this stimulus appeared, and $K_{c,n}(t)$ represents the Stimulus kernel function of this stimulus for neuron n . M represents the set of movement times and $K_{m,n}(t)$ represents the Action kernel for neuron n . D_m represents the direction of movements (-1 for CCW, +1 for CW), and $K_{D,n}(t)$ represents the Choice kernel for neuron n . There are two context kernels: $K_{Z_v,n}(t)$ and $K_{Z_a,n}(t)$, which represent the Context-Stim kernel and Context-Premotor kernel for neuron n . Z represents the value of the latent context variable (Fig. 2.1G) which is normalized to range from -1 to 1. $S_{c=0}$ represents the set of times for which the center stimulus appeared, and thus the Context-Stim kernel is only active over center stimulus trials. There is additionally an

interaction term between the Context-Premotor kernel and the Choice kernel. Finally, p represents the two pupil position coordinates, and $P_{p,n}$ is the Pupil predictor for neuron n . Likewise, f represents the top 16 face motion SVD components, and $F_{f,n}$ is the Face predictor for neuron n . The Pupil and Face predictors were unlagged, which has been found to maximize variance explained in neural activity (Stringer et al., 2019). The Vision and Context-Stim kernels are supported over the window 0 to 0.4 s relative to stimulus onset. The Action, Choice, and Context-Premotor kernels are supported over the window -0.25 to 0 s relative to movement onset. For trials with reaction times shorter than 400 or 250 ms, we trimmed the respective kernels such that only timepoints before movement onset and after stimulus onset were included in the model.

To fit each kernel, we constructed a Toeplitz predictor matrix of size $T \times l$, in which T is the total number of time bins and l is the number of lags required for the kernel. The predictor matrix contains diagonal stripes starting each time an event occurs and 0 otherwise. The predictor matrices were horizontally concatenated to yield a global prediction matrix $\mathbf{P}$ of size $T \times L$ containing all predictor kernels. Rate vectors of all N neurons in a given session were horizontally concatenated to form $\mathbf{F}$, a $T \times N$ matrix. If kernels extended before stimulus onset or after movement onset on a given trial, these rows were removed from $\mathbf{P}$ and $\mathbf{F}$.

To prevent noisy and overfit kernels, we implemented reduced-rank regression (Izenman, 1975; Steinmetz et al., 2019), which allows regularized estimation by factorizing the kernel matrix $\mathbf{K}$ into the product of a $L \times r$ matrix $\mathbf{B}$ and a $r \times N$ matrix $\mathbf{W}$, minimizing the total error: $E = \|\mathbf{F} - \mathbf{PBW}\|^2$. The $T \times r$ matrix $\mathbf{PB}$ consists of a set of ordered temporal basis functions that can be linearly combined to estimate the neuron's firing

rate over the whole training set and which results in the best possible prediction from any rank r matrix. To estimate each neuron's kernel functions, we generated the reduced rank predictor matrix $\mathbf{PB}$ for $r=20$ and estimated the weights $\mathbf{w}_n$ to minimize the squared error $E_n = \|\mathbf{f}_n - \mathbf{PBw}_n\|^2$ with elastic net regularization (using the package `cvglmnet` for Matlab https://hastie.su.domains/glmnet_matlab/index.html) with parameters $\alpha=0.5$ and using ten-fold cross-validation to determine the optimal value for λ for each neuron. The kernel functions for neuron n were then unpacked from the L -length vector obtained by multiplying the first $r=20$ columns of $\mathbf{B}$ by $\mathbf{w}_n$ (i.e. $\mathbf{k}_n = \mathbf{B}_{1:L, 1:20} \mathbf{w}_n$).

To determine whether a given kernel significantly contributed to a neuron's firing activity, we generated empirical null distributions for each neuron using a shuffling approach. The kernel matrix was refit whilst shuffling either the Stimulus, Choice, or Context kernels. The Stimulus kernel was shuffled by randomly permuting the stimulus position labels of c within their respective Blocks (i.e., center and Left stimuli shuffled with left Blocks, center and right stimuli shuffled within right Blocks). The Choice kernel was shuffled by randomly permuting sequences of choices (to preserve temporal structure) in D . The Context kernels were shuffled using the context representations generated from pseudosessions (described in detail in the Pseudosessions section below) in place of the true context trace in L . Shuffles were performed 500 times, with only one kernel being shuffled at a time in each iteration. The Context-Stim and Context-Premotor kernels were shuffled both together and separately to assess the total effect of contextual modulation on trial firing rates as well as its effects on stimulus and premotor-related activity specifically. Thus, a total of $500 \times 5 = 2500$ shuffled kernel matrices were fit per neuron. We used the same five-fold cross-validation (splits

performed across trials) for both the true and shuffled kernel regressions, and the cross-validated variance explained was used as the test statistic (p values and effect sizes were derived as described in Chapter 1.4.7).

3.2.7 Drift-diffusion modeling of stimulus-action remapping mechanisms

To examine whether each of our three hypothesized mechanisms (context-dependent shift in attractor dynamics, baseline motor bias, and center stimulus input angle) could theoretically solve the stimulus-action remapping problem posed by the task, we instantiated each mechanism using drift-diffusion models (DDMs) (Ratcliff and McKoon, 2008).

We implemented a stochastic DDM for a one-dimensional decision variable representing position in state space along a CCW to CW axis. For each mechanism, we simulated 1000 trials with simulation parameters chosen to qualitatively match observed neural dynamics. Simulations were run for 200 ms using a timestep (dt) of 0.01 s. The decision variable at each timestep was updated according to:

$$v(t + 1) = v(t) + \eta(t + 1) + \alpha(v(t) - \theta)$$

where η represents Gaussian random noise (standard deviation = 0.1), θ is a mechanism-specific threshold parameter that determines the attractor dynamics, and α (0.02) scales the influence of the current state on future updates.

The initial state of the decision variable for each simulation was drawn from a normal distribution with mechanism-specific mean starting value and standard deviation of 0.3. At the first timestep after $t = 0$, a stimulus-specific input was applied, representing the sensory input to the system.

For the dynamics shift mechanism, we kept the starting value constant at 0 for all trial types, applied stimulus-specific inputs ($[1 \ 0 \ 0 \ -1]$ for left, center left Block, center right Block, and right stimuli respectively), and set thresholds to induce opposite dynamics between Blocks ($[1 \ 1 \ -1 \ -1]$; dynamics biased toward CCW action in left Block and CW in right Block). For the baseline shift mechanism, we varied the starting values ($[-0.3 \ -0.3 \ 0.3 \ 0.3]$) to bias initial positions toward CCW in left Blocks and CW in right Blocks, while keeping the threshold parameter at 0 and using the same stimulus-specific inputs. For the stimulus input shift mechanism, we kept starting values and thresholds at 0 for all trial types, but modified the stimulus inputs ($[0.75 \ -0.375 \ 0.375 \ -0.75]$) to induce trajectories toward CCW or CW actions for center stimuli in left or right Blocks.

For each mechanism, we visualized the simulation results in three ways: 1) individual decision variable trajectories for 30 randomly selected trials of each type, 2) mean trajectories across all 1000 trials per type, and 3) distributions of final decision variable values. Each mechanism was able to solve the task, with performance comparable to that of the mice at the tuned parameters detailed above.

3.2.8 Pseudosessions

To generate an empirical null distribution for the context variable, we used pseudosessions (Harris, 2020), wherein the same random distribution from which the Block sequence was drawn (random uniform distribution between 40 and 70 non-repeat trials per Block; first Block is between 1 and 70) was used to generate “pseudoblock” sequences. As trials in the true session were filtered for reaction times (75-400 ms), we

filtered pseudosessions in the same way: Reaction times were assigned to each pseudosession trial by sampling with replacement from trials of matching trial type (i.e. left stim, center stim left Block, center stim right Block, right stim).

To get shuffled context representations, i.e., the latent context variables from the context belief behavioral model generated on pseudosessions, we used the fitted weights for each mouse and fed in pseudosession trial sequences (Fig. 2.1H). The model's output probability distribution was sampled based on the previous observation and latent value to predict the choice on each trial. Choices were fed back into the model to update the latent value and predict the choice on each subsequent trial. Following the same rule as in the behavioral task, trials with incorrect choices were repeated.

3.2.9 Viral injections and brain clearing

Viral injections were performed with the Nanoject III (Drummond Scientific). Glass pipettes were pulled using the Narashige pipette puller PE-22 with 1.0 mm OD glass and the following settings: heater 70.5, sub magnet 20.9, main magnet 100.7. Pipettes were 5-6 mm in length and broken back to a tip diameter of ~30 microns.

After performing a small (~0.5x0.5 mm) craniotomy at the desired AP and ML coords (Table 1), the pipette was driven down to and 100 um past the target injection depth. It was then drawn back 100 um to create a pocket for the injection. Injection volumes ranged from 50-75 nl with an injection speed of 1 nl/s. After the injection was completed, there was a 2 minute wait time, after which there were 2 sets of withdrawing 100 um waiting 1 minute between each, then 5 sets of withdrawing 100 um waiting 30

seconds between each. The pipette was then fully withdrawn out of the brain over about 2 minutes.

Virus	Virus volume	Brain region/coordinates (mm from bregma)
AAV2-hSyn-EGFP (Addgene, viral titer 2×10^{11} - 4×10^{11} vg/ml (1:8-1:12 dilutions))	50-75 nl	• MRF anterior, (AP -2.90, ML $\pm$ 0.85, DV 3.85 mm) • MRF intermediate anterior, (AP -3.50, ML +1.25, DV 3.65) • MRF intermediate posterior, (AP -4.10, ML +1.05, DV 3.15) • MRF posterior, (AP -4.70, ML +0.5, DV 3.8)

To visualize whole-brain structure while preserving fluorescence of virally induced fluorescent proteins, we performed tissue clearing using stabilization under harsh conditions via intramolecular epoxide linkages to prevent degradation (SHIELD) (Park et al., 2019) with the Clear+ passive clearing kit and protocol from LifeCanvas Technologies. The protocol is explained in detail by LifeCanvas Technologies (<https://sites.google.com/lifecanvastech.com/protocol/delipidation/passive-delipidation-protocol>); here we briefly describe each step. Brains were extracted immediately following perfusion and post-fixed in 4% paraformaldehyde for 24 h. SHIELD preservation was performed by first incubating in fresh SHIELD OFF solution (25% DI H₂O, 25% SHIELD-buffer solution, 50% SHIELD-Epoxy solution) for 24 hours at 4°C with shaking for 4 days. Following this incubation, brains were incubated in SHIELD ON Buffer at 37°C with shaking for 24 hours. To clear brains by delipidation, brains were incubated in delipidation buffer at 45°C with shaking for 7-8 days. Finally, brains were index matched at 37°C with shaking using 1:2 diluted EasyIndex solution for 24 hours,

followed by 100% EasyIndex solution for 24-48 hours (until transparent).

Brains were mounted (horizontal orientation) onto custom 3D-printed holders by embedding into an EasyIndex agarose solution. Once brains were mounted in the agarose, the whole sample was immersed in EasyIndex overnight and imaged the following day. Brains were imaged on a light sheet microscope (LifeCanvas Technologies SmartSPIM) using two-sheet illumination (100% width) and the 3.6x objective (1.8x1.8 μm resolution). Images were gathered from the dorsal surface of the brain to the ventral surface in 2 μm steps in 488 nm (EGFP, ~15% power) and 639 nm (autofluorescence, ~70% power) excitation channels. The full resolution image stacks as well as CCF-transformed stacks (using procedure as described above) were then used for visualization (Fig. 3.1B).

3.3 Results

3.3.1 Stimulus, choice, and context-dependent dynamics in MRF

The MRF is the 3rd largest subcortical brain region (Fig. 3.1A, based on the Allen CCFv3 atlas (Q. Wang et al., 2020)), and sends ascending and descending projections to decision- and motor-related circuitry (Fig. 3.1B). The MRF sits medial and dorsal to the substantia nigra pars reticulata (SNr) and ventral to the superior colliculus (SC). Its boundaries are heterogenous and delineated around more well-defined structures including the red nucleus (RN) (embedded within the ventral-anterior aspect of MRF) and superior cerebellar peduncle (courses through ventral-posterior aspect of MRF) (Fig. 3.1A). To confirm the outputs of MRF, we injected anterograde viral vectors (AAV2-hSyn-EGFP) followed by brain clearing (Park et al., 2019) and atlas-registered

light-sheet imaging. Consistent with prior literature, MRF axons were found most prominently in SC (particularly intermediate and deep layers, collectively SCm (Benavidez et al., 2021)), zona incerta (Kolmac et al., 1998; Oh et al., 2014), reticular, parafascicular (PF), and mediodorsal (MD) thalamus (Inagaki et al., 2022; Kolmac and Mitrofanis, 1998; Oh et al., 2014), and lower parts of the reticular formation in the pons and medulla (Oh et al., 2014; Perkins et al., 2014) (Fig. 3.1B). Thus, the MRF's connectivity places it within the recurrent loops underlying flexible decisions, receiving inputs from frontal and motor cortex (Gao et al., 2022; Peng et al., 2021; Winnubst et al., 2019), SC (Benavidez et al., 2021), and SNr (McElvain et al., 2021; Oh et al., 2014) and sending outputs to frontal/motor-projecting thalamic nuclei and SC.

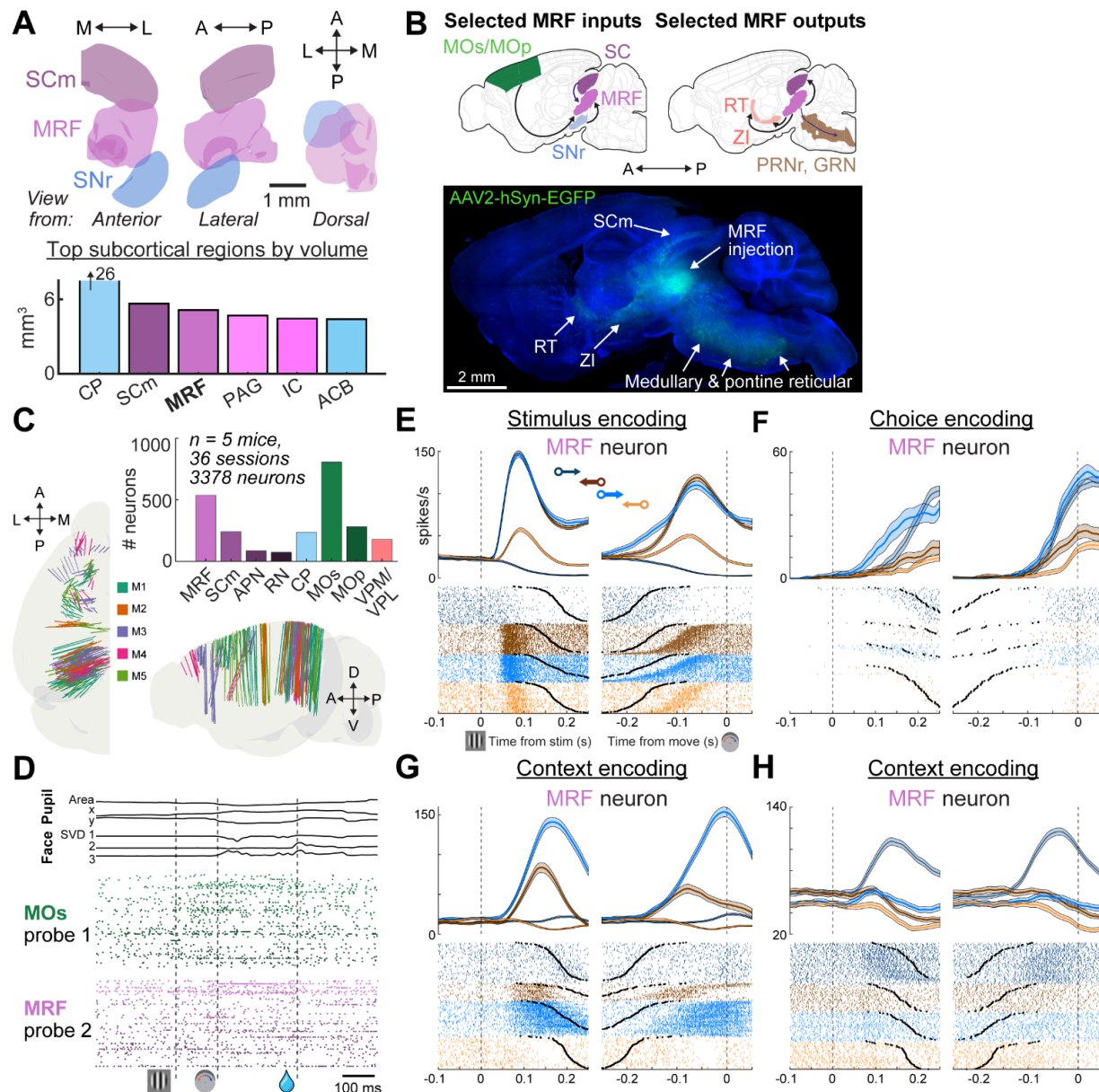


Figure 3.1. Dense electrophysiological recordings of MRF and connected regions during the flexible stimulus-action remapping task show diverse trial response dynamics. **A)** Top: 3D renderings of MRF along with adjacent SCm and SNr for reference. Bottom: Top 6 subcortical brain regions by volume according to Allen CCF atlas. **B)** Selected input and output connections of MRF based on previous studies and confirmatory viral tracing data. Example images of MRF outputs from viral tracing experiments with AAV2-hSyn-EGFP injected into MRF are shown. **C)** 3D view of all Neuropixels 2.4 insertions from task recordings from dorsal (left) and lateral (bottom right). All insertions are projected onto left hemisphere for visualization. Top right, number of neurons recorded from a set of 8 regions after applying single unit quality control and task-responsiveness metrics. **D)** Example spike rasters during a dual probe recording in MRF and MOs, with task events overlaid and face video motion energy SVDs, pupil position, and pupil area above. **E)** PETHs by trial type (filtered for non-repeat correct trials) of an example stimulus-responsive SCm neuron. Spikes are aligned to stimulus onset on the left and wheel turn onset on the right. Top traces are trial-averaged PETHs with 1 SE shaded, bottom are spike rasters sorted by trial type (color) and reaction time within each trial type (black dots are wheel turn times on left

plot and stimulus times on right plot). Inset is a color legend for the 4 trial types as in Fig. 2.1B (top to bottom: left stim, center stim left Block, center stim right Block, right stim). **F)-H)** Same plots as in **E)** but for choice-responsive MRF neuron, contextually modulated MRF neuron, and contextually modulated CP neuron respectively.

Based on MRF anatomy, connectivity, and prior decision-making literature, we performed targeted Neuropixels 2.4 (Steinmetz et al., 2021b) recordings in the forebrain and midbrain (n = 5 mice, 36 sessions, 3378 task-responsive quality-controlled neurons, 2444 neurons amongst the 8 regions in Fig. 3.1C). Regions that strongly encoded decision variables in related tasks include secondary motor cortex (MOs), primary motor cortex (MOp), deep layers of the superior colliculus (SCm), and caudoputamen (CP) (Chen et al., 2024; Duan et al., 2021; Guo et al., 2014b; International Brain Laboratory et al., 2023a; Khilkevich et al., 2024; Pasupathy and Miller, 2005; Steinmetz et al., 2019). Two nearby midbrain regions, the anterior pretectal nucleus (APN) and RN respectively, as well as the somatosensory parts of the thalamus, ventral posteromedial and ventral posterolateral thalamus (VPM/VPL), were targeted for recordings as comparison regions (Fig. 3.1C). To reliably characterize the spatial coding of task variables within MRF, recordings were targeted to densely tile the entire region (Fig. 3.1C). During all electrophysiological recordings, we simultaneously recorded the subject's eye and front body using high-speed videography to assess the contributions of pupil dynamics and spontaneous body movements to neural activity (Fig. 3.1D) (Musall et al., 2019b; Stringer et al., 2019).

MRF neurons responded to task stimuli and choice, as has been observed previously (Chen et al., 2024; International Brain Laboratory et al., 2023a; Khilkevich et al., 2024; Steinmetz et al., 2019). Peri-event time histograms (PETH) were constructed for each neuron recorded in the task, aligned to both stimulus onset and movement

onset, with reaction time variation allowing for the assessment of stimulus vs. movement-aligned activity. Several MRF neurons had stimulus-related activity, which was aligned to stimulus onset and often selective for one or two of the stimuli (if the receptive field is narrower than the 160° spanned by the gabors) (example neurons in Fig. 3.1E, G, H). MRF neurons also contained premotor activity selective for the impending action (i.e. choice activity). This activity was aligned to movement onset and had differential firing between CW and CCW movements independent of stimulus identity (e.g., Fig. 3.1F).

MRF neurons appeared to encode task context both in baseline firing rates and trial responses. We observed that some MRF neurons had trial responses that could not be straightforwardly explained by the stimulus position (left, center, or right), the direction of the action (CW or CCW wheel turn), or a combination of the two. Rather, it appeared that these neurons had stimulus and/or premotor activity that was modulated by context. For example, the neuron in Figure 3.1G had responses following center stimuli which varied between Blocks (right Block center stimulus response greater than left Block). This difference could be due to context-dependent changes in visual stimulus input or context-dependent changes in choice dynamics (i.e., strong CW preference in left Block, weak or no CW preference in right Block); as this is difficult to ascertain qualitatively, we use a kernel encoding model described in the following section to disentangle coding of each task variable. Context coding in the baseline (pre-stimulus) period was more straightforward: firing rates before stimulus onset were higher in one Block than the other (e.g. Fig. 3.1H, left Block baseline firing rates greater than right Block).

To rigorously quantify context coding in the subsequent neural analyses of this study, we used pseudosessions (Harris, 2020) to create null distributions of context on each session. A key potential pitfall in looking for relationships between context and neuronal firing rates is the risk of finding “nonsense correlations”: since both variables have slowly varying components and high autocorrelation, they will often covary by chance leading to excess false positives with standard statistical tests (Harris, 2020; Yule, 1926). To address this confound, we designed the task to have randomized Block lengths (Fig. 2.1C), which allowed us to use the same generating distribution that generated the true Block sequence to create “pseudosessions”. Pseudosessions were fed to the behavioral model to generate a null distribution of latent context variables, preserving the temporal statistics of the original context time series while disrupting its precise sequence (Fig. S1H, Chapter 3.2.8). In the remainder of the study, all quantifications of relationships between context and neuronal activity are compared against a null distribution based on context traces from pseudosessions.

Having identified single MRF neurons that coded for context both at baseline and in modulation of trial responses, we next investigated how these context representations during different periods might contribute to solving the stimulus-action remapping problem in this task.

3.3.2 Stimulus-action remapping is mediated by baseline context representations in MRF, SCm, CP, and MOs

We hypothesized three possible mechanisms by which a neural population can solve the stimulus-action remapping problem posed by the flexible stimulus-action remapping task (Fig. 3.2A). We used a framework wherein there are equal strength

attractors for the CCW and CW actions which drive the neural population dynamics during the trial (Fig. 3.2A, circled black X's). The visual stimulus provides an initial input to the system (Fig. 3.2A, first vector from origin), with the direction and magnitude of the input dependent on stimulus identity (left stimulus pushed toward CW ($<45^\circ$), right stimulus toward CCW ($>45^\circ$), center stimulus neutral ($\sim 45^\circ$)). Additionally, the system occupies an initial position in the 2D space equidistant from the CCW and CW attractors. In the first hypothesized mechanism for center stimulus-action remapping, the initial position of the system is biased toward the CCW or CW attractor in the left or right Blocks respectively (Fig. 3.2A, "Baseline motor bias"), allowing for the center stimulus trials to flow toward opposite actions between Blocks with neutral stimulus input and symmetrical action dynamics. In the second hypothesized mechanism, the center stimulus input angle is not neutral but rather shifts to $>45^\circ$ in the left Block and $<45^\circ$ in the right Block (Fig. 3.2A, "Stim input angle shift") while initial positions and attractor strengths are equal and constant. In the third hypothesized mechanism, there is a context dimension orthogonal to the CW and CCW dimensions which mediates a shift in the attractor strengths and thus action dynamics between Blocks (CW $>$ CCW, right Block, CCW $>$ CW, left Block; Fig. 3.2A, "Dynamics shift mediated by orthogonal context dimension"). Notably, this hypothesis requires the context dimension to be instantiated during the baseline (pre-stimulus) period in order to set the attractor strengths which drive the ensuing trial dynamics. To validate the above framework, we instantiated each of the three hypothesized mechanisms as drift diffusion models and found that each model could adequately solve the task (Fig. S2A-C, Chapter 3.2.7).

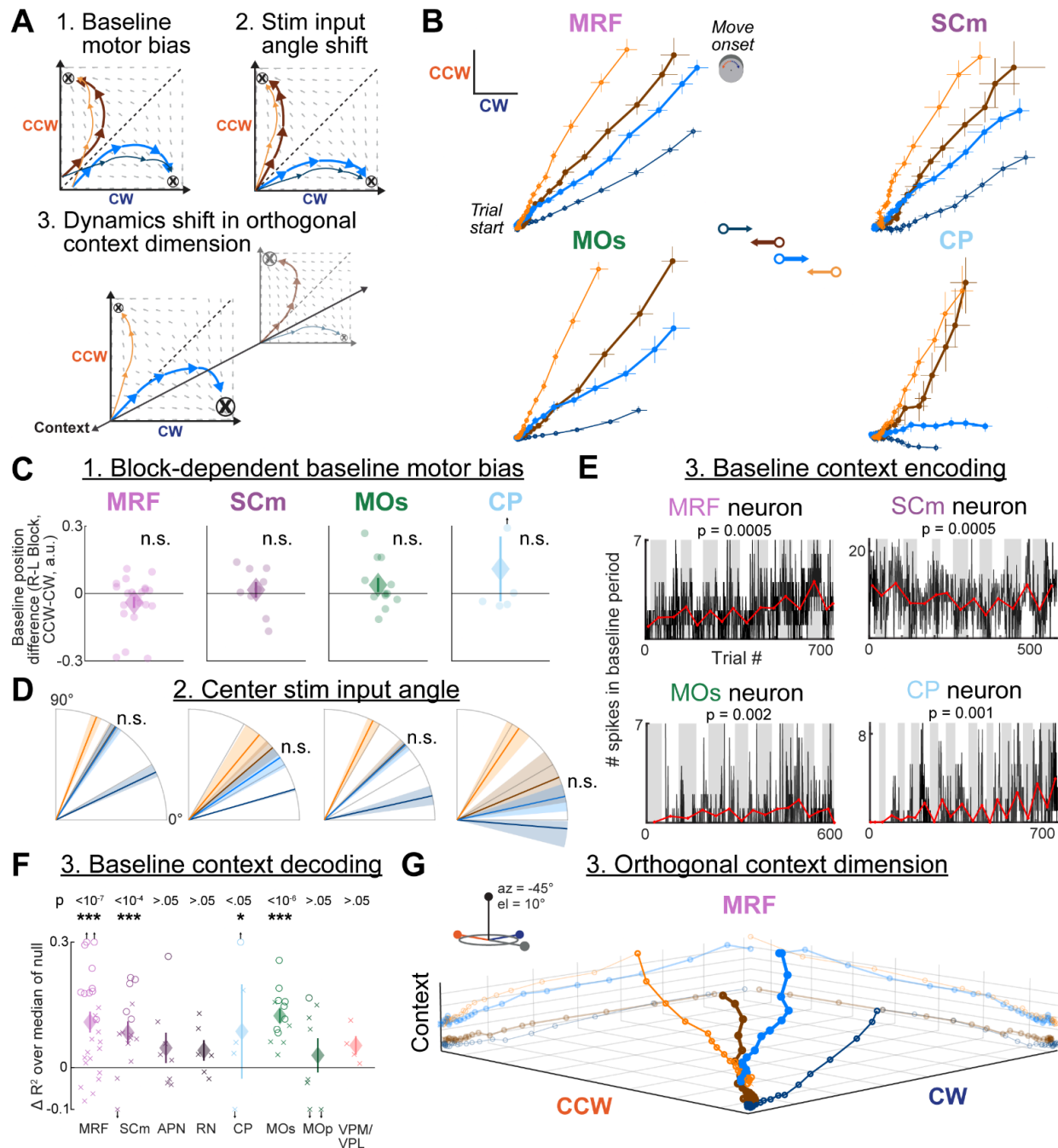


Figure 3.2. Stimulus-action remapping mechanism is most consistent with a context-dependent shift in action dynamics in MRF, SCm, CP, and MOs. **A)** Schematics of three hypothesized mechanisms by which a neural population can solve the stimulus-action remapping problem in state space. Trajectories are colored by trial type and evolve according to the underlying flow field dynamics (as determined by attractor strengths for CW and CCW actions, marked by circled black X's), baseline position, and stimulus input angle. 1. Baseline motor bias shift mechanism (left Block biased toward CCW, right Block biased toward CW). center stimulus input angle is 45° and unchanged between Blocks; action attractor strengths are equal. 2. center stimulus input angle shift ($\sim 60^\circ$ left Block, 30° right Block). Baseline position along CCW and CW dimensions is the same between Blocks; action attractor strengths are equal. 3. Context-dependent shift in dynamics (action attractor strengths shift between Blocks). A third

orthogonal context dimension mediates an increase in CCW attractor strength and decrease in CW attractor strength in the left Block, and vice-versa in the right Block. Baseline position along CCW and CW dimensions is equal between Blocks; center stimulus input angle is 45° and unchanged between Blocks. **B)** 2D (CCW vs CW) state space trajectories (median across all trials standardized and combined from all sessions; error bars are 95% confidence interval (CI)) in MRF, MOs, SCm, and CP. Each point is a timepoint within a correct trial. Trial starts at the bottom left. Trajectories span from -250 to 0 ms aligned to movement onset in a sliding window of 75 ms bins with 10 ms step sizes. **C)** Difference between baseline positions in CCW-CW state space between right Block and left Block. Each circle is a session, diamond is mean, error bar is SEM. No significant differences between Blocks were found in any region using Wilcoxon signed-rank test. **D)** Stimulus input angles across the four trial types (median $\pm$ 95% CI across all trials). Angles formed between the x axis and the line between averaged positions of 0-20 ms bins and 100-150 ms bins (aligned to stimulus onset) in CCW vs CW state space. No significant differences between left and right Block center stimulus angles were found in any region using rank sum test. **E)** Example neurons with significant correlations ($p < 0.05$) between baseline firing (300 to 50 ms before stimulus onset) and context. Black trace is trial to trial baseline spike count, Block identity is indicated by shaded vs. unshaded patches, and red trace is the mean baseline spike count in each Block. P values are derived from null using pseudosessions. **F)** Decoding of context from baseline firing (300 to 50 ms before stimulus onset). Diamonds are mean $\pm$ SEM ΔR^2 across sessions. Circles are individual sessions with significant decoding ($p < 0.05$ against pseudosession null); x's are sessions with non-significant decoding ($p > 0.05$). P values per region are derived from Fisher's combined probability test. **G)** 3D state space trajectory of MRF with the added third dimension of context. As opposed to the 2D case, center stimulus trajectories are separable across Blocks starting at baseline.

We did not observe a context-dependent motor bias in the state space trajectories of any of the examined regions. To define CW, CCW, and context coding dimensions, we used logistic regression to decode CCW and CW actions from trial firing rates, and context from baseline firing rates, across all neurons in a population. These three dimensions were orthogonalized, and trial-by-trial binned neural activity was then projected onto each dimension. To test hypotheses 1 and 2 (Fig. 3.2A), we first visualized the average trajectories in MRF, SCm, MOs, CP, and MOp (median across trials from all sessions) for each of the four trial types in a 2D CW vs. CCW state space (Fig. 3.2B, Fig. S2E; APN, RN, and VPM/VPL were excluded due to insufficient sessions passing inclusion criteria, see Chapter 3.2.3). The trajectories were inconsistent with the context-dependent motor bias mechanism: Baseline positions did not significantly differ between Blocks in any region (Fig. 3.2C, $p > 0.05$ signed-rank

test; MOp not shown).

Center stimulus input angles showed no significant differences across Blocks in any region (Fig. 3.2D). We defined the stimulus input angle for each trial type as the angle formed between the x axis and the line between the baseline and 125 ms bins (aligned to stimulus onset) in CCW vs CW state space. The left and right stimulus input angles consistently pushed the trajectory toward CW and CCW actions respectively (MRF = 25.6° left stimulus, 65.7° right stimulus; median across trials standardized and combined across all sessions). The center stimulus input drove a neutral trajectory angle in between that of the left and right stimulus (MRF = 57.6° left Block, 57.0° right Block), with no significant differences in center stimuli between Blocks in any region (Fig. 3.2D, $p > 0.05$ rank-sum test; MOp not shown).

The trajectories of MRF, SCm, CP, and MOs populations were most consistent with a context-dependent shift in action attractor dynamics established during the baseline period. We first considered whether context was represented at baseline (-300 to 0 ms before stimulus onset) in each region. Trial-by-trial baseline firing rates in single neurons of MRF, SCm, CP, and MOs were significantly correlated with context (Fig. 3.2E, Fig. S2D, RN, APN, MOp, VPM/VPL below chance). Context was also found to be decodable from the baseline firing of neural populations in MRF, SCm, CP, and MOs (Fig. 3.2F $p < 10^{-7}$, 10^{-4} , 0.05, 10^{-6} , Fisher's combined probability test). The decodability of context in each session was not dependent on the number of neurons in the population (Fig. S3A) nor on the behavioral performance of the subject (only sessions with significantly above chance performance on center stimulus trials were included, but there was variability within these sessions; Fig. S3B, Chapter 2.2.3). We next projected

neural population activity onto CW and CCW dimensions as well as a third orthogonal context dimension. This third dimension was defined using logistic regression to classify Block identity (rather than the latent context variable from the behavioral model, see Chapter 3.2.3) from baseline firing rates of the neural population. As opposed to the 2D state space, center stimulus trajectories between Blocks in this 3D state space were separable at baseline (Fig. 3.2G — MRF, Fig. S2G-I — MOs, SCm, and CP). In contrast, whereas MOp population trajectories looked similar to MRF, MOs, SCm, and CP in 2D state space, in 3D state space, there was not a significant baseline separation between Blocks along the context dimension (Fig. S2E-F).

Uninstructed task movements and pupil dynamics do not account for the context representations observed here. While mice were instructed to hold the wheel still during the baseline period used for this decoding (pre-stim wheel quiescence condition, Chapter 2.2.1), it is nonetheless possible that there were subthreshold movements and/or uninstructed movements elsewhere on the mouse's body that correlated with context. Given the high degree of variance in neural activity that can be explained by body movements (Musall et al., 2019b; Stringer et al., 2019), the interpretation of significant neural correlations with context may be confounded by possible context-aligned uninstructed movements. To sensitively detect whether movements were significantly aligned with context during the baseline pre-stimulus period, we used logistic regression to decode trial-by-trial context from face video motion energy singular value decompositions (SVDs) (tracking face, body, and wheel movements, Fig. 2.1A) and pupil position traces aligned to the 250 ms before stimulus onset (Fig. S3C). There was no relationship between the decoding scores from uninstructed movements/pupil

and decoding scores from neural activity per session (Fig. S3D). Moreover, out of 34 sessions, context was significantly decodable from neural data in 16 sessions and from uninstructed movements/pupil in 8 sessions. Only 2 sessions showed significant decoding from both sources, which approximates the expected 3.8 sessions of overlap if these variables were independent (Fig. S3E).

We hypothesized that the shifting context-dependent attractor dynamics apparent in population trajectories would result in contextual modulation of premotor activity. To test this directly, we next analyze single neuron trial responses to quantify how context modulates stimulus- and choice-related activity.

To disentangle and statistically quantify the contribution of stimulus, premotor, and context variables to trial responses, we used a kernel regression approach to fit trial-by-trial single neuron activity as a function of task-related events and behavioral measures (Park et al., 2014; Steinmetz et al., 2019) (Fig. 3.3A). Context can be represented in trial responses of neurons by either modulating stimulus-aligned activity to the center stimulus (Fig. 3.3A, “Context-Stim” predictor) or by modulating movement (i.e., wheel-turn) aligned activity (Fig. 3.3A, “Context-Premotor” predictor), independent of combinatory effects of the stimulus and choice on firing rates alone. We used a cross-validated (CV) nested approach wherein the model was refit with either the stimulus, choice, or context predictors shuffled to identify the unique contributions of each variable to a neuron’s trial response against an empirical null distribution (Fig. 3.3A-E, Fig. S4, Chapter 3.2.6). As the visual stimulus kernel is shuffled by permuting the stimulus labels (left, center, or right), visual stimulus coding is specifically defined here by responses which are selective for one or two of the stimuli rather than a

nonspecific response to all stimuli. Additionally, the model included face video motion energy SVDs and pupil position traces as regressors to help account for any possible motor differences between Blocks during the trial period (Fig. 3.3A, “Spontaneous behavior” predictors).

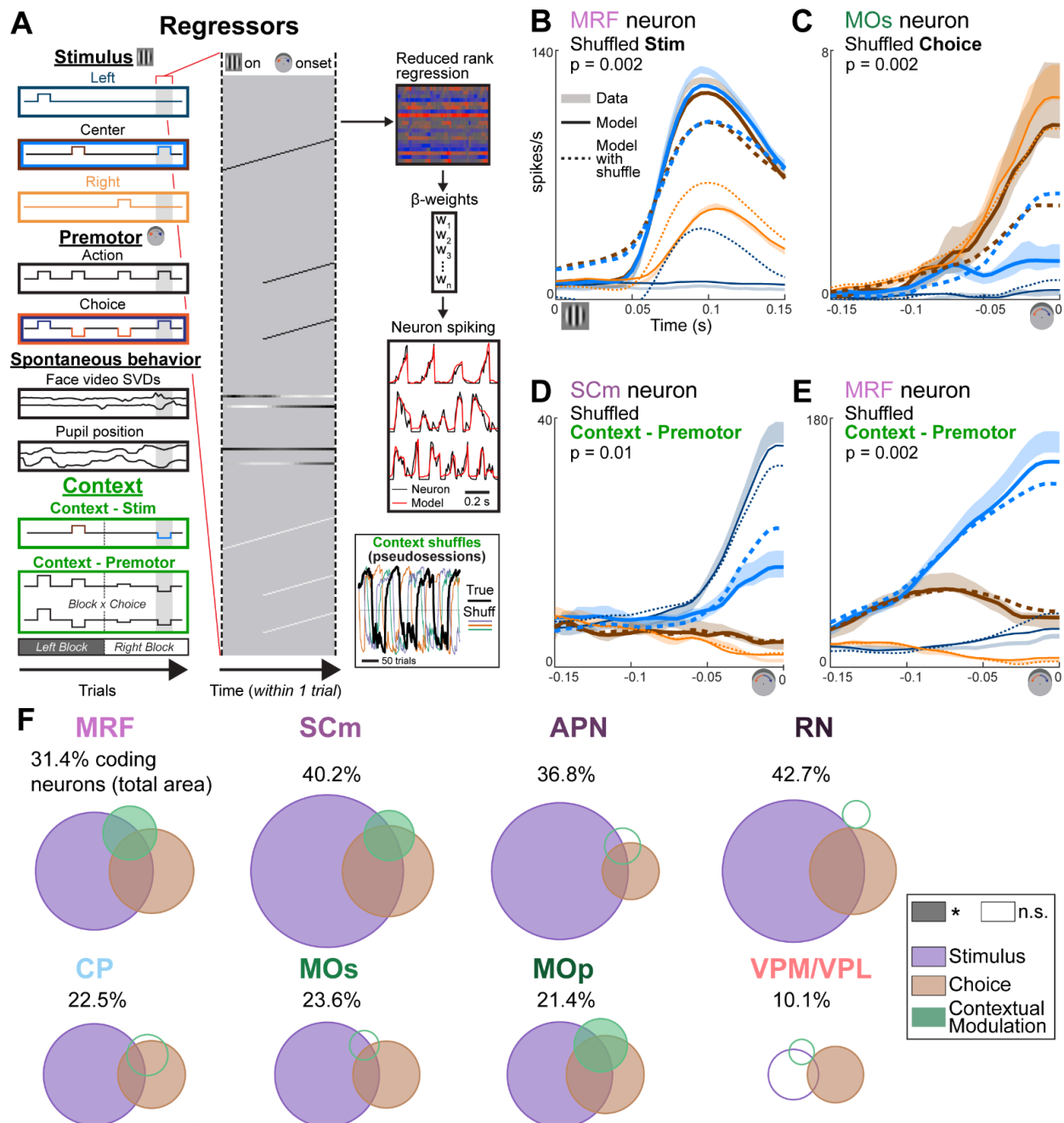


Figure 3.3. Kernel regression encoding model reveals enhanced midbrain stimulus coding, widespread choice coding, and contextually modulated dynamics. A) Kernel regression encoding

model for single neuron activity within trial periods (i.e., between stimulus and movement onset on each trial) during the flexible stimulus-action remapping task. Left: task events across 4 example trials aligned to stimulus onset or movement onset, as well as ongoing spontaneous behavior traces, throughout each trial. To the right of the events is a zoomed in view of the regressors for one trial (marked by gray shading and red bracket) which are convolved with a temporal kernel aligned to either stimulus onset or movement onset. Spontaneous behavior regressors are not lagged. White indicates a regressor value of -1, black is +1. Predictor matrix is reduced to 20 predictors using reduced rank regression, where target variable is binned spiking from all neurons in a session. Ridge regression is then used to fit the binned spiking of one neuron. Bottom right: schematic of pseudosession approach to create a shuffled null distribution for task context. True context representation is plotted on top of 3 example context representations generated from pseudosessions. **B)** Trial-averaged $\pm$ 1SEM PETHs aligned to stimulus onset of an MRF neuron encoding the center and right visual stimuli. Data, model, and model with shuffled visual stimulus labels are plotted. The shuffle iteration with the median R^2 among the shuffle distribution was used for plotting, and its fit to the data is diminished compared to the true model. A neuron is deemed to encode a task variable if the true model R^2 is greater than the 98th percentile R^2 of the shuffle distribution. **C)-E)** Same as B) but for MOs, SCm, and MRF neurons which encode choice, context-premotor, and context-premotor variables respectively. PETHs are aligned to wheel movement onset. **F)** Venn diagrams of percentages of neurons across the 8 target regions encoding stimulus, choice, and/or trial context (i.e. contextual modulation: Context - Stim and Context - Premotor combined), defined as $p < 0.02$ against respective shuffled null. Filled circles indicate above chance coding (Fig. S4B-C), empty circles indicate below chance coding. Value below the region labels is the percentage of neurons in a region represented by the total area encompassed by the three circles.

Regions across the midbrain contained the highest proportion of visual stimulus encoding neurons, whereas choice was more widely distributed across regions (Fig. 3.3F). Similar to previous studies (Chen et al., 2024; International Brain Laboratory et al., 2023a; Steinmetz et al., 2019), we found choice-encoding neurons across the midbrain and forebrain, with significant (lower 95% confidence interval (CI) above chance level; Fig. S4B and Chapter 1.4.7) coding in every region we recorded from and the highest proportions of choice neurons in SCm followed by RN, MRF, MOp, MOs, and CP (range of 8.1-14.9% at $p < 0.02$; Fig. 3.3F, Supp Fig. S4B). Choice coding was lower (slightly above chance) in APN (5.8%) and VPM/VPL (5.6%). The visual stimulus was more strongly encoded than choice (mean across all eight regions 22.8% stimulus vs. 10% choice) and was also distributed, with significant coding in all forebrain (CP, MOp, MOs) and midbrain (MRF, SCm, APN, RN) regions, but not VPM/VPL. Midbrain

regions had a significantly higher proportion of visual stimulus coding neurons than the forebrain (mean 31.3% neurons in midbrain, 19.1% in forebrain, $p = 2.4 \times 10^{-11}$, chi-squared test). In the midbrain and the forebrain, neurons exhibited mixed selectivity for stimulus and choice (Fig. 3.3F, 6.4% of neurons jointly encoding in midbrain, 3.8% expected based on independent probabilities; 4.2% and 1.8% in forebrain).

A significant proportion of neurons in MRF, SCm, and MOp had contextually modulated premotor activity (Fig. 3.3F, Supp Fig. S4C). Neurons with contextually modulated responses (also referred to as “trial context” hereafter) were rarer than stimulus or choice encoding neurons and were encoded in similar proportions across MRF, SCm, and MOp (4.8%, 6.2%, 5.0% of total neurons, 14.0%, 13.5%, 18.9% of coding neurons). Surprisingly, MOs and CP did not code for trial context above chance level (Fig. 3.3F, Fig. S4C). The Context-Stim and Context-Premotor kernels were shuffled both together and separately to assess the overall contribution of context to trial responses and the stimulus vs. premotor nature of the modulation. Context almost solely modulated premotor activity, with the combined Context kernels shuffle yielding roughly the same proportion of coding neurons as the shuffle of the Context-Premotor kernel alone (0.2% difference across MRF, SCm, and MOp). Stimulus-aligned activity was not significantly modulated by context in any region (Fig. S4C). As we hypothesized, these results are consistent with context-dependent action attractor dynamics rather than context-dependent shifts in stimulus input angle (Fig. 3.2A).

Neurons that encoded baseline context formed a largely distinct group from those encoding stimulus, choice, or trial context (Fig. S5A-E). To assess whether the same neurons encode multiple task variables, we calculated the joint encoding

probability for variable pairs. In all regions that encode baseline context (MRF, SCm, CP, MOs), the chance that neurons encoding stimulus, choice, or trial context also encoded baseline context (and vice versa) did not exceed what would be expected if the coding of each variables was independent (Fig. S5D). In contrast, there was a strong overlap between trial context and choice-coding neurons: In MRF, SCm, and MOp, neurons encoding choice were significantly more likely to also encode trial context (and vice versa) than expected by chance (Fig. S5B,C). This finding supports our observation that contextual modulation of trial responses specifically affects premotor activity, putatively resulting from shifting action attractor strengths with context (Fig. S4B, Fig. 3.2).

3.4 Discussion

Here, we investigated the neural correlates of flexible, context-dependent decision-making by performing large-scale electrophysiological recordings across multiple brain regions, including dense sampling of the MRF, during a novel visual stimulus-action remapping task. Our analyses revealed that MRF, in a network with MOs, CP, and SCm, contains robust representations of the abstract task context. State-space trajectory analysis showed that population activity in these four regions, projected onto choice and context dimensions, separated along the context dimensions during the baseline period, i.e. prior to stimulus onset. This baseline separation appears to set the initial conditions for a context-dependent shift in the subsequent action dynamics, enabling the identical center stimulus trials to flow toward opposite motor actions in the choice dimensions (CW vs. CCW) (Fig. 3.2, Fig. S2). Complementary

single-neuron analyses using kernel regression supported the mechanism of a context-dependent shift in action attractor dynamics, as premotor activity, but not center stimulus-related activity, was modulated by context (Fig. 3.3, Fig. S4).

The mechanism neural populations in MRF, SCm, MOs, and CP used to solve the context-dependent stimulus-action remapping problem is analogous to the dynamical mechanism found in primate PFC to enact context-dependent sensory selection in the landmark study by Mante & Sussillo et al. (Mante et al., 2013). They found that context-dependent selection and evidence accumulation were both present in the population dynamics of PFC: A context dimension orthogonal to each sensory dimension (color and motion) effectively rotated the dynamical landscape to integrate sensory evidence along one dimension over the other. While the stimulus-action remapping task here did not require evidence accumulation, the context dimension was similarly orthogonal to the choice dimensions and putatively set the initial conditions to shape the ensuing dynamical trajectories in the choice dimensions, rather than simply manifesting as a baseline motor bias in the choice dimensions towards one action or the other (Fig. 3.2G, Fig. S2G-I). Crucially, as in Mante et al. 2013, we did not find evidence for a top-down “early” gating or modulatory mechanism for visual stimuli in any of the recorded regions (Fig. 3.2A,D) (Noudoost et al., 2010). This argues against a feedforward scheme wherein sensory input is modulated by context in upstream regions and subsequently fed to downstream motor and associative areas. Notably, the dynamical mechanism we observed here was distributed across a network encompassing midbrain (MRF, SCm) and forebrain (MOs, CP) structures, all of which are known to be densely and recurrently connected (Foster et al., 2021b; Oh et al.,

2014). Thus, rather than purely relying on the internal recurrent connectivity of cortex, the results here instead support a model wherein multiregional recurrent interactions across the forebrain and midbrain collectively establish the context-appropriate dynamical state (Okazawa and Kiani, 2023; Wang, 2008).

The initial conditions of the population dynamics in MRF, MOs, CP, and SCm were in large part set by stable baseline context representations. Using a careful task design and robust null distributions based on pseudosession statistics (Harris, 2020), we observed significant baseline context representations in both single neuron firing rates and via decoding at the population level (Fig. 3.2E-F, Fig. S2D). Such sustained, context-related activity during baseline periods has been observed in other tasks requiring tracking of hidden states or rules, such as in mPFC and primary auditory cortex during an auditory stimulus selection task (Rodgers and DeWeese, 2014) or in mPFC during a dynamic foraging task where signals reflect relative action values (Bari et al., 2019). As described in **Chapter 2**, the signal here is likely representative of an abstract rule belief rather than being directly related to motor output or stimulus/action values. This representation is also distinct from the widely distributed subjective prior expectation signals found in the recent International Brain Laboratory brain-wide map companion study (Findling et al., 2023), as the imbalance in rewarded actions between blocks may similarly represent value-based or motor planning-related processes.

Though baseline context decoding in MRF, SCm, CP, and MOs was highly significant in aggregative statistics across sessions, it was not decodable in every single recording session within each region (Fig. 3.2F). This session-to-session variability is a common observation in large-scale electrophysiological studies of decision-making

tasks (International Brain Laboratory et al., 2023a) and could arise from multiple factors. Crucially, we found it was not explained by differences in behavioral performance (within the included range of proficient sessions), the number of neurons within a population, nor context-correlated uninstructed movements (Fig. S3). Potential contributing factors include variations in probe placement within the broader region (explored further in **Chapter 5**), inherent noise in neuronal firing and recordings, natural variations in internal state between subjects and sessions, and the often sparse coding of cognitive and task representations, wherein only a small subset of neurons within a population carry the relevant signal (Hanks et al., 2015; International Brain Laboratory et al., 2023a; Steinmetz et al., 2019; Waydo et al., 2006). While our current analyses did not reveal significant differences in the basic trajectory shapes or baseline context decoding strength between MRF, MOs, SCm, and CP, it is possible that more sophisticated methods allowing for analysis of single-trial population dynamics, such as Latent Factor Analysis via Dynamical Systems (LFADS) (Pandarinath et al., 2018), might uncover more subtle distinctions in how these regions contribute to the overall computation. It is notable, however, that other recorded regions, namely MOp, APN, RN, and VPM/VPL, did not show significant baseline context coding, providing some specificity to the core network identified (Fig. 3.2F).

Our single-neuron encoding analyses revealed two key differences amongst regions. First, midbrain regions (MRF, SCm, APN, RN) exhibited significantly stronger coding of the visual stimulus compared to forebrain areas (MOs, MOp, and CP), echoing recent findings of particularly enhanced visual representations emerging in MRF after training of a visual evidence accumulation task which were linked to

sensorimotor transformation (Fig. 3.3F) (Khilkevich et al., 2024). Though our task lacked an evidence accumulation component, the significant proportion of midbrain neurons coding for both stimulus and choice may reflect a similar role in linking the stimulus identity to the motor plan, especially for outside stimulus trials which had a fixed motor contingency (Fig. S5A-B). Second, while baseline context coding was found in MRF, SCm, MOs, and CP, the contextual modulation of ongoing premotor activity during the trial was primarily observed in MRF, SCm, and MOp (but not MOs or CP). These context-modulated premotor neurons significantly overlapped with choice-selective neurons but formed a distinct population from those encoding baseline context (Fig. 3.2F, Fig. S5B-E). This suggests a potential functional division within the recorded regions: one set of populations (in MRF, SCm, MOs, CP) establishes the stable baseline context representation, which then influences a separate set of populations closer to motor output pathways (in MRF, SCm, MOp) to implement the context-dependent shift in action selection dynamics.

There are several limitations of the study presented here. First, our analysis relies on observing population trajectories consistent with attractor dynamics but does not provide direct proof of attractor states. More definitive evidence could be borne out of dynamical systems modeling such as Recurrent Switching Linear Dynamical Systems (rSLDS) (Glaser et al., 2020) and perturbation-based methods that directly test the stability properties of putative attractors (Inagaki et al., 2019). Most significantly, the current study lacks causal manipulations, precluding definitive conclusions about the necessity of the contextual representations in MRF or other regions for performance of the task. However, the presence of robust, specific context signals during the baseline

period, which was equally present in canonical flexible decision-related regions as it was in MRF and absent in control regions, supports MRF, SCm, MOs, and CP each being key nodes in the context computation circuitry rather than passive relays of the signal. We additionally acknowledge that other brain regions not recorded here are likely involved and warrant investigation. mPFC is a prime candidate, given its established role in rule representation and cognitive control in related paradigms (Duan et al., 2015; Huda et al., 2020; Liu et al., 2014; Merre et al., 2021; Reinert et al., 2021; Rodgers and DeWeese, 2014). Orbitofrontal cortex (OFC) is another candidate, although its function might relate more strongly to the value-based associative learning aspects of task acquisition rather than the dynamic rule application focused on here (Miller et al., 2022; Padoa-Schioppa and Assad, 2006; Wilson et al., 2014). Moreover, areas connected with and related to the MRF—pontine and medullary aspects of the reticular formation—also showed intriguing task correlates in recent brain-wide surveys and could potentially contribute to integrating context with motor output pathways (Chen et al., 2024; International Brain Laboratory et al., 2023a).

Overall, these results provide the first direct evidence of MRF computing abstract contextual rules during flexible visual decision-making. They suggest that the MRF participates within a distributed network alongside SCm, MOs, and CP to establish context-dependent shifts in sensorimotor transformations. This points towards a higher-order cognitive function for the MRF than traditionally ascribed. Future studies involving spatially targeted causal manipulations (discussed further in **Chapter 6**), detailed analysis of functional organization within MRF (which we begin to conduct in

Chapter 5), and broader brain-wide recordings will be crucial for fully elucidating the MRF's specific contributions to the circuitry underlying mammalian flexible behavior.

Chapter 4: The MRF develops behaviorally relevant visual stimulus representations as a result of task learning

4.1 Introduction

Early studies recording from MRF neurons found responses to a diversity of sensory stimuli (Bell et al., 1964; Büttner-Ennever and Büttner, 1978; Scheibel et al., 1955; Steriade and Glenn, 1982). Additionally, three recent studies performing brain-wide recordings of neural activity during performance of visual decision-making tasks found significant representations of visual stimuli in the MRF (International Brain Laboratory et al., 2023a; Khilkevich et al., 2024; Steinmetz et al., 2019). One of these studies found the visual representations in MRF to only be present in trained, but not task-naïve, mice (Khilkevich et al., 2024). However, it remains unclear whether the visual responses in MRF of trained subjects are representations of the specific task stimuli or reflect MRF neurons more generally processing the sensory attributes of visual stimuli due to visual task training. To answer this question, we perform passive stimulus presentation experiments, containing both task and non-task stimuli, in task-naïve mice and mice trained on the flexible visual stimulus-action remapping task while densely recording from MRF and SCm.

4.2 Methods

The methods below consist of technical details of the passive stimulus presentation protocol and statistical approaches to quantify neural encoding of the stimuli. Context and rationale for use of these methods are described in the results section (**Chapter 4.3**).

4.2.1 Passive stimulus presentation protocol

At the end of each task recording session (from **Chapter 3**), we performed a passive stimulus presentation experiment while continuing to record from the same electrodes. This experiment was also displayed for recordings from task-naïve mice. Mice were presented with stimuli from the task: full contrast gabors at -80° , 0° , and 80° , go cue beeps, reward tones, and white noise bursts. The gabor stimuli were also presented at azimuths in between those present in the final stage of the task (-80° to 80° with 13.33° step increments in between). Gabor stimuli lasted for 0.5 seconds. These stimuli were replayed at 1-3 s randomized intervals for 20 randomly interleaved repetitions each.

4.2.2 Quantification of stimulus encoding during passive task presentation

To robustly assess the encoding of passively presented visual and auditory stimuli (Fig. 4.1, Fig. S6), we used the Zenith of Event-based Time-locked Anomalies (ZETA) test (Montijn et al., 2021) for each neuron and stimulus type (Matlab code, <https://github.com/JorritMontijn/ZETA>). The ZETA test is parameter- and binning-free, generating a cumulative distribution of spike times relative to stimulus onset as a function of time and measuring deviation from a linear baseline. The ZETA test was applied using a 1 second window after stimulus onset, with 60 trials per stimulus type (for visual stimuli, each stimulus azimuth was considered separately, and each of 3 contrasts were considered together (0.25, 0.5, 1)). P values were obtained by comparing to a null distribution generated by jittering stimulus-onset times ($n = 500$ jitters). To broadly assess whether a neuron encoded visual stimuli irrespective of

azimuths (Fig. 4.1D), we controlled the false discovery rate (FDR) by applying the Benjamini-Hochberg procedure to the p-values from each azimuth. We then used the minimum FDR value among azimuths to determine stimulus encoding.

4.3 Results

The MRF and SCm exhibited enhanced responses to visual, but not auditory, task stimuli in a passive context following training. To assess the effects of task training on MRF stimulus responsivity, we performed a separate set of Neuropixels 2.4 recordings in task-naive mice wherein a passive stimulus protocol was presented (Fig. 4.1A). The protocol was also displayed at the end of each recording session in trained mice. Stimuli present in the task (left, center, and right gabors; go cue, miss noise, and reward tone) were displayed in random order and timing and could not be moved by the wheel. Gabor stimuli at intermediate azimuths (i.e., between -80° , 0° , and 80°) were also displayed. Randomly timed sucrose rewards were given to ensure alertness. Similar to trained mouse recordings, recordings in task-naive mice were targeted to densely tile MRF to assess the spatial dependency of variable coding (Fig. 4.1B). We found that there were significantly more neurons in the MRF of trained mice which responded to gabor stimuli compared to task naive mice (Fig. 4.1C,D 36/717 neurons ZETA (Montijn et al., 2021) $p < 0.01$ in trained mice, 4/666 neurons in task-naive mice, $p = 9.5 \times 10^{-7}$ chi-squared test). MRF gabor stimulus encoding was approximately at chance level (0.6% coding neurons, chance level is 1% at $p = 0.01$) in task-naive mice, suggesting that stimulus responsiveness in MRF occurs as a result of task training. In contrast to visual stimuli, the MRF did not increase its responsiveness to auditory stimuli

with task training: the reward tone was coded near chance in both task-naïve and trained mice (Fig. 4.1D, 8/717 trained, 8/666 task-naïve, $p=0.88$). Similar to MRF, neurons in SCm also showed enhanced gabor stimulus responsiveness with training (24/370 trained, 1/224 task-naïve, $p=3.8 \times 10^{-4}$) and a small decrease in reward tone responsiveness (2/370 trained, 6/224 task-naïve, $p=0.029$). The training-induced visually responsive neurons were broadly distributed throughout the MRF and SCm (Fig. 4.1E).

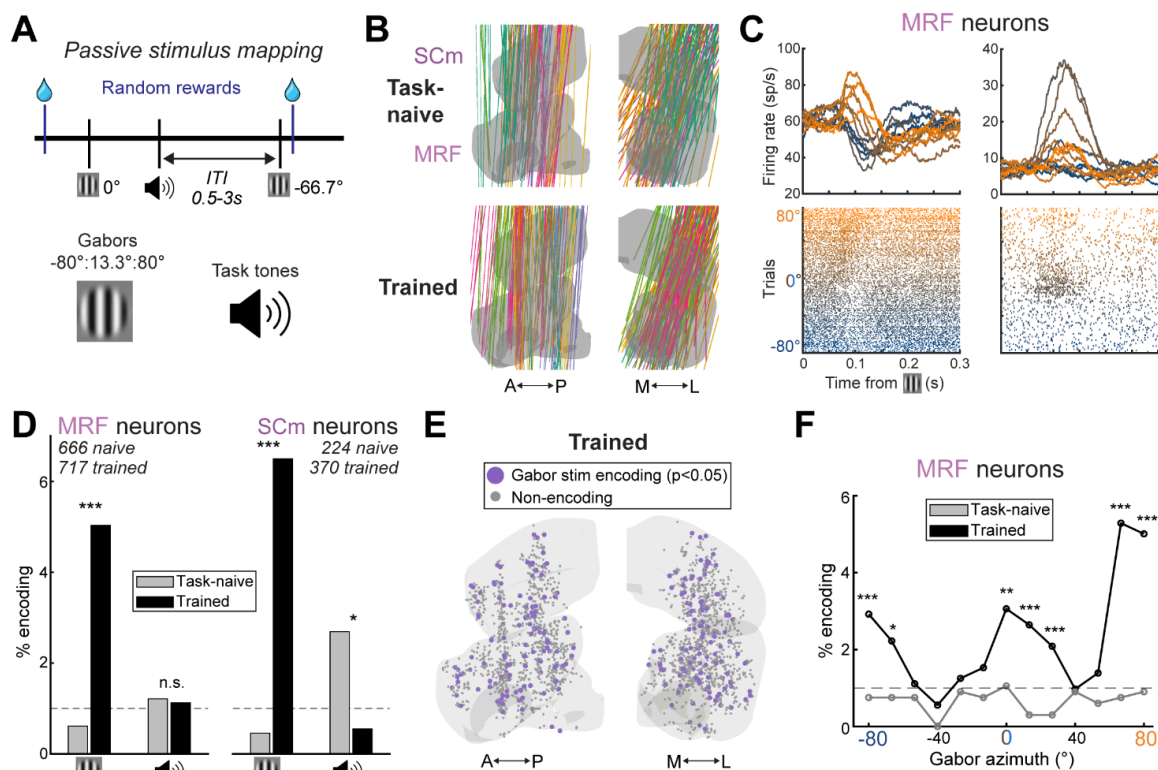


Figure 4.1. Task-specific visual stimulus representations emerge across MRF and SCm as a result of task training. **A)** Passive stimulus mapping protocol schematic. Task stimuli (visual: gabors at -80° , 0° , or 80° , audio: go cue, reward tone, miss noise) and control stimuli (visual: gabors at intermediate azimuths, expanding or contracting black discs, sparse noise receptive field mapping) were displayed in random order with random timing, and randomly timed sucrose rewards were given. **B)** Lateral (left) and anterior (right) 3D view of Neuropixels 2.4 recording insertions tiling SCm and MRF in task-naïve (top) and trained (bottom) mice. **C)** PSTH and ZETA plot of an example visually responsive SCm neuron. Left: Mean $\pm$ SEM PSTH spike raster of PSTH aligned to passively presented 80° gabor stimulus. Right: ZETA test plot showing the deviation of the cumulative fractional spike positions of the true data vs. a jittered null, from which the ZETA p-value is derived. **D)** Percentage of neurons in MRF (left) or SCm (right) encoding gabor stimuli (all azimuths) or reward tone in task-naïve and trained mice. P values are derived

from the ZETA test (Montijn et al., 2021). **E)** Lateral (left) and anterior (right) 3D views of locations of gabor stimulus encoding and non-encoding neurons in MRF and SCm. 50 μm jitter was added to AP and ML coordinates of neurons for visualization. **F)** Percentage of MRF neurons encoding gabor stimuli split by azimuth in task-naïve or trained mice. Above chance coding was only observed in trained mice, at or near task stimulus locations (-80° , 0° , 80°). P values are derived from chi-squared test between trained and task-naïve mice at matched azimuths.

The enhanced visual stimulus responsiveness in MRF and SCm was specific to behaviorally relevant task stimuli. Gabor stimuli were presented at 13 different azimuths ranging from -80° to 80° during the passive stimulus protocol. Strikingly, MRF stimulus encoding was enhanced after training only at the azimuths at or near those present in the task (Fig. 4.1F, $p < 0.05$ at left -80° , center 0° , and right 80° stimuli as well as -66.7° , 13.3° , 23.7° , and 66.7°). This same pattern of stimulus responsiveness was present in SCm but not forebrain regions, where passively presented stimuli were coded near chance (Fig. S6B,C). This finding corroborates the enhanced midbrain encoding of visual stimuli during the task (Fig. 3.3F, Fig. S6).

Behaviorally relevant visual stimulus responsiveness in MRF was not caused by stimulus-evoked movements. To test whether the passively presented stimuli evoked wheel turns or movements in trained mice, we compared face video motion energy and rotary encoder traces before and after stimulus onset. We found that of the gabor stimuli present in the task, only the right stimulus significantly evoked movement responses on just a subset of trials (Fig. S6A, $p < 0.05$ signed-rank test baseline vs. post-stimulus period). The left and center stimuli did not significantly evoke movement responses. Thus, it is unlikely that visual responses to the behaviorally relevant stimuli during the passive task are a result of stimulus-evoked movements.

4.4 Discussion

In this chapter, we investigated how training on a flexible visual decision-making task altered sensory representations within the MRF and the adjacent SCm. By comparing neural responses during passive stimulus presentation in task-naïve versus trained mice, we found a significant emergence of visual responsiveness specifically in the trained subjects. This training-induced plasticity was selective for visual stimuli; responsiveness to task-related auditory stimuli (reward tones) did not differ between naïve and trained groups in the MRF (Fig. 4.1D). Furthermore, this enhanced visual coding was strikingly specific to the spatial locations (azimuths) of the visual stimuli used during the active behavioral task (-80° , 0° , $+80^\circ$ azimuths), with little to no enhancement for intermediate, untrained locations (Fig. 4.1F). Importantly, control analyses indicated these visual responses were unlikely to be driven by stimulus-evoked movements (Fig. S6A). These visually responsive neurons appeared broadly distributed throughout the sampled regions of MRF and SCm in trained animals (Fig. 4.1E).

These findings demonstrate that the MRF undergoes significant functional plasticity as a result of task learning, developing representations tuned to behaviorally relevant visual inputs. This complements the findings from **Chapter 3**, which highlighted the MRF's role in rapid, dynamic computations related to abstract context within a well-learned task, by suggesting the MRF also exhibits longer-term adaptations reflecting task experience. The task experience-dependent adaptations likely account for recent findings of high proportions of MRF neurons encoding visual stimuli in mice trained on visual decision-making tasks (International Brain Laboratory et al., 2023a;

Steinmetz et al., 2019). Our results align with and extend recent observations by Khilkevich et al. (Khilkevich et al., 2024), who also noted emergent visual responses in MRF following training on a visual evidence accumulation task. The key novelty here is the demonstration of visual responses in a *passive* setting. Khilkevich et al. showed that the increased visual responses after task training in MRF, as well as in other associative and motor regions, were present during task performance and were functionally related to the visuomotor transformation required by the task. The additional presence of visual responses in a passive setting observed here is consistent with a potential subcortical plasticity of visual pathways (Duménieu et al., 2021) to produce visual responses in MRF. We further demonstrate that these visual responses are tuned specifically to the learned task stimuli locations, suggesting they represent more than just a general increase in visual sensitivity. Instead, they may relate to the behavioral significance or learned associations of specific stimuli within the task context.

As suggested previously (Khilkevich et al., 2024), the emergence of these task-specific visual representations in midbrain structures like MRF and SCm likely reflects their roles in sensorimotor transformation—linking relevant sensory inputs to appropriate motor outputs. The tuning to specific task stimuli could facilitate the rapid selection context-appropriate actions upon stimulus presentation. Future studies could further probe the nature of these representations by testing responses to a wider range of visual stimuli varying in features unrelated to the learned task dimensions (e.g., orientation, spatial frequency) and by performing detailed receptive field mapping to characterize the tuning properties of these emergent visually responsive neurons more comprehensively. Additionally, to more definitively determine the source of the passively

evoked, training-induced visual signals in MRF, future work could employ pathway-specific manipulations targeting potential upstream visual structures, such as the superficial layers of the SC or the visual cortex, to pinpoint which pathway(s) specifically mediate this effect.

Chapter 5: Context coding and cortical axon terminals spatially align in MRF

5.1 Introduction

Having found that MRF codes for abstract context in a network with MOs, SCm, and CP during flexible visual decision-making (**Chapter 3**), an important question is raised: Is the context coding in MRF spatially organized, and if so, what are the organizing principles? As reviewed in **Chapter 1.2**, the MRF is large, poorly differentiated, and heterogenous, raising the intriguing possibility that subregions specialized for different functions may be present within the broadly defined structure, similar to the functionally defined subregions of MLR in locomotion (Ferreira-Pinto et al., 2021) and cMRF in saccades (Graf and Ugolini, 2006).

Given that the MRF receives substantial cortical input (Gao et al., 2022; Oh et al., 2014; Peng et al., 2021; Winnubst et al., 2019) and appears to compute context alongside frontal cortex, a direct hypothesis emerges: the specific MRF subregions encoding context should correspond spatially to the areas receiving direct cortical axon terminals. This spatial alignment would putatively facilitate the feedforward/feedback interactions characteristic of the recurrent computations underlying flexible decision-making processes (Foster et al., 2021a; Pasupathy and Miller, 2005; Steinmetz et al., 2019). To test this hypothesis, here we examine the spatial distribution of baseline context-coding neurons identified in our dense recordings within the MRF. We then compare this functional map to the anatomical distribution of cortical axon terminals obtained from reconstructed neuron morphologies, assessing the degree of spatial alignment between context representation and cortical afferents.

5.2 Methods

The methods below consist of technical details about the preprocessing of open single neuron morphology datasets and statistical approaches for assessing non-uniformity and alignment of context coding neurons and cortical terminals in MRF. Further context and rationale for use of these methods are described in the results section (**Chapter 5.3**).

5.2.1 Preprocessing of single neuron morphology datasets

We integrated three open datasets of fully reconstructed single neuron morphologies (Gao et al., 2022; Peng et al., 2021; Winnubst et al., 2019), using custom analysis code in Matlab to analyze the files in a SWC format (<https://swc-specification.readthedocs.io/en/latest/swc.html>). The datasets primarily sampled cortical neurons, therefore our analyses focused solely on these cells. We only considered soma and axon terminal locations. To determine if a point was an axon terminal, we filtered for points whose type was axon (column 2 “Type” == 2) and which weren’t a “parent” to any other sample (no entry in column 7 “Parent”, i.e. no other points branched from this point). As the coordinates for all points were aligned to the CCF, we downsampled the coordinates to 10 um resolution and filtered the datasets for neurons whose somas were anywhere within the CCF-defined cerebral cortex and with at least 1 axon terminal in the CCF-defined midbrain reticular nucleus. These neurons and the CCF coordinates of their axon terminals were then analyzed for their alignment with spatial coding in MRF.

5.2.2 Non-uniformity testing of cortical terminals

To test whether cortical axon terminals exhibited non-uniform distribution within the MRF (Fig. 5.1C), we developed a statistical approach comparing observed terminal density distributions against a uniform null hypothesis. We first identified all axon terminals within the MRF from cortically-projecting neurons. For each spatial dimension (anterior-posterior i.e. sagittal, dorsal-ventral i.e. horizontal, medial-lateral i.e. coronal), we created histograms of terminal density in MRF slices using a sliding window with bin width of 250 μm and step size of 150 μm . Terminal density was calculated as the number of terminals within each slice divided by the MRF volume (in mm^3) within that slice.

The metric we used to test distance from a uniform distribution in each spatial dimension was to compute the sum of squared deviations of the observed number of terminals in each slice from the expected number of terminals based on a uniform density. The expected density based on a uniform distribution was calculated as the total number of terminals divided by total MRF volume. This density was then multiplied by the volume of a given slice to get the expected terminal count. The advantage of calculating deviations between observed and expected counts versus observed and expected densities is that regions with higher sampling are given more weight in the calculation (this especially important for more uneven sampling of recording probe locations, as described next for baseline context coding spatial distribution).

To assess the probability that the observed distances from uniform arose by chance, we formed a null distribution of randomly positioned cortical axon terminals in MRF. For each of 1,000 permutations, we randomly sampled positions from all possible

MRF voxels to create a surrogate terminal distribution with the same number of terminals as in the original dataset. For each permutation, we calculated terminal densities and distances from uniform using the same sliding window and sum of squared deviation approaches. This provided an empirical null distribution of deviation scores. This test was performed both for the complete terminal dataset and for a subsampled version containing one terminal per neuron to account for potential bias from the non-uniform branching patterns of neurons with many MRF terminals. In each case, cortical terminal distributions were significantly non-uniform across all three spatial dimensions.

5.2.3 Non-uniformity testing of baseline context coding

To determine whether neurons encoding baseline context showed spatial organization within the MRF (Fig. 5.1D), we performed non-uniformity testing similar to that applied to cortical terminals. Neurons were classified as encoding baseline context if they exhibited significant correlation ($p < 0.05$ against pseudosession null) between their baseline firing rate and task context (Fig. 3.2E, Fig. S2D). The spatial distribution of these context-encoding neurons was assessed using the same sliding window approach (bin width of 250 μm , step size of 150 μm) applied separately along anterior-posterior, dorsal-ventral, and medial-lateral dimensions.

For each MRF slice, we calculated the proportion of context-encoding neurons relative to all recorded neurons. The distance from a uniform distribution was again computed as the sum of squared deviations between the observed and expected counts of coding neurons in a slice. The expected uniform density was the overall fraction of

neurons in MRF encoding baseline context, and the expected count was this density multiplied by the number of recorded neurons in a slice. To assess the probability that the observed distances from uniform arose by chance, we performed 1,000 permutations. In each permutation, we randomly reassigned the "context-encoding" label among all recorded neurons while maintaining their spatial positions, then recalculated the proportion of context-encoding neurons and distances from uniform in each slice.

5.2.4 Voxelization and spatial alignment of cortical terminals and baseline context coding

To directly compare the spatial distributions of cortical terminals and baseline context-encoding neurons (Fig. 5.1E), we implemented a voxelization approach that partitioned the MRF into overlapping spherical voxels (radius 125 μm , grid spacing 150 μm). These voxels were composed of varying numbers of CCF-defined subvoxels (10 μm^3) depending on MRF topography around the voxel. To avoid edge artifacts, we excluded volumes containing fewer than 50 CCF-defined MRF subvoxels.

For each voxel, we calculated: (1) density of cortical axon terminals (terminals per mm^3), and (2) proportion of neurons encoding baseline context. Voxels with no recorded neurons in them were excluded. As before, we calculated expected terminal densities by dividing the total number of cortical terminals in MRF by total MRF volume. We calculated expected coding proportions by dividing the total number of coding neurons by the total recorded neuron count in MRF. Expected terminal and coding neuron counts were then calculated by multiplying the expected densities by the volume or total number of recorded neurons within each voxel. We then computed the distance

from uniform for each voxel using the sum of squared deviation of observed counts from expected counts as described above.

To assess the degree of spatial alignment between cortical terminals and baseline context coding, we computed Spearman's rank correlation coefficient (ρ) between the distances from uniform of terminals and context-coding neurons across voxels. Statistical significance was determined using the same permutation approach for context coding neurons as described above, wherein the "context-encoding" label was randomly reassigned amongst all recorded neurons 1,000 times. Voxel context coding proportions, deviations from uniform, and Spearman's ρ were recalculated for each permutation to form a null of ρ values. This approach preserved the spatial autocorrelation structure of the recorded neurons while testing the specific relationship between terminals and context coding.

5.3 Results

Single neuron morphological reconstructions displayed dense cortical inputs to MRF primarily originating from frontal, motor, and somatosensory cortex. To analyze the degree and spatial distribution of cortical inputs to MRF, we compiled single neuron morphologies from three CCF-aligned open datasets (Gao et al., 2022; Peng et al., 2021; Winnubst et al., 2019). We filtered the neurons in each dataset for those with their cell body in the neocortex and at least one axon terminal within MRF. MRF received prominent inputs from prefrontal (anterior cingulate (ACA), prelimbic (PL), orbital (ORB)), motor, and somatosensory cortical areas (Fig. 5.1A). Terminals from cortical regions with 10 or more neurons projecting to MRF were included for further analysis.

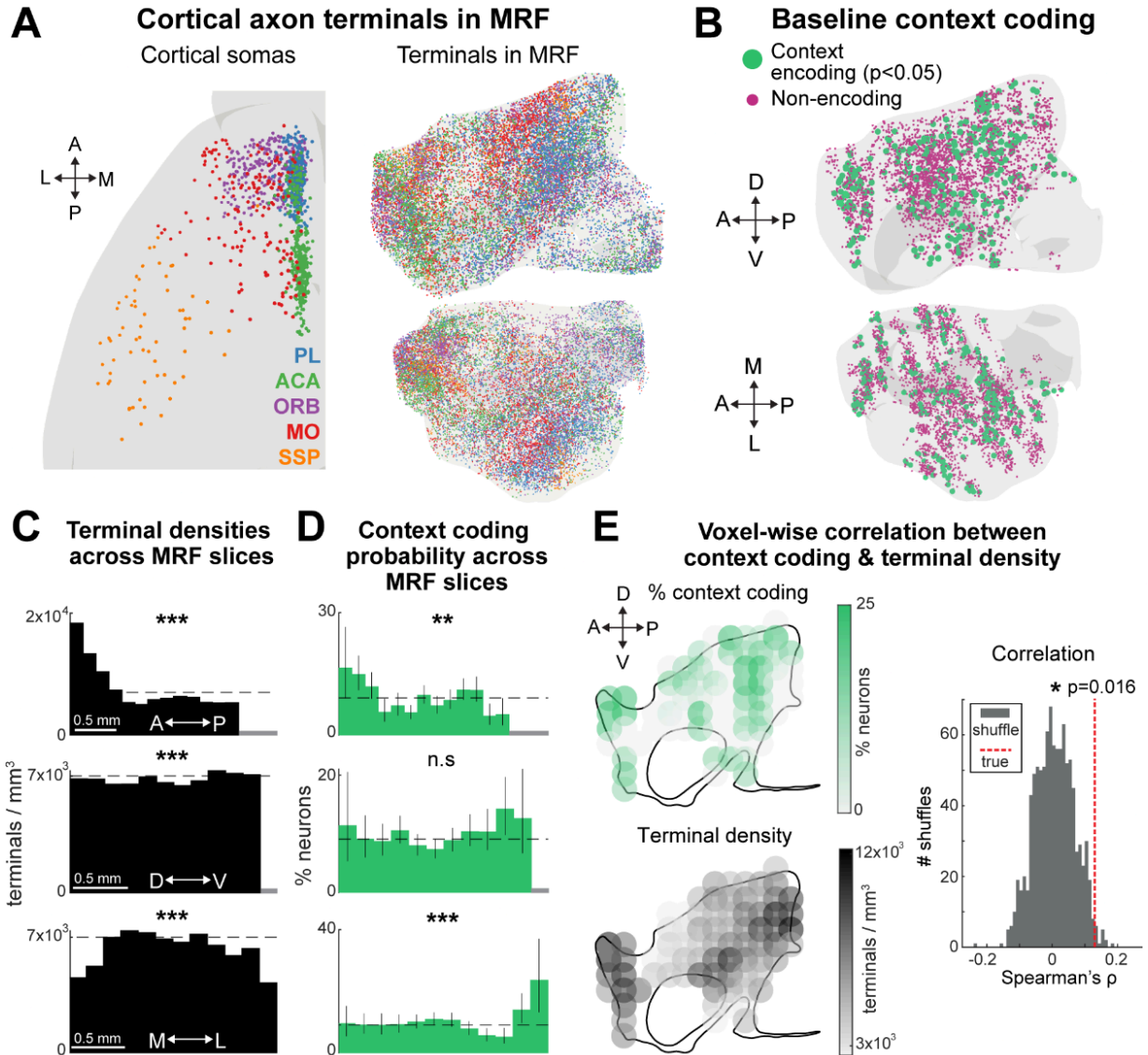


Figure 5.1. Spatial distributions of cortical axon terminals and context coding neurons in MRF are non-uniform and correlated. **A)** Left: Spatial locations of cortical neuron cell bodies which project to MRF (≥ 1 axon terminal in MRF). Each point is a cell body. Right: 3D views of MRF from lateral (top) or dorsal (bottom) with axon terminal locations of cortical neurons plotted. Each point is an axon terminal. Cell bodies and terminals are colored by cortical region from which neuron originates (PL = prelimbic, ACA = anterior cingulate, ORB = orbital area, MO = primary and secondary motor, SSP = primary sensory). Data compiled from open single neuron morphology datasets (Gao et al., 2022; Peng et al., 2021; Winnubst et al., 2019). **B)** 3D views of MRF with recording locations of baseline context coding neurons ($p < 0.05$; Fig. 3.2E & Fig.S2D) and non-encoding neurons. 50 μm jitter was added to AP and ML coordinates of neurons for visualization. **C)** Histograms of cortical axon terminal densities across coronal, horizontal, and sagittal slices of MRF (top, middle, bottom plots). Slice binning is 250 microns wide with 150 micron spacing. Significance stars above plots indicate non-uniformity of distribution (Chapter 5.2.2). **D)** Same as C) but for proportion of neurons encoding context across MRF slices. Error bars are 95% binomial proportion CIs. **E)** Left: Sagittal slice of MRF at 1.04 mm M/L showing proportions of context coding neurons (top) and terminal densities (bottom) per voxel. Voxels had a 125 micron radius and 150 micron separation. Right: True vs. shuffle null Spearman's rank correlation coefficient for correlations

between voxelized context coding proportions and terminal densities throughout MRF. Shuffles consisted of random permutations of context coding neuron locations (Chapter 5.2.4).

The axon terminal projection patterns of cortical neurons into MRF were significantly non-uniform. To characterize whether cortical neurons had a spatially biased or uniform targeting of MRF, we binned the MRF into 250 μm slices with 150 μm spacing along the coronal, horizontal, and sagittal planes and calculated the density of axon terminals in each slice (Fig. 5.1C). To assess non-uniformity, the distance of the density distributions from the expected densities based on the global mean was calculated for each plane. We found that the distances of the spatial densities from a uniform distribution were significant along each plane compared to a null distribution wherein terminal positions were randomly distributed throughout MRF (Chapter 5.2.2, $p < 0.01$ coronal, horizontal, sagittal). The most prominent feature of the non-uniform cortical terminal distribution was the particularly high density in the anterior aspect of MRF (Fig. 5.1C).

Baseline context coding showed a similarly non-uniform spatial distribution in MRF (Fig. 5.1B,D). To determine spatial distributions of neurons encoding baseline context, we performed the same slice binning procedure as before and calculated the proportion of coding neurons within each slice. We performed the same non-uniformity test against a null distribution which was formed by permuting coding neuron labels amongst all neuron recording positions in MRF. The distribution of baseline context coding neurons was significantly in the sagittal and coronal planes (Fig. 5.1D). There were high proportions of baseline context neurons in the anterior and lateral aspects of MRF (Fig. 5.1D).

Baseline context neurons and cortical axon terminals in MRF were spatially aligned (Fig. 5.1E). We divided MRF into 125 micron radius voxels with 150 micron spacing and calculated the density of cortical terminals and proportion of baseline context neurons in each voxel. The voxel-wise context coding proportions and terminal densities were then correlated using Spearman's rank correlation and were compared to a null wherein the labels for context coding vs. non-coding were randomly permuted among the recorded neuron locations. There was a modest but significant correlation (Spearman's $\rho = 0.12$, $p = 0.016$ against shuffle, Fig. 5.1E), with both baseline context coding and cortical terminals sharing increased incidence in the anterior aspect of MRF (Fig. 5.1C,D).

5.4 Discussion

In this chapter, we examined the spatial distributions of abstract context representations and cortical axon terminals within MRF and assessed the relationship between the two. Leveraging our dense electrophysiological recordings across the MRF during the flexible stimulus-action remapping task and open datasets of single neuron morphologies, we generated a multimodal 3D map of MRF task coding neurons and cortical inputs (Fig. 5.1A-B). Our analyses revealed that both the spatial distributions of cortical terminals and neurons encoding baseline context were significantly non-uniform (Fig. 5.1C-D). Furthermore, using voxel-based correlation analysis, we found a significant, albeit modest, spatial alignment between the density of cortical terminals and the proportion of context-coding neurons (Fig. 5.1E). Both features showed enrichment in the anterior aspect of MRF. Our findings provide the first evidence, to our

knowledge, for a spatially organized representation of a cognitive variable like task context within the MRF.

Recalling the circuitry reviewed in **Chapter 1**, the MRF is a node within the recurrent connections between cortex, basal ganglia, and SC, regions collectively implicated in flexible, context-dependent decision-making. The observed spatial alignment between cortical inputs and baseline context coding in the MRF lends support to the hypothesis that direct anatomical connections facilitate the functional interactions necessary for the context computations identified in **Chapter 3**. Specifically, the co-localization suggests that cortical inputs may preferentially target MRF subregions involved in establishing or maintaining the context signal, participating in the broader feedback and feedforward communication in the cortico-basal ganglia-thalamic-midbrain network underlying flexible behavior (Benavidez et al., 2021; Duan et al., 2021; Foster et al., 2021b; Okazawa and Kiani, 2023; Pasupathy and Miller, 2005; Steinmetz et al., 2019). Indeed, the distributed nature of the context computation across MRF, MOs, SCm, and CP (**Chapter 3**) supports the notion that recurrence between these interconnected regions facilitates the collective establishing of context-dependent dynamical states.

The enrichment of both cortical terminals and context-coding neurons in the anterior MRF is intriguing. This finding hints at a potential large-scale organizational gradient within the MRF, transitioning from functions related more to cognitive control or context-setting rostrally, towards more established motor control functions caudally. This aligns broadly with the known caudal location of key motor centers like the MLR (involving the adjacent PPN/CnF) within the MRF territory and the origin of major

descending reticulospinal tracts and other motor-related neurons from more caudal reticular formation levels in the pons and medulla (Arber and Costa, 2022b; Ferreira-Pinto et al., 2021; Roseberry et al., 2016). However, it is crucial to emphasize that our data suggest a loose gradient rather than sharp functional divisions: Baseline context-coding neurons were found throughout the MRF, albeit at lower proportions caudally (Fig. 5.1B,D), and motor-related signals (like choice) are also present rostrally (**Chapter 3**).

These results open several avenues for future investigation into the organizing principles of the MRF, combining spatial mapping of connectivity, molecular cell types, and diverse functional correlates, which we explore further in **Chapter 6**. An immediate next step stemming from the current dataset would be to test whether the termination patterns of specific cortical source regions (e.g., prefrontal vs. motor cortex) show differential alignment with context coding, as well as whether other functions within the MRF (stimulus, choice, and trial context-coding neurons) exhibit non-uniformity and distinct spatial relationships with cortical inputs. Such analyses will be crucial for gaining a more comprehensive understanding of structure-function relationships in this large and poorly differentiated midbrain region.

Chapter 6: General discussion

In sum, the work presented in this dissertation provides multiple lines of evidence supporting the role of MRF as a novel node within the flexible visual decision-making circuitry. We began by developing a novel behavioral task for head-fixed mice specifically designed to require the use of abstract contextual rules to guide sensorimotor remapping (**Chapter 2**). Behavioral analysis and modeling revealed that mice successfully performed this task, employing strategies consistent with tracking a latent context belief variable updated by both stimulus-specific action-outcomes history and implicit contextual cues related to the underlying task structure. Using large-scale electrophysiological recordings, we demonstrated that neurons within the MRF, alongside populations in MOs, CP, and SCm, robustly code for abstract context belief (**Chapter 3**). Analysis of population dynamics suggested a mechanism whereby a baseline context representation potentially sets the initial action attractor states to shape subsequent stimulus-action trajectories within a distributed network across MRF, MOs, CP, and SCm. Further experiments revealed that the MRF also exhibits long-term contextual adaptations in the form of plasticity, developing specific representations of behaviorally relevant visual stimuli locations after task training (**Chapter 4**). Finally, we found that the representation of abstract context within the MRF is spatially organized, showing significant non-uniform distribution and aligning spatially with the axon terminal input patterns from cortex, particularly within the anterior MRF (**Chapter 5**).

Collectively, these findings significantly update our understanding of MRF function. Moving beyond its traditionally ascribed role in basic motor control, this work

provides evidence that is consistent with the MRF's participation in higher-order computation of abstract contextual states driving flexible behavior.

6.1 Updated theory of MRF function and organization

In this section, we discuss a potential theory of specific MRF function within the broader circuitry underlying flexible visual decisions. While grounded in our findings and existing literature, this should be regarded as a highly speculative framework intended to guide future research. We outline potential future directions to test these hypotheses in the subsequent section.

Decades of research have established that cortico-basal ganglia-thalamic loops are critical for learning and executing specific stimulus-action associations or policies (Ding and Gold, 2013; Foster et al., 2021b; International Brain Laboratory et al., 2021; Lee et al., 2020; Pasupathy and Miller, 2005; Xiong et al., 2015). The context-dependent remapping mechanism of these stimulus-action associations has been shown to arise within the population dynamics in cortical regions, especially PFC (Aoi et al., 2020; Mante et al., 2013; Okazawa and Kiani, 2023; Reinert et al., 2021). However, the mechanism to correctly set the dynamical regime in cortex based on the current, often abstract or internally inferred, context remains less clear. For example, examining the classical study from Mante & Susillo et al. 2013, where do the contextual controlling signals come from that rotate the dynamics in PFC to integrate along the color dimension in one context and the motion context in another? Some theories propose that this control function emerges intrinsically within higher cortical areas like the PFC, wherein these areas maintain representations of task rules and exert top-down

control on downstream circuits (Miller and Cohen, 2001). Back during the era of prominence for the reticular formation in the 1950's and 60's, some hypothesized that it was the source of “function setting” (Kilmer et al., 1968). However, the precise source and nature of such context-setting signals remain debated (Fraser and Holland, 2019).

Here, we posit a theory that the MRF may be a key controller of dynamic context-setting, orchestrating the state of forebrain circuits to enable flexible behavior. Crucially, the contextual signal provided by the MRF would not need to determine the precise structure of the ensuing forebrain dynamics. Its primary function would instead be to reliably drive the forebrain network into a distinct region of its state space for each unique context. Establishing these separable states would then allow the appropriate context-specific dynamics, which implement the required stimulus-action mappings, to be independently learned or engaged from these unique starting points. Orthogonal representations of different contexts within MRF itself, as demonstrated here in **Chapter 3**, could be computationally advantageous, minimizing potential interference between the dynamics associated with each context even if the contexts share many underlying features.

The MRF's connectivity strongly supports its suitability for this role. As described in **Chapter 1.2.1**, a unique aspect of MRF's connectivity relative to that of other nodes in the flexible decision-making circuitry is the input profile it receives from SNr. Indeed, the MRF serves as a remarkable convergence point for the parallel output streams of the basal ganglia, receiving input via the SNr from channels that otherwise maintain relative segregation in their thalamic and brainstem targets (McElvain et al., 2021). This convergence theoretically allows the MRF to monitor the outcomes of action selection

and reinforcement learning processes across multiple parallel cortico-basal ganglia-thalamic streams (Foster et al., 2021b; Lee et al., 2020). Combined with the extensive reciprocal connections with SC (Benavidez et al., 2021), which plays a large role in competitive interactions between opposing actions and sensorimotor transformation (Duan et al., 2021; Jun et al., 2021; Lee and Sabatini, 2021; Thomas et al., 2023), as well as inputs from cortex itself (Gao et al., 2022; Oh et al., 2014; Peng et al., 2021), the MRF appears ideally situated to synthesize a comprehensive representation of current contextual state.

MRF's output pathways seem equally well-suited for implementing context-dependent control of forebrain dynamics. The substantial projections to the inhibitory controllers of all thalamic regions, RT and ZI, provide a potent mechanism for globally modulating thalamocortical information flow (Crick, 1984; Kolmac et al., 1998; Kolmac and Mitrofanis, 1998; X. Wang et al., 2020). By adjusting activity in RT/ZI, the MRF could effectively reconfigure dynamics within the forebrain loop towards those appropriate for the current context. Projections to intralaminar (PF, CL) and MD thalamic nuclei offer additional routes to influence cortical state and prefrontal processing (Krout et al., 2002; Oh et al., 2014; Steriade and Glenn, 1982). Of course, the intrinsic dynamics in each of these nuclei to which MRF projects likely significantly contribute to the context-setting signals themselves as well.

Further circumstantial evidence for the theory of MRF posited here can be found in a recent study showing that in an olfactory delayed-match-to-sample task, ALM was necessary for mapping the test odor to the correct action, but crucially, the sample odor serving as context for the test odor-action mapping was specifically represented via

likely external inputs to layer 2 ALM neurons (Wu et al., 2020). Superficial cortical layers receive prominent input from intralaminar nuclei like PF and CL (Weiler et al., 2023), which are precisely the thalamic nuclei most densely targeted by the MRF (Oh et al., 2014) and which also receive significant, topographically organized input from the SNr (McElvain et al., 2021). This convergence is consistent with a hypothetical pathway wherein MRF integrates cortical and basal ganglia signals, transforms these signals into overarching contextual states, and relays these states via modulatory projections to RT/ZI and routing projections through intralaminar thalamus to superficial cortical layers. This would provide the necessary contextual information to enact the appropriate sensorimotor mappings within circuits in deeper cortical layers.

Our finding of orthogonal baseline context representations in MRF, MOs, SCm, and CP (**Chapter 3**) is consistent with the hypothesized role for MRF here, but it does not provide any additional evidence for it beyond establishing MRF as one of the nodes coding for abstract contextual information. Indeed, we did not observe any significant differences between the context representations or dynamics amongst the four regions, nor did we perform causal perturbations to test whether MRF could change activity along the context dimension in forebrain regions. The finding that context coding spatially aligns with cortical inputs in anterior MRF (**Chapter 5**) supports the idea that the parts of MRF receiving more cortical information play more of a role in establishing this contextual state. Providing direct evidence for this theory will require more sophisticated experiments and analysis approaches, which we review in the next section.

The dynamic context-setting view of MRF posited here does not necessarily exclude previous functional accounts of the MRF but rather can reframe them. Part of the MLR's function could be viewed as establishing a “locomotor” context for the rest of the brain, and indeed, subthreshold stimulation of the MLR has been found to increase the gain of visual cortical responses akin to those seen in a locomoting state (Lee et al., 2014). Similarly, stimulation of a subset of thalamus-projecting cells in MRF and PPN during a “go” cue-triggered delayed licking task was found to reorganize dynamics in ALM by shifting the state from motor planning to motor execution (Inagaki et al., 2022). Our findings suggest that beyond these roles, the MRF, especially parts interacting heavily with cortex, may engage in setting more abstract cognitive contexts critical for flexible decision-making.

6.2 Future directions and implications

As outlined in **Chapter 1.2.2**, the MRF had historically been a central focus of neuroscientists in the 1950's and 60's but subsequently experienced a sharp decline in interest in the following decades (“International Symposium: Reticular Formation of the Brain,” 1957; Scheibel, 1984). The work presented in this dissertation, implicating MRF in higher-order contextual processing beyond the motor-related functions typically ascribed to it, we hope will serve as a catalyst for a resurgence in interest and new studies on this poorly understood brain region. As such, there are a myriad of future directions related to this work, many more than we can focus on here. Indeed, there is little known about fundamental aspects of MRF including its local connectivity and the resulting internal dynamics, the morphology, dendritic structure, and biophysics of MRF

neurons, etc. Here, we will lay out more concrete future directions related to the MRF's role in context setting and flexible decision-making as well as its spatial organization.

A immediately fruitful line of future research would be a fine-grained parcellation of the MRF across multiple modalities. The MRF, serving as a “negative space” of the midbrain with its poorly delineated cytoarchitectural boundaries drawn around other key structures, is likely to be diverse in terms of function, cell types, and connectivity. Indeed, our results hint at anterior-posterior functional gradients (**Chapter 5**), and spatial transcriptomic data suggest over 700 cell types may exist with MRF, about which almost nothing is known (Yao et al., 2023). A concerted effort combining high-resolution input-output connectivity mapping (tracing specific cortical sources, basal ganglia channels, or thalamic/brainstem targets), spatial transcriptomics for defining molecular cell types, and expanded functional mapping with electrophysiological recordings as we have begun to do here (**Chapters 3, 5**) would likely yield great insight into the diverse roles MRF may play within a variety of neural circuits and computations. Integrating these datasets could reveal distinct subregions, interleaved cell populations, and/or gradients within MRF with unique connectivity, cell types, and functional properties. Crucially, this approach would also provide anatomical and genetic targets for specific manipulations within MRF, which are likely to be significantly confounded if performed in a non-specific manner as we describe below. Overall, parcellating MRF based on functional, anatomical, and genetic characteristics would move beyond treating it as a “negative space” of the midbrain to establish a mechanistically grounded organizational framework which would reshape our understanding of brainstem circuitry more broadly.

Rigorously testing the causal role of the MRF in context processing during flexible decisions is crucial but also requires overcoming significant experimental hurdles. The MRF's deep location beneath critical structures for decision and motor-related processing like the SC, its involvement in basic motor functions, and the prevalence of fibers of passage, many of which contain significant neuromodulatory influence, would make conventional manipulation techniques (lesions, non-specific pharmacology or optogenetics via implanted fibers) difficult to interpret unambiguously within the likely subtle and nuanced behavioral deficits needing to be measured for disruption of specific higher-order cognitive functions. In other words, observed overt behavioral deficits could easily stem from damage to overlying areas from the implanted fiber, disruption of passing axons, or general effects on alertness or motor control, potentially overshadowing and confounding the intended specific impairments of contextual processing induced by the manipulation.

Addressing these challenges requires more advanced techniques offering a high degree of target specificity. For example, leveraging a potential parcellation of MRF described above, cell-type-specific manipulations using Cre-driver lines combined with targeted viral delivery of opsins would allow perturbation of molecularly defined MRF populations, ideally those which are more aligned with the function of interest, while sparing fibers of passage (Yizhar et al., 2011; Zeng and Madisen, 2012). Similarly, projection-specific targeting, using retrograde viruses injected into downstream targets like RT, ZI, intralaminar thalamus, SC, etc. and/or anterograde viruses injected into specific MRF subregions to express opsins in defined output pathways, could test the differential roles these pathways in mediating contextual processing. Combinatorial

approaches targeting inputs of MRF as well could be used, perturbing MRF populations with distinct input/output connectivity (Inagaki et al., 2022). This approach would be especially powerful when combined with recordings in other nodes of the circuit, e.g. MOs, to test the hypothesis that MRF perturbation impairs contextual processing elsewhere within the network.

While moving in the right direction for specificity, the above methods still do not address the confounds of implanted fibers significantly damaging key overlying structures as well as ensuring that perturbations specifically disrupt contextual processing. Perhaps the most promising approach involves closed-loop experiments combining recording and manipulation, potentially harnessing cutting-edge tools like Neuropixels Opto probes (Lakunina et al., 2025). Not only do these probes allow for dense electrophysiological recording and restricted deliverance of light from specific sites along the probe, they have the additional advantage of minimally damaging tissue with just a 70 micron shank width. One could identify MRF neurons based on their functional properties (e.g., baseline context coding) during the task and then deliver spatially precise optogenetic stimulation targeted at these specific neuronal populations during critical task epochs on interleaved trials. An even more powerful approach would involve manifold-specific perturbations, wherein the population activity in state space is defined online and the pattern of perturbation is defined to manipulate the activity “on-manifold” (Jazayeri and Afraz, 2017; Sadtler et al., 2014). For example, if perturbing the activity of the population defining the baseline context dimension we observed in **Chapter 3**, one could perform an on-manifold manipulation to specifically enhance representation of one context over the other while leaving activity in other premotor and

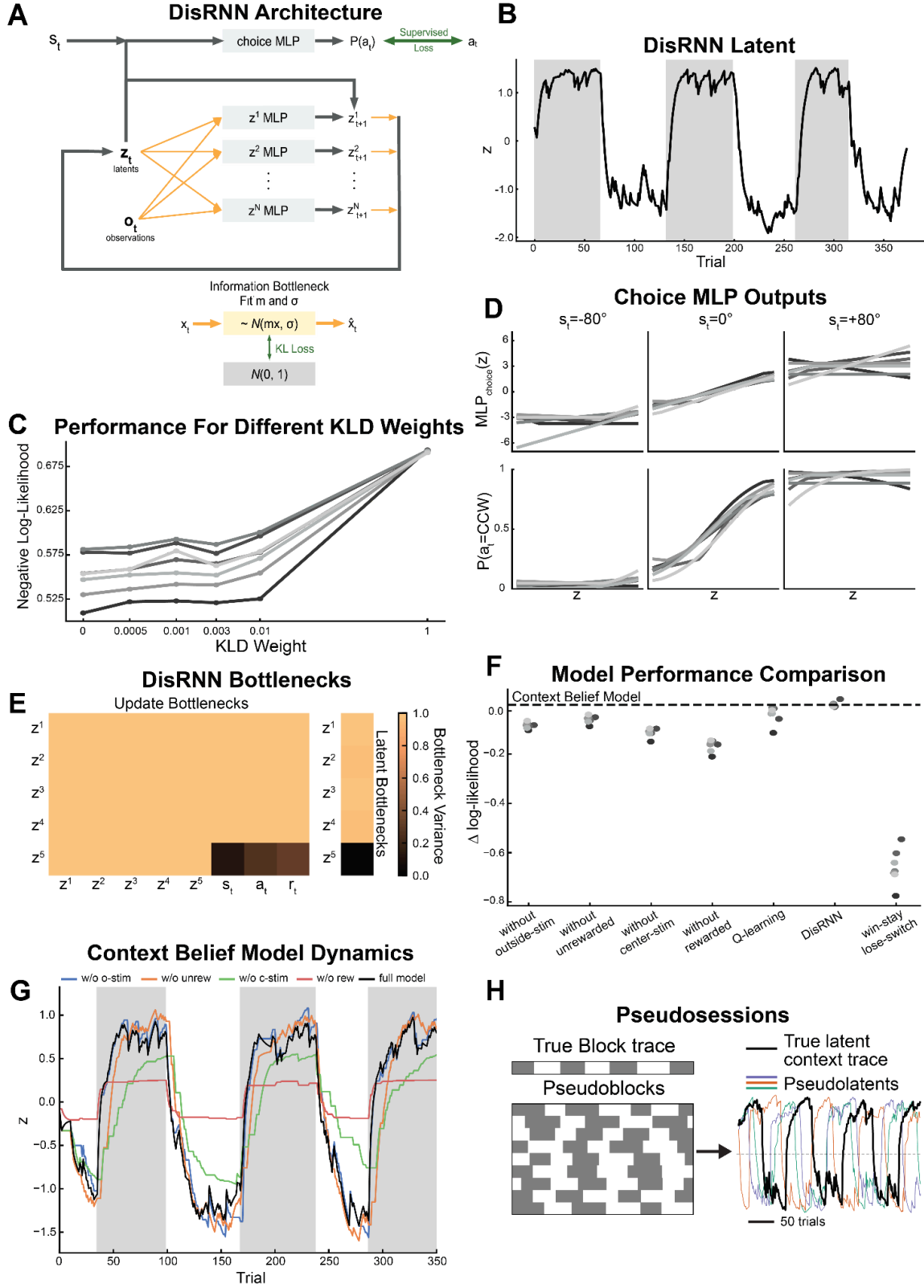
stimulus-related dimensions relatively unperturbed. Observing whether selectively perturbing activity along the context dimension in MRF specifically disrupts context-dependent choices (i.e., on center stimulus trials) while potentially sparing responses on unambiguous trials (i.e. outside stimulus trials) or basic motor execution would provide strong causal evidence for the hypothesis that this population of MRF neurons plays a key role in dynamic context-setting, circumventing many of the confounds associated with less targeted methods.

Understanding the MRF's role in contextual processing and flexible behavior has many potential broader implications. Dysfunction in circuits involving the MRF, given its extensive connections to cortical, basal ganglia, and thalamic regions, could contribute to deficits seen in various neurological and psychiatric disorders. For example, impaired context processing is a feature of schizophrenia (Floresco et al., 2009; Yang et al., 2013), while motor control deficits involving related brainstem circuits are central to Parkinson's disease. The use of Deep Brain Stimulation (DBS) targeting the adjacent PPN area for Parkinsonian gait issues (Thevathasan et al., 2018) highlights the therapeutic potential of modulating this region, but also the risks; variable outcomes and side effects related to cognition or alertness might stem from non-specific stimulation affecting diverse populations within the complex MRF territory. A detailed understanding of MRF subregions and functions could inform more precise therapeutic targeting.

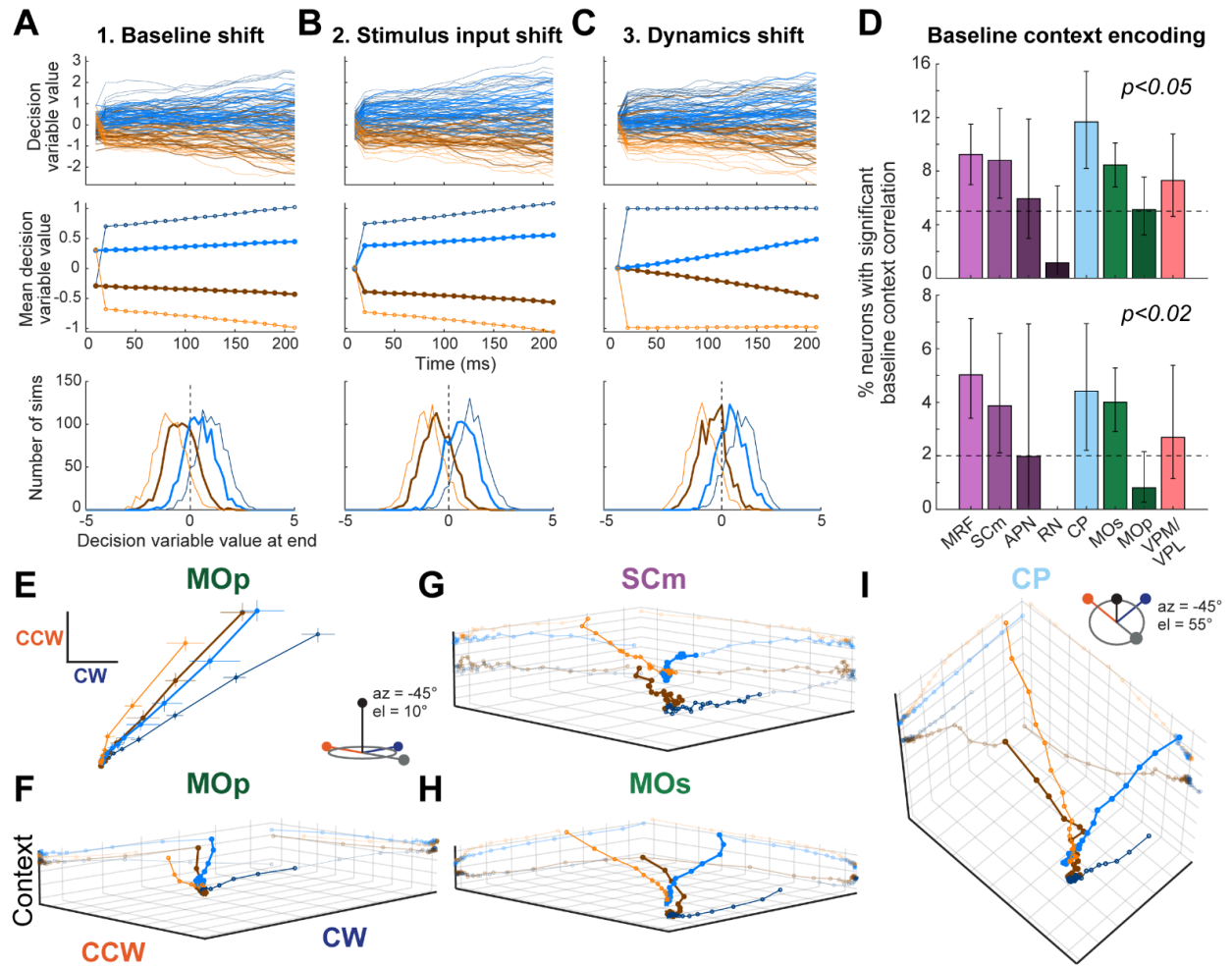
In conclusion, the work presented here challenges the traditional motor-related view of MRF by providing evidence for its role as a node in the network underlying abstract contextual processing during flexible visual decisions. This new perspective

provides the impetus and groundwork for future MRF studies to clarify its specific roles in the brain-wide circuits producing context-dependent behavior.

Supplementary Figures

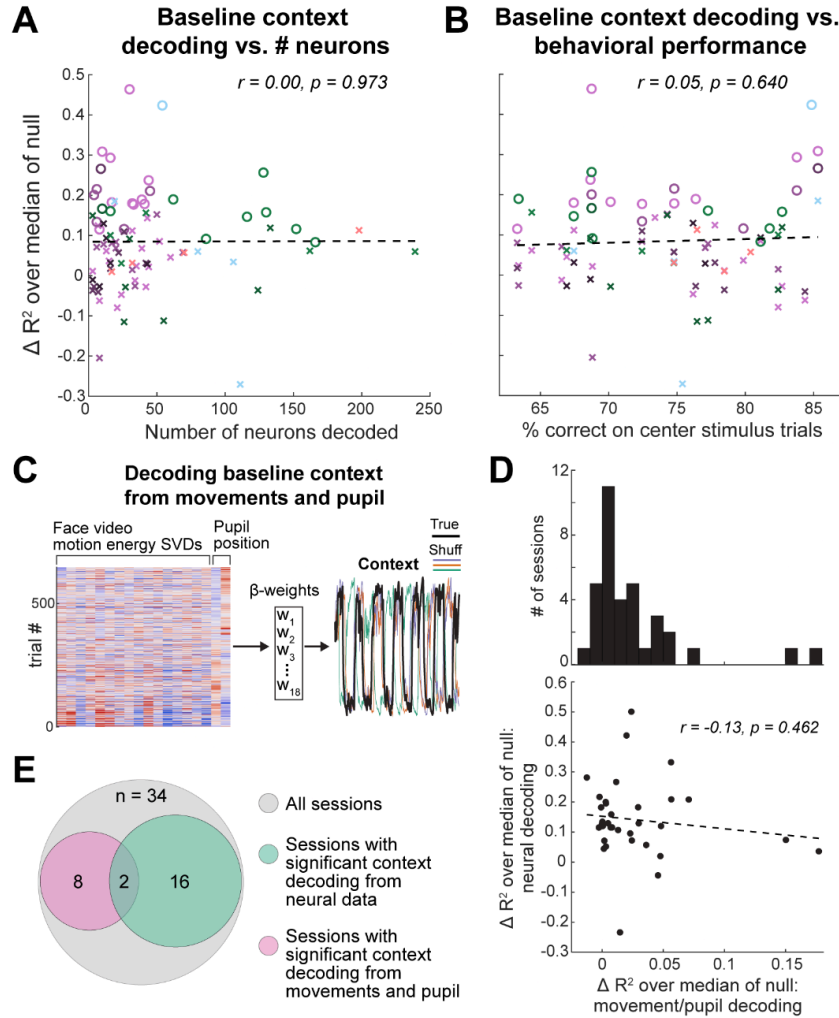


Supplementary Figure 1. Disentangled RNNs discover a model class which captures task context, related to Figure 2.1. All panels were created by J. Schroter, with the exception of panel H. **A)** Disentangled RNN model schematic, see Chapter 2.2.4 for more details. **B)** Representative evolution from an example session of the latent variable discovered by Disentangled RNN overlaid on task Blocks (shaded = left Block, unshaded = right Block). **C)** center stimulus trial performance of Disentangled RNNs for different KL-Divergence weightings. Performance across weightings are plotted for each of the 7 subjects. **D)** Choice MLP output and probabilities for all subjects over a sweep of latent variable values. **E)** Representative heatmap from an example subject for disentangled RNN bottlenecks. Large bottleneck variance value corresponds to a closed bottleneck, meaning those variables are not used by the model. The pattern of a single latent variable solely dependent on the observations was consistently found across subjects and corresponds to a latent update equation that is linear, conditioned on the observation. This leads to piecewise linear dynamics driven by discrete inputs. **F)** Performance of full context-belief model vs. context-belief models with dynamics fixed on outside stimulus, all unrewarded, center stimulus, or all rewarded trials. Q-Learning, DisRNN, and win-stay-lose-switch performances are also shown. For win-stay-lose-switch probabilities of 0.99 and 0.01 were chosen for model predictions in order to compare to other models. **G)** Full context-belief latent trace along with latent traces with dynamics fixed on different subsets of trials. **H)** Schematic illustrating pseudosession approach: “Pseudoblock” traces are created using the same generating distribution as the true Block. These are fed into each mouse’s fitted context belief model, and the choice on each trial is predicted from the model’s probabilities based on the previous trial type and latent value. Predicted choices are fed back into the model to sample the subsequent trial.



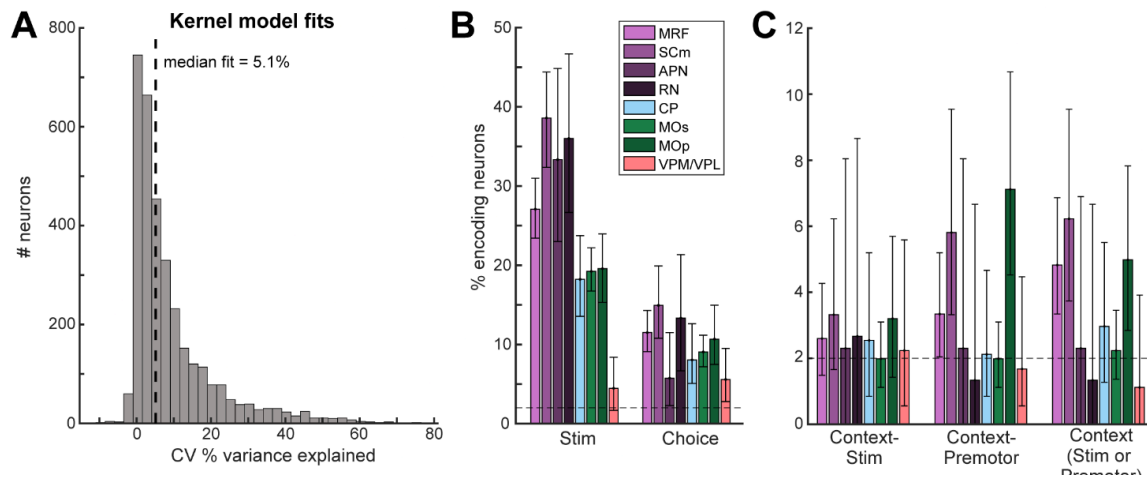
Supplementary Figure 2. Drift-diffusion modeling and population trajectories, related to Figure 3.2.

A) Instantiation of hypothesized context-dependent baseline motor bias mechanism in drift-diffusion model. Left Block trials start at -0.3 along decision variable axis (positive values = CW, negative = CCW), right Block trials start at 0.3. Top: Trajectory of decision variable on 30 individual trials per trial type. Middle: Averaged trajectories for each trial type. Bottom: Histograms of decision variable values at final time point for each trial type across 1000 simulations. **B)** Same as A) for the stimulus input shift mechanism. Left Block center stimulus trials receive negative (i.e. toward CCW) input, right Block center stimuli receive positive input. **C)** Same as A) and B) for shifting attractor dynamics mechanism. At each time point, left Block trials receive small negative inputs and right Block trials receive small positive inputs. **D)** Percentage of neurons encoding baseline (pre-stimulus) context by region. Baseline firing rates per trial were correlated with latent context traces from the context belief behavioral model. Significance was determined against a pseudosession null. Error bars are 95% confidence intervals. Top plot uses $\alpha = 0.05$ threshold to determine encoding, bottom $\alpha = 0.02$. **E)** Population trajectory by trial type in MOp in 2D CW vs. CCW state space. Each point is a timepoint within a correct trial (median across all trials standardized and combined from all sessions; error bars are 95% confidence interval (CI)). Trial starts at the bottom left. Trajectories span from -250 to 0 ms aligned to movement onset in a sliding window of 75 ms bins with 10 ms step sizes. **F-I)** 3D (CW, CCW, context) state space trajectories of MOp, SCm, MOs, and CP.

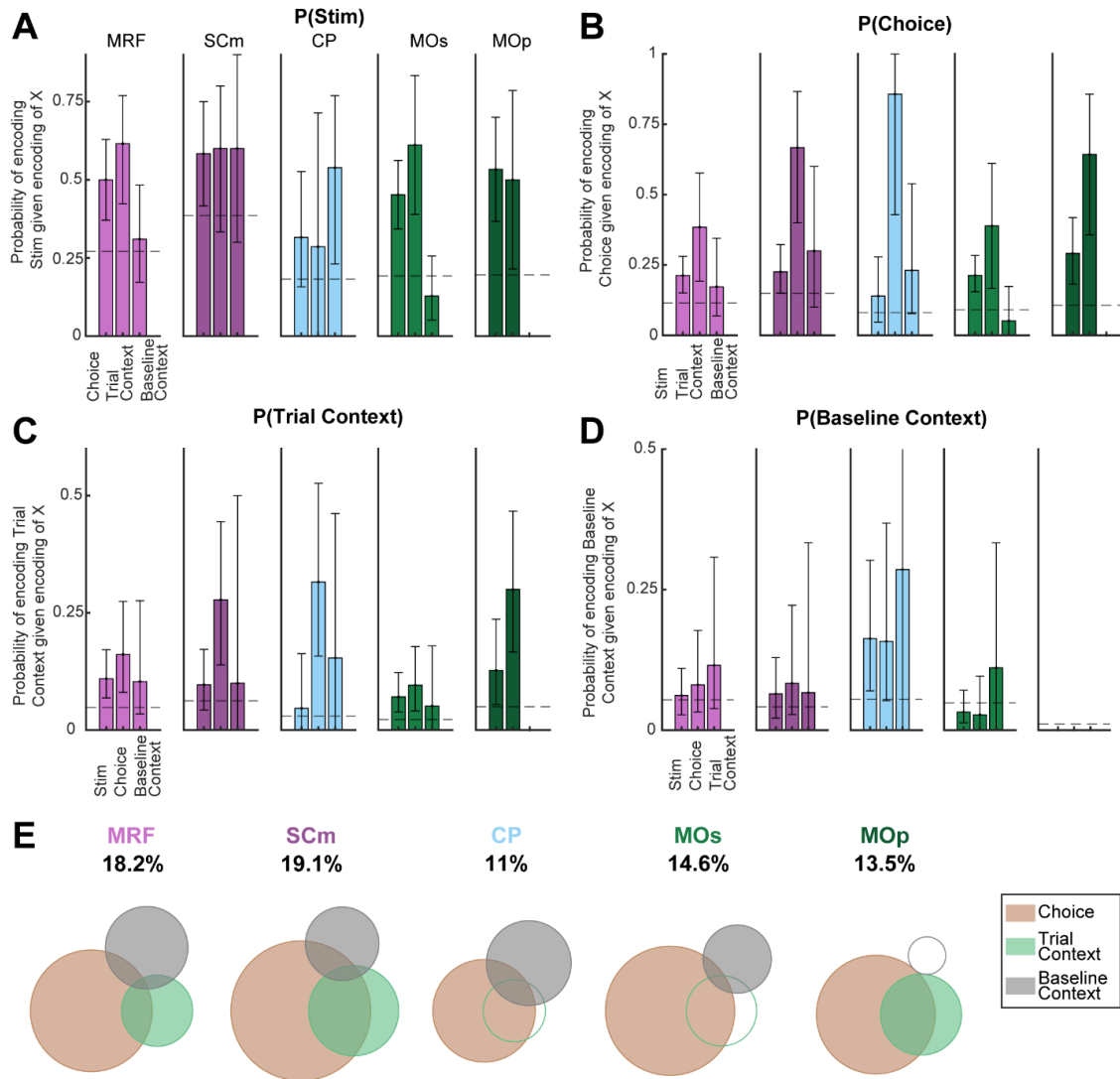


Supplementary Figure 3. Baseline context decoding is not related to number of neurons in the population, behavioral performance, or uninstructed movements, related to Figure 3.2. **A)** Scores of decoding context from baseline pre-stimulus firing rates of neural populations across regions (ΔR^2 over median of pseudosession null) as a function of number of neurons in the decoded population. Each symbol represents a neural population from a single session, with colors representing brain region as in Figures 2-4. Open circles indicate sessions with significant decoding ($p < 0.05$), x's indicate $p > 0.05$. Fitted regression line shows no relationship between number of neurons decoded and baseline context decoding score ($r = 0, p = 0.97$). **B)** Same as A) but plotting baseline context decoding score as a function of session behavioral performance metric (% correct on center stimulus trials). Fitted regression line shows no significant relationship ($r = 0.05, p = 0.64$). **C)** Schematic showing decoding of context from top 16 PCs of face video motion energy SVDs and x and y positions of center of pupil (derived from tracking with Deep Lab Cut (Mathis et al., 2018)). Values of face SVDs and pupil positions across the session were standardized with robust Z-scoring, then mean values were taken 300 to 50 ms before stimulus onset on each trial. True context was decoded along with context traces from 1000 pseudosessions to create a null distribution for each session. **D)** Top: distribution of baseline context decoding scores (ΔR^2 over median of pseudosession null). Bottom: Baseline context decoding score from neural activity as a function of baseline context decoding score from uninstructed movements and pupil position. Each point is a session; when multiple populations were decoded in a session, the population with the highest decoding score was plotted. Fitted regression line shows no significant relationship ($r = -0.13, p = 0.462$). **E)** Overlap

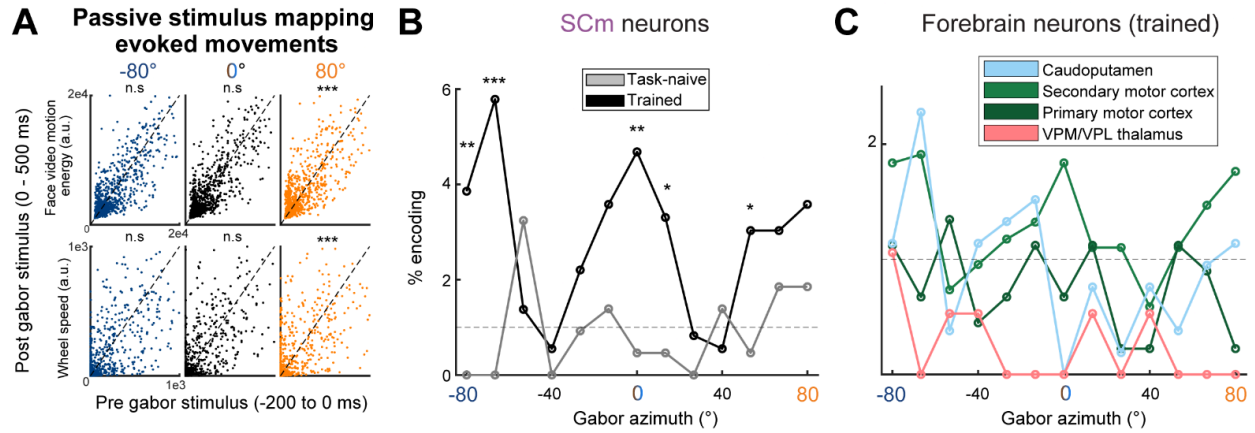
of sessions with significant baseline context decoding from neural activity and sessions with significant baseline context decoding from uninstructed movements and pupil dynamics. 2 sessions are overlapping, and an overlap of 3.8 sessions would be expected if the variables were independent.



Supplementary Figure 4. Kernel encoding model fits and coding of task variables by region, related to Figure 3.3. **A)** Distribution of all individual neuron fits from the kernel encoding model as measured by cross-validated variance explained. **B)** Percentage of neurons encoding stimulus or choice variables by region as in Fig. 4F. Encoding was determined by fit of kernel encoding model against null with specific variable shuffled ($p < 0.02$, Chapters 1.4.7, 3.2.6, & 3.2.8). Dotted line indicates false discovery rate, i.e. chance level. Error bars 95% CIs; bars with lower CI above chance level is deemed as a significant encoding of the variable by the region. **C)** Same as B) but for contextual modulation variables. Shuffles were created by using pseudosession context traces to construct context-stim kernel, context-premotor kernel, or context-stim and context-premotor kernels together.



Supplementary Figure 5. Increased probability of joint coding of trial context and choice, related to Figure 3.3. **A)** Probability of encoding stimulus given encoding of choice, trial context, or baseline context variables in MRF, SCm, CP, MOs, and MOp; i.e., percentage of choice, trial context, or baseline context coding neurons which also encode stimulus. Dotted line is the overall proportion of neurons in a given region encoding stimulus, representing the baseline probability of stimulus encoding. Encoding of stimulus, choice, and trial context were determined by kernel model (Fig. 4, Supp Fig. KERNEL); encoding baseline context determined by correlating baseline firing rates with context traces (Fig. 3E, Supp Fig. TRAJECTORIES_D); $p < 0.02$. Error bars are 95% CIs. **B)-D)** Same as A), but for probability of encoding choice, trial context, or baseline context given encoding of other 3 variables. **E)** Venn diagrams displaying overlap of populations encoding choice, trial context, and/or baseline context across regions. Filled circles indicate above chance coding, empty circles indicate below chance coding (Supp Fig. KERNEL_B,C, Supp Fig. TRAJECTORIES_D). Value below the region labels is the percentage of neurons in a region represented by the total area encompassed by the three circles.



Supplementary Figure 6. Stimulus responsiveness and evoked movements during passive stimulus mapping, related to Figure 4.1. **A)** Mean trial-by-trial face motion energy or wheel speed for task stimuli (split into -80°, 0°, 80°) in the pre-gabor stimulus period vs. period after gabor stimulus onset. Each point is a trial, compiled across all passive task sessions for trained mice. P values are derived from Wilcoxon signed rank test; only the right stimulus evoked significant motion and wheel turn responses to gabor stimuli. **B)** Percentage of SCm neurons encoding gabor stimuli split by azimuth in task-naïve or trained mice. P values are derived from chi-squared test between trained and task-naïve mice at matched azimuths. **C)** Percentage of CP, MOs, MOp, and VPM/VPL neurons encoding gabor stimuli split by azimuth in trained mice (regions were not targeted in task-naïve recordings). Coding at all azimuths for each region was not above chance (lower 95% CI overlapping with chance level; not shown).

References

- Aoi, M.C., Mante, V., Pillow, J.W., 2020. Prefrontal cortex exhibits multidimensional dynamic encoding during decision-making. *Nat. Neurosci.* 23, 1410–1420. <https://doi.org/10.1038/s41593-020-0696-5>
- Arber, S., Costa, R.M., 2022a. Networking brainstem and basal ganglia circuits for movement. *Nat. Rev. Neurosci.* 23, 342–360. <https://doi.org/10.1038/s41583-022-00581-w>
- Arber, S., Costa, R.M., 2022b. Networking brainstem and basal ganglia circuits for movement. *Nat. Rev. Neurosci.* 23, 342–360. <https://doi.org/10.1038/s41583-022-00581-w>
- Arlt, C., Barroso-Luque, R., Kira, S., Bruno, C.A., Xia, N., Chettih, S.N., Soares, S., Pettit, N.L., Harvey, C.D., 2022. Cognitive experience alters cortical involvement in goal-directed navigation. *eLife* 11, e76051. <https://doi.org/10.7554/eLife.76051>
- Asaad, W.F., Rainer, G., Miller, E.K., 2000. Task-Specific Neural Activity in the Primate Prefrontal Cortex. *J. Neurophysiol.* 84, 451–459. <https://doi.org/10.1152/jn.2000.84.1.451>
- Balleine, B.W., Delgado, M.R., Hikosaka, O., 2007. The Role of the Dorsal Striatum in Reward and Decision-Making. *J. Neurosci.* 27, 8161–8165. <https://doi.org/10.1523/JNEUROSCI.1554-07.2007>
- Bari, B.A., Cohen, J.Y., 2021. Dynamic decision making and value computations in medial frontal cortex. *Int. Rev. Neurobiol.* 158, 83–113. <https://doi.org/10.1016/bs.irn.2020.12.001>
- Bari, B.A., Grossman, C.D., Lubin, E.E., Rajagopalan, A.E., Cressy, J.I., Cohen, J.Y., 2019. Stable Representations of Decision Variables for Flexible Behavior. *Neuron* 103, 922–933.e7. <https://doi.org/10.1016/j.neuron.2019.06.001>
- Basso, M.A., May, P.J., 2017. Circuits for Action and Cognition: A View from the Superior Colliculus. *Annu. Rev. Vis. Sci.* 3, 197–226. <https://doi.org/10.1146/annurev-vision-102016-061234>
- Basso, M.A., Wurtz, R.H., 1998. Modulation of neuronal activity in superior colliculus by changes in target probability. *J. Neurosci. Off. J. Soc. Neurosci.* 18, 7519–7534. <https://doi.org/10.1523/JNEUROSCI.18-18-07519.1998>
- Basso, M.A.M.A., Bickford, M.E.M.E., Cang, J., 2021. Unraveling circuits of visual perception and cognition through the superior colliculus. *Neuron* 109, 918–937. <https://doi.org/10.1016/J.NEURON.2021.01.013>
- Bell, C., Sierra, G., Buendia, N., Segundo, J.P., 1964. SENSORY PROPERTIES OF NEURONS IN THE MESENCEPHALIC RETICULAR FORMATION. *J. Neurophysiol.* 27, 961–987. <https://doi.org/10.1152/jn.1964.27.6.961>
- Benavidez, N.L., Bienkowski, M.S., Zhu, M., Garcia, L.H., Fayzullina, M., Gao, L., Bowman, I., Gou, L., Khanjani, N., Cotter, K.R., Korobkova, L., Becerra, M., Cao, C., Song, M.Y., Zhang, B., Yamashita, S., Tugangui, A.J., Zingg, B., Rose, K., Lo, D., Foster, N.N., Boesen, T., Mun, H.-S.S., Aquino, S., Wickersham, I.R., Ascoli, G.A., Hintiryan, H., Dong, H.-W.W., 2021. Organization of the inputs and outputs of the mouse superior colliculus. *Nat. Commun.* 12, 1–20. <https://doi.org/10.1038/s41467-021-24241-2>
- Birman, D., Yang, K.J., West, S.J., Karsh, B., Browning, Y., Laboratory, the I.B., Siegle, J.H., Steinmetz, N.A., 2023. Pinpoint: trajectory planning for multi-probe electrophysiology and injections in an interactive web-based 3D environment. *eLife* 12. <https://doi.org/10.7554/eLife.91662.1>
- Biró, S., Lasztóczy, B., Klausberger, T., 2019. A Visual Two-Choice Rule-Switch Task for Head-Fixed Mice. *Front. Behav. Neurosci.* 13. <https://doi.org/10.3389/fnbeh.2019.00119>
- Bohlen, M.O., Warren, S., May, P.J., 2022. Is the central mesencephalic reticular formation a purely horizontal gaze center? *Brain Struct. Funct.* 227, 2367–2393. <https://doi.org/10.1007/s00429-022-02532-8>

- Bradfield, L.A., Hart, G., 2020. Rodent medial and lateral orbitofrontal cortices represent unique components of cognitive maps of task space. *Neurosci. Biobehav. Rev.* 108, 287–294. <https://doi.org/10.1016/J.NEUBIOREV.2019.11.009>
- Burgess, C.P., Lak, A., Steinmetz, N.A., Zatzka-Haas, P., Bai Reddy, C., Jacobs, E.A.K., Linden, J.F., Paton, J.J., Ranson, A., Schröder, S., Soares, S., Wells, M.J., Wool, L.E., Harris, K.D., Carandini, M., 2017. High-Yield Methods for Accurate Two-Alternative Visual Psychophysics in Head-Fixed Mice. *Cell Rep.* 20, 2513. <https://doi.org/10.1016/J.CELREP.2017.08.047>
- Büttner-Ennever, J.A., Büttner, U., 1978. A cell group associated with vertical eye movements in the rostral mesencephalic reticular formation of the monkey. *Brain Res.* 151, 31–47.
- Cáceda, R., Nemeroff, C.B., Harvey, P.D., 2014. Toward an understanding of decision making in severe mental illness. *J Neuropsychiatry Clin Neurosci* 26, 196–213. <https://doi.org/10.1176/appi.neuropsych.12110268>
- Carandini, M., Churchland, A.K., 2013. Probing perceptual decisions in rodents. *Nat. Neurosci.* 16, 824. <https://doi.org/10.1038/NN.3410>
- Cella, M., Dymond, S., Cooper, A., 2010. Impaired flexible decision-making in Major Depressive Disorder. *J. Affect. Disord.* 124, 207–210. <https://doi.org/10.1016/j.jad.2009.11.013>
- Chen, B., May, P.J., 2000. The feedback circuit connecting the superior colliculus and central mesencephalic reticular formation: a direct morphological demonstration. *Exp. Brain Res.* 131, 10–21. <https://doi.org/10.1007/s002219900280>
- Chen, G., King, J.A., Burgess, N., O’Keefe, J., 2013. How vision and movement combine in the hippocampal place code. *Proc Natl Acad Sci U A* 110, 378–383.
- Chen, S., Liu, Y., Wang, Z.A., Colonell, J., Liu, L.D., Hou, H., Tien, N.-W., Wang, T., Harris, T., Druckmann, S., Li, N., Svoboda, K., 2024. Brain-wide neural activity underlying memory-guided movement. *Cell* 187, 676–691.e16. <https://doi.org/10.1016/j.cell.2023.12.035>
- Chen, T.-W., Li, N., Daie, K., Svoboda, K., 2017. A Map of Anticipatory Activity in Mouse Motor Cortex. *Neuron* 94, 866–879.e4. <https://doi.org/10.1016/j.neuron.2017.05.005>
- Cisek, P., 2021a. Evolution of behavioural control from chordates to primates. *Philos. Trans. R. Soc. B Biol. Sci.* 377, 20200522. <https://doi.org/10.1098/rstb.2020.0522>
- Cisek, P., 2021b. An Evolutionary Perspective on Embodiment, in: Robinson, M.D., Thomas, L.E. (Eds.), *Handbook of Embodied Psychology: Thinking, Feeling, and Acting*. Springer International Publishing, Cham, pp. 547–571. https://doi.org/10.1007/978-3-030-78471-3_24
- Cohen, B., Matsuo, V., Fradin, J., Raphan, T., 1985. Horizontal saccades induced by stimulation of the central mesencephalic reticular formation. *Exp. Brain Res.* 57, 605–616. <https://doi.org/10.1007/BF00237847>
- Cole, N., Harvey, M., Myers-Joseph, D., Gilra, A., Khan, A.G., 2024. Prediction-error signals in anterior cingulate cortex drive task-switching. *Nat. Commun.* 15, 7088. <https://doi.org/10.1038/s41467-024-51368-9>
- Crick, F.C., 1984. Function of the thalamic reticular complex: The searchlight hypothesis. *Proc. Natl. Acad. Sci.* 81, 4586–4590.
- Cromer, J.A., Waitzman, D.M., 2006. Neurones associated with saccade metrics in the monkey central mesencephalic reticular formation. *J Physiol* 570, 507–523. <https://doi.org/10.1113/jphysiol.2005.096834>
- Cruz, K.G., Leow, Y.N., Le, N.M., Adam, E., Huda, R., Sur, M., 2022. Cortical-Subcortical Interactions in Goal-directed Behavior.
- Daan, S., Spoelstra, K., Albrecht, U., Schmutz, I., Daan, M., Daan, B., Rienks, F., Poletaeva, I., Dell’Omo, G., Vyssotski, A., Lipp, H.-P., 2011. Lab mice in the field: unorthodox daily activity and effects of a dysfunctional circadian clock allele. *J. Biol. Rhythms* 26, 118–129. <https://doi.org/10.1177/0748730410397645>

- Dahlstrom, A., 1964. Evidence for the existence of monoamine-containing neurons in the central nervous system. I. Demonstration of monoamines in the cell bodies of brain stem neurons. *Acta Physiol Scand* 62, 1–55.
- Daw, N.D., Niv, Y., Dayan, P., 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711. <https://doi.org/10.1038/nn1560>
- D’Cruz, A.-M., Ragozzino, M.E., Mosconi, M.W., Shrestha, S., Cook, E.H., Sweeney, J.A., 2013. Reduced behavioral flexibility in autism spectrum disorders. *Neuropsychology* 27, 152–160. <https://doi.org/10.1037/a0031721>
- Devineni, A.V., Scaplen, K.M., 2022. Neural Circuits Underlying Behavioral Flexibility: Insights From *Drosophila*. *Front. Behav. Neurosci.* 15. <https://doi.org/10.3389/fnbeh.2021.821680>
- Ding, L., Gold, J.I., 2013. The basal ganglia’s contributions to perceptual decision making. *Neuron* 79, 640–649. <https://doi.org/10.1016/j.neuron.2013.07.042>
- Ding, L., Gold, J.I., 2010. Caudate Encodes Multiple Computations for Perceptual Decisions. *J. Neurosci.* 30, 15747–15759. <https://doi.org/10.1523/JNEUROSCI.2894-10.2010>
- Driscoll, L.N., Pettit, N.L., Minderer, M., Chettih, S.N., Harvey, C.D., 2017. Dynamic Reorganization of Neuronal Activity Patterns in Parietal Cortex. *Cell* 170, 986–999.e16. <https://doi.org/10.1016/j.cell.2017.07.021>
- Duan, C.A., Erlich, J.C., Brody, C.D., 2015. Requirement of Prefrontal and Midbrain Regions for Rapid Executive Control of Behavior in the Rat. *Neuron* 86, 1491–1503. <https://doi.org/10.1016/j.neuron.2015.05.042>
- Duan, C.A., Pagan, M., Piet, A.T., Kopec, C.D., Akrami, A., Riordan, A.J., Erlich, J.C., Brody, C.D., 2021. Collicular circuits for flexible sensorimotor routing. *Nat. Neurosci.* <https://doi.org/10.1038/s41593-021-00865-x>
- Duménieu, M., Marquèze-Pouey, B., Russier, M., Debanne, D., 2021. Mechanisms of Plasticity in Subcortical Visual Areas. *Cells* 2021 Vol 10 Page 3162 10, 3162. <https://doi.org/10.3390/CELLS10113162>
- Durstewitz, D., Vittoz, N.M., Floresco, S.B., Seamans, J.K., 2010. Abrupt Transitions between Prefrontal Neural Ensemble States Accompany Behavioral Transitions during Rule Learning. *Neuron* 66, 438–448. <https://doi.org/10.1016/j.neuron.2010.03.029>
- Ebitz, R.B., Hayden, B.Y., 2021. The population doctrine in cognitive neuroscience. *Neuron*. <https://doi.org/10.1016/J.NEURON.2021.07.011>
- Elsayed, G.F., Lara, A.H., Kaufman, M.T., Churchland, M.M., Cunningham, J.P., 2016. Reorganization between preparatory and movement population responses in motor cortex. *Nat. Commun.* 7, 1–15. <https://doi.org/10.1038/ncomms13239>
- Ertürk, A., Becker, K., Jährling, N., Mauch, C.P., Hojer, C.D., Egen, J.G., Hellal, F., Bradke, F., Sheng, M., Dodt, H.-U., 2012. Three-dimensional imaging of solvent-cleared organs using 3DISCO. *Nat. Protoc.* 7, 1983–1995. <https://doi.org/10.1038/nprot.2012.119>
- Essig, J., Hunt, J.B., Felsen, G., 2021. Inhibitory neurons in the superior colliculus mediate selection of spatially-directed movements. *Commun. Biol.* 4, 1–14. <https://doi.org/10.1038/s42003-021-02248-1>
- Evenden, J.L., Robbins, T.W., 1984. Win-Stay Behaviour in the Rat. *Q. J. Exp. Psychol. Sect. B* 36, 1–26. <https://doi.org/10.1080/14640748408402190>
- Felsen, G., Mainen, Z.F., 2012. Midbrain contributions to sensorimotor decision making. *J. Neurophysiol.* 108, 135. <https://doi.org/10.1152/JN.01181.2011>
- Felsen, G., Mainen, Z.F., 2008. Neural substrates of sensory-guided locomotor decisions in the rat superior colliculus. *Neuron* 60, 137–148. <https://doi.org/10.1016/j.neuron.2008.09.019>
- Ferreira-Pinto, M.J., Kanodia, H., Falasconi, A., Sigrist, M., Esposito, M.S., Arber, S., 2021. Functional diversity for body actions in the mesencephalic locomotor region. *Cell* 184, 4564–4578.e18. <https://doi.org/10.1016/j.cell.2021.07.002>

- Findling, C., Hubert, F., Laboratory, I.B., Acerbi, L., Benson, B., Benson, J., Birman, D., Bonacchi, N., Carandini, M., Catarino, J.A., Chapuis, G.A., Churchland, A.K., Dan, Y., DeWitt, E.E., Engel, T.A., Fabbri, M., Faulkner, M., Fiete, I.R., Freitas-Silva, L., Gerçek, B., Harris, K.D., Häusser, M., Hofer, S.B., Hu, F., Huntenburg, J.M., Khanal, A., Krasniak, C., Langdon, C., Latham, P.E., Lau, P.Y.P., Mainen, Z., Meijer, G.T., Miska, N.J., Mrsic-Flogel, T.D., Noel, J.-P., Nylund, K., Pan-Vazquez, A., Paninski, L., Pillow, J., Rossant, C., Roth, N., Schaeffer, R., Schartner, M., Shi, Y., Socha, K.Z., Steinmetz, N.A., Svoboda, K., Tessereau, C., Urai, A.E., Wells, M.J., West, S.J., Whiteway, M.R., Winter, O., Witten, I.B., Zador, A., Dayan, P., Pouget, A., 2023. Brain-wide representations of prior information in mouse decision-making. <https://doi.org/10.1101/2023.07.04.547684>
- Floresco, S.B., Zhang, Y., Enomoto, T., 2009. Neural circuits subserving behavioral flexibility and their relevance to schizophrenia. *Behav. Brain Res.* 204, 396–409. <https://doi.org/10.1016/j.bbr.2008.12.001>
- Flossmann, T., Rochefort, N.L., 2021. Spatial navigation signals in rodent visual cortex. *Curr. Opin. Neurobiol.* 67, 163–173. <https://doi.org/10.1016/j.conb.2020.11.004>
- Foster, N.N., Barry, J., Korobkova, L., Garcia, L., Gao, L., Becerra, M., Sherfat, Y., Peng, B., Li, X., Choi, J.-H., Gou, L., Zingg, B., Azam, S., Lo, D., Khanjani, N., Zhang, B., Stanis, J., Bowman, I., Cotter, K., Cao, C., Yamashita, S., Tugangui, A., Li, A., Jiang, T., Jia, X., Feng, Z., Aquino, S., Mun, H.-S., Zhu, M., Santarelli, A., Benavidez, N.L., Song, M., Dan, G., Fayzullina, M., Ustrell, S., Boesen, T., Johnson, D.L., Xu, H., Bienkowski, M.S., Yang, X.W., Gong, H., Levine, M.S., Wickersham, I., Luo, Q., Hahn, J.D., Lim, B.K., Zhang, L.I., Cepeda, C., Hintiryan, H., Dong, H.-W., 2021a. The mouse cortico–basal ganglia–thalamic network. *Nature* 598, 188–194. <https://doi.org/10.1038/s41586-021-03993-3>
- Foster, N.N., Barry, J., Korobkova, L., Garcia, L., Gao, L., Becerra, M., Sherfat, Y., Peng, B., Li, X., Choi, J.-H., Gou, L., Zingg, B., Azam, S., Lo, D., Khanjani, N., Zhang, B., Stanis, J., Bowman, I., Cotter, K., Cao, C., Yamashita, S., Tugangui, A., Li, A., Jiang, T., Jia, X., Feng, Z., Aquino, S., Mun, H.-S., Zhu, M., Santarelli, A., Benavidez, N.L., Song, M., Dan, G., Fayzullina, M., Ustrell, S., Boesen, T., Johnson, D.L., Xu, H., Bienkowski, M.S., Yang, X.W., Gong, H., Levine, M.S., Wickersham, I., Luo, Q., Hahn, J.D., Lim, B.K., Zhang, L.I., Cepeda, C., Hintiryan, H., Dong, H.-W., 2021b. The mouse cortico–basal ganglia–thalamic network. *Nature* 598, 188–194. <https://doi.org/10.1038/s41586-021-03993-3>
- Fraser, K.M., Holland, P.C., 2019. Occasion setting. *Behav. Neurosci.* 133, 145–175. <https://doi.org/10.1037/bne0000306>
- Fujita, H., Kodama, T., Du Lac, S., 2020. Modular output circuits of the fastigial nucleus for diverse motor and nonmotor functions of the cerebellar vermis. *Elife* 9, e58613.
- Fuster, J.M., Uyeda, A.A., 1962. Facilitation of tachistoscopic performance by stimulation of midbrain tegmental points in the monkey. *Exp. Neurol.* 6, 384–406.
- Gao, L., Liu, S., Gou, L., Hu, Y., Liu, Y., Deng, L., Ma, D., Wang, H., Yang, Q., Chen, Z., Liu, D., Qiu, S., Wang, Xiaofei, Wang, D., Wang, Xinran, Ren, B., Liu, Q., Chen, T., Shi, X., Yao, H., Xu, C., Li, C.T., Sun, Y., Li, A., Luo, Q., Gong, H., Xu, N., Yan, J., 2022. Single-neuron projectome of mouse prefrontal cortex. *Nat. Neurosci.* 25, 515–529. <https://doi.org/10.1038/s41593-022-01041-5>
- Ghosh, S., Zador, A.M., 2021. Corticostriatal Plasticity Established by Initial Learning Persists after Behavioral Reversal. *eNeuro* 8. <https://doi.org/10.1523/ENEURO.0209-20.2021>
- Gire, D.H., Kapoor, V., Arrighi-Allisan, A., Seminara, A., Murthy, V.N., 2016. Mice develop efficient strategies for foraging and navigation using complex natural stimuli. *Curr. Biol.* CB 26, 1261–1273. <https://doi.org/10.1016/j.cub.2016.03.040>
- Glaser, J., Whiteway, M., Cunningham, J.P., Paninski, L., Linderman, S., 2020. Recurrent Switching Dynamical Systems Models for Multiple Interacting Neural Populations, in:

- Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 14867–14878.
- Gleichgerricht, E., Ibáñez, A., Roca, M., Torralva, T., Manes, F., 2010. Decision-making cognition in neurodegenerative diseases. *Nat. Rev. Neurol.* 6, 611–623. <https://doi.org/10.1038/nrneurol.2010.148>
- Glickfeld, L.L., Reid, R.C., Andermann, M.L., 2014. A mouse model of higher visual cortical function. *Curr. Opin. Neurobiol.* 24, 28–33. <https://doi.org/10.1016/J.CONB.2013.08.009>
- Goard, M.J., Pho, G.N., Woodson, J., Sur, M., 2016. Distinct roles of visual, parietal, and frontal motor cortices in memory-guided sensorimotor decisions. *eLife* 5, e13764. <https://doi.org/10.7554/eLife.13764>
- Gold, J.I., Shadlen, M.N., 2007. The Neural Basis of Decision Making. *Annu. Rev. Neurosci.* 30, 535–574. <https://doi.org/10.1146/annurev.neuro.29.051605.113038>
- Gold, J.I., Shadlen, M.N., 2000. Representation of a perceptual decision in developing oculomotor commands. *Nature* 404, 390–394. <https://doi.org/10.1038/35006062>
- Gold, J.I., Stocker, A.A., 2017. Visual Decision-Making in an Uncertain and Dynamic World. *Annu Rev Vis Sci* 3, 227–250. <https://doi.org/10.1146/annurev-vision-111815-114511>
- Graf, W.M., Ugolini, G., 2006. The central mesencephalic reticular formation: Its role in space-time coordinated saccadic eye movements. *J. Physiol.* 570, 433–434. <https://doi.org/10.1113/jphysiol.2005.103184>
- Grantyn, A., Berthoz, A., 1987. Reticulo-spinal neurons participating in the control of synergic eye and head movements during orienting in the cat. I. Behavioral properties. *Exp. Brain Res.* 66, 339–354. <https://doi.org/10.1007/BF00243309>
- Gruner, P., Pittenger, C., 2017. Cognitive inflexibility in Obsessive-Compulsive Disorder. *Neuroscience* 345, 243–255. <https://doi.org/10.1016/j.neuroscience.2016.07.030>
- Guo, Z.V., Hires, S.A., Li, N., O'Connor, D.H., Komiyama, T., Ophir, E., Huber, D., Bonardi, C., Morandell, K., Gutnisky, D., Peron, S., Xu, N., Cox, J., Svoboda, K., 2014a. Procedures for Behavioral Experiments in Head-Fixed Mice. *PLOS ONE* 9, e88678. <https://doi.org/10.1371/journal.pone.0088678>
- Guo, Z.V., Inagaki, H.K., Daie, K., Druckmann, S., Gerfen, C.R., Svoboda, K., 2017. Maintenance of persistent activity in a frontal thalamocortical loop. *Nature* 545, 181–186. <https://doi.org/10.1038/nature22324>
- Guo, Z.V., Li, N., Huber, D., Ophir, E., Gutnisky, D., Ting, J.T., Feng, G., Svoboda, K., 2014b. Flow of Cortical Activity Underlying a Tactile Decision in Mice. *Neuron* 81, 179–194. <https://doi.org/10.1016/j.neuron.2013.10.020>
- Haber, S.N., 2003. The primate basal ganglia: parallel and integrative networks. *J. Chem. Neuroanat., Special Issue on the Human Brain - The Structural Basis for understanding Human Brain function and dysfunction* 26, 317–330. <https://doi.org/10.1016/j.jchemneu.2003.10.003>
- Hanes, D.P., Schall, J.D., 1996. Neural control of voluntary movement initiation. *Science* 274, 427–430. <https://doi.org/10.1126/science.274.5286.427>
- Hanks, T.D., Kopec, C.D., Brunton, B.W., Duan, C.A., Erlich, J.C., Brody, C.D., 2015. Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature* 520, 220–223. <https://doi.org/10.1038/nature14066>
- Hanks, T.D., Summerfield, C., 2017. Perceptual Decision Making in Rodents, Monkeys, and Humans. *Neuron* 93, 15–31. <https://doi.org/10.1016/J.NEURON.2016.12.003>
- Hao, Y., Thomas, A.M., Li, N., 2021. Fully autonomous mouse behavioral and optogenetic experiments in home-cage. *eLife* 10, e66112. <https://doi.org/10.7554/eLife.66112>
- Hara, T., Favale, E., Rossi, G.F., Sacco, G., 1961. Responses in mesencephalic reticular formation and central gray matter evoked by somatic peripheral stimuli. *Exp. Neurol.* 4, 297–309. [https://doi.org/10.1016/0014-4886\(61\)90056-5](https://doi.org/10.1016/0014-4886(61)90056-5)
- Harris, K.D., 2020. Nonsense correlations in neuroscience.

- Harris, K.D., Miller, K.J., 2023. Conditional Randomization Tests for Behavioral and Neural Time Series. <https://doi.org/10.48550/arXiv.2311.03554>
- Haruno, M., Kawato, M., 2006. Different Neural Correlates of Reward Expectation and Reward Expectation Error in the Putamen and Caudate Nucleus During Stimulus-Action-Reward Association Learning. *J. Neurophysiol.* 95, 948–959. <https://doi.org/10.1152/jn.00382.2005>
- Hayashi, Y., Yoshii, N., 1966. MESENCEPHALIC RETICULAR INFLUENCE ON ELECTRICAL ACTIVITIES OF THE PONTINE RETICULAR FORMATION. *Jpn. J. Physiol.* 16, 335–353. <https://doi.org/10.2170/jjphysiol.16.335>
- Heekeren, H.R., Marrett, S., Ungerleider, L.G., 2008. The neural systems that mediate human perceptual decision making. *Nat Rev Neurosci* 9, 467–479. <https://doi.org/10.1038/nrn2374>
- Hervig, M.E., Fiddian, L., Piilgaard, L., Božič, T., Blanco-Pozo, M., Knudsen, C., Olesen, S.F., Alsö, J., Robbins, T.W., 2020. Dissociable and Paradoxical Roles of Rat Medial and Lateral Orbitofrontal Cortex in Visual Serial Reversal Learning. *Cereb. Cortex* 30, 1016–1029. <https://doi.org/10.1093/cercor/bhz144>
- Hintiryan, H., Foster, N.N., Bowman, I., Bay, M., Song, M.Y., Gou, L., Yamashita, S., Bienkowski, M.S., Zingg, B., Zhu, M., Yang, X.W., Shih, J.C., Toga, A.W., Dong, H.-W., 2016a. The mouse cortico-striatal projectome. *Nat. Neurosci.* 19, 1100–1114. <https://doi.org/10.1038/nn.4332>
- Hintiryan, H., Foster, N.N., Bowman, I., Bay, M., Song, M.Y., Gou, L., Yamashita, S., Bienkowski, M.S., Zingg, B., Zhu, M., Yang, X.W., Shih, J.C., Toga, A.W., Dong, H.-W., 2016b. The mouse cortico-striatal projectome. *Nat. Neurosci.* 19, 1100–1100.
- Horwitz, G.D., Batista, A.P., Newsome, W.T., 2004. Representation of an Abstract Perceptual Decision in Macaque Superior Colliculus. *J. Neurophysiol.* 91, 2281–2296. <https://doi.org/10.1152/JN.00872.2003/ASSET/IMAGES/LARGE/Z9K0050438290012.JPEG>
- Horwitz, G.D., Newsome, W.T., 2001. Target selection for saccadic eye movements: prelude activity in the superior colliculus during a direction-discrimination task. *J Neurophysiol* 86, 2543–2558. <https://doi.org/10.1152/jn.2001.86.5.2543>
- Horwitz, G.D., Newsome, W.T., 1999. Separate signals for target selection and movement specification in the superior colliculus. *Science* 284, 1158–1161. <https://doi.org/10.1126/science.284.5417.1158>
- Hoy, J.L., Yavorska, I., Wehr, M., Niell, C.M., 2016. Vision Drives Accurate Approach Behavior during Prey Capture in Laboratory Mice. *Curr Biol* 26, 3046–3052. <https://doi.org/10.1016/j.cub.2016.09.009>
- Huberman, A.D., Niell, C.M., 2011. What can mice tell us about how vision works? *Trends Neurosci.* 34, 464–473. <https://doi.org/10.1016/j.tins.2011.07.002>
- Huda, R., Sipe, G.O., Breton-Provencher, V., Cruz, K.G., Pho, G.N., Adam, E., Gunter, L.M., Sullins, A., Wickersham, I.R., Sur, M., 2020. Distinct prefrontal top-down circuits differentially modulate sensorimotor behavior. *Nat. Commun.* 11, 1–17. <https://doi.org/10.1038/s41467-020-19772-z>
- Hunnicutt, B.J., Jongbloets, B.C., Birdsong, W.T., Gertz, K.J., Zhong, H., Mao, T., 2016. A comprehensive excitatory input map of the striatum reveals novel functional organization. *eLife* 5, 1–32. <https://doi.org/10.7554/eLife.19103>
- Hutmacher, F., 2019. Why Is There So Much More Research on Vision Than on Any Other Sensory Modality? *Front. Psychol.* 10, 2246. <https://doi.org/10.3389/fpsyg.2019.02246>
- Inagaki, H.K., Chen, S., Ridder, M.C., Sah, P., Li, N., Yang, Z., Hasanbegovic, H., Gao, Z., Gerfen, C.R., Svoboda, K., 2022. A midbrain-thalamus-cortex circuit reorganizes cortical dynamics to initiate movement. *Cell* 0. <https://doi.org/10.1016/J.CELL.2022.02.006>
- Inagaki, H.K., Fontolan, L., Romani, S., Svoboda, K., 2019. Discrete attractor dynamics

- underlies persistent activity in the frontal cortex. *Nature* 566, 212–217.
<https://doi.org/10.1038/s41586-019-0919-7>
- International Brain Laboratory, Aguillon-Rodriguez, V., Angelaki, D., Bayer, H., Bonacchi, N., Carandini, M., Cazettes, F., Chapuis, G., Churchland, A.K., Dan, Y., Dewitt, E., Faulkner, M., Forrest, H., Haetzel, L., Häusser, M., Hofer, S.B., Hu, F., Khanal, A., Krasniak, C., Laranjeira, I., Mainen, Z.F., Meijer, G., Miska, N.J., Mrosovsky, T.D., Murakami, M., Noel, J.-P., Pan-Vazquez, A., Rossant, C., Sanders, J., Socha, K., Terry, R., Urai, A.E., Vergara, H., Wells, M., Wilson, C.J., Witten, I.B., Wool, L.E., Zador, A.M., 2021. Standardized and reproducible measurement of decision-making in mice. *Elife* 10. <https://doi.org/10.7554/eLife.63711>
- International Brain Laboratory, Banga, K., Benson, J., Bhagat, J., Biderman, D., Birman, D., Bonacchi, N., Bruijns, S.A., Buchanan, K., Campbell, R.A., Carandini, M., Chapuis, G.A., Churchland, A.K., Davatolhagh, M.F., Lee, H.D., Faulkner, M., Gerçek, B., Hu, F., Huntenburg, J., Hurwitz, C., Khanal, A., Krasniak, C., Langfield, C., Meijer, G.T., Miska, N.J., Mohammadi, Z., Noel, J.-P., Paninski, L., Pan-Vazquez, A., Roth, N., Scharfner, M., Socha, K., Steinmetz, N.A., Svoboda, K., Taheri, M., Urai, A.E., Wells, M., West, S.J., Whiteway, M.R., Winter, O., Witten, I.B., 2024. Reproducibility of in vivo electrophysiological measurements in mice. *eLife* 13. <https://doi.org/10.7554/eLife.100840.1>
- International Brain Laboratory, Benson, B., Benson, J., Birman, D., Bonacchi, N., Carandini, M., Catarino, J.A., Chapuis, G.A., Churchland, A.K., Dan, Y., 2023a. A Brain-Wide Map of Neural Activity during Complex Behaviour. *bioRxiv* 2023–07.
- International Brain Laboratory, Benson, B., Benson, J., Birman, D., Bonacchi, N., Carandini, M., Catarino, J.A., Chapuis, G.A., Churchland, A.K., Dan, Y., Dayan, P., DeWitt, E.E., Engel, T.A., Fabbri, M., Faulkner, M., Fiete, I.R., Findling, C., Freitas-Silva, L., Gerçek, B., Harris, K.D., Häusser, M., Hofer, S.B., Hu, F., Hubert, F., Huntenburg, J.M., Khanal, A., Krasniak, C., Langdon, C., Lau, P.Y.P., Mainen, Z.F., Meijer, G.T., Miska, N.J., Mrosovsky, T.D., Noel, J.-P., Nylund, K., Pan-Vazquez, A., Pouget, A., Rossant, C., Roth, N., Schaeffer, R., Scharfner, M., Shi, Y., Socha, K.Z., Steinmetz, N.A., Svoboda, K., Urai, A.E., Wells, M.J., West, S.J., Whiteway, M.R., Winter, O., Witten, I.B., 2023b. A Brain-Wide Map of Neural Activity during Complex Behaviour. <https://doi.org/10.1101/2023.07.04.547681>
- International Symposium: Reticular Formation Of The Brain, 1957. . *Henry Ford Hosp. Med. J.* 5, 58–63.
- Izenman, A.J., 1975. Reduced-rank regression for the multivariate linear model. *J. Multivar. Anal.* 5, 248–264. [https://doi.org/10.1016/0047-259X\(75\)90042-1](https://doi.org/10.1016/0047-259X(75)90042-1)
- Izquierdo, A., Brigman, J.L., Radke, A.K., Rudebeck, P.H., Holmes, A., 2017. The neural basis of reversal learning: An updated perspective. *Neuroscience* 345, 12–26. <https://doi.org/10.1016/j.neuroscience.2016.03.021>
- Jazayeri, M., Afraz, A., 2017. Navigating the Neural Space in Search of the Neural Code. *Neuron* 93, 1003–1014. <https://doi.org/10.1016/j.neuron.2017.02.019>
- Jefferson, G., 1958. The reticular formation and clinical neurology, in: *Reticular Formation of the Brain*. Little, Brown Boston, pp. 729–738.
- Jones, A.R., Overly, C.C., Sunkin, S.M., 2009. The Allen Brain Atlas: 5 years and beyond. *Nat Rev Neurosci* 10, 821–828.
- Jun, E.J., Bautista, A.R., Nunez, M.D., Allen, D.C., Tak, J.H., Alvarez, E., Basso, M.A., 2021. Causal role for the primate superior colliculus in the computation of evidence for perceptual decisions. *Nat. Neurosci.* 24, 1121–1131. <https://doi.org/10.1038/s41593-021-00878-6>
- Jun, J.J., Steinmetz, N.A., Siegle, J.H., Denman, D.J., Bauza, M., Barbarits, B., Lee, A.K., Anastassiou, C.A., Andrei, A., Aydin, Ç., Barbic, M., Blanche, T.J., Bonin, V., Couto, J.,

- Dutta, B., Gratiy, S.L., Gutnisky, D.A., Häusser, M., Karsh, B., Ledochowitsch, P., Lopez, C.M., Mitelut, C., Musa, S., Okun, M., Pachitariu, M., Putzeys, J., Rich, P.D., Rossant, C., Sun, W.L., Svoboda, K., Carandini, M., Harris, K.D., Koch, C., O'Keefe, J., Harris, T.D., 2017a. Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551, 232–236. <https://doi.org/10.1038/nature24636>
- Jun, J.J., Steinmetz, N.A., Siegle, J.H., Denman, D.J., Bauza, M., Barbarits, B., Lee, A.K., Anastassiou, C.A., Andrei, A., Aydin, Ç., Barbic, M., Blanche, T.J., Bonin, V., Couto, J., Dutta, B., Gratiy, S.L., Gutnisky, D.A., Häusser, M., Karsh, B., Ledochowitsch, P., Lopez, C.M., Mitelut, C., Musa, S., Okun, M., Pachitariu, M., Putzeys, J., Rich, P.D., Rossant, C., Sun, W., Svoboda, K., Carandini, M., Harris, K.D., Koch, C., O'Keefe, J., Harris, T.D., 2017b. Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551, 232–236. <https://doi.org/10.1038/nature24636>
- Kaufman, M.T., Churchland, M.M., Ryu, S.I., Shenoy, K.V., 2014. Cortical activity in the null space: permitting preparation without movement. *Nat. Neurosci.* 17, 440–448. <https://doi.org/10.1038/nn.3643>
- Khilkevich, A., Lohse, M., Low, R., Orsolic, I., Bozic, T., Windmill, P., Masic-Flogel, T.D., 2024. Brain-wide dynamics linking sensation to action during decision-making. *Nature* 1–11. <https://doi.org/10.1038/s41586-024-07908-w>
- Kilmer, W.L., McCulloch, W.S., Blum, J., 1968. Some Mechanisms for a Theory of the Reticular Formation, in: Mesarović, M.D. (Ed.), *Systems Theory and Biology*. Springer, Berlin, Heidelberg, pp. 286–375. https://doi.org/10.1007/978-3-642-88343-9_14
- Kim, H., Sul, J.H., Huh, N., Lee, D., Jung, M.W., 2009. Role of Striatum in Updating Values of Chosen Actions. *J. Neurosci.* 29, 14701–14712. <https://doi.org/10.1523/JNEUROSCI.2728-09.2009>
- Kim, J.N., Shadlen, M.N., 1999. Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nat. Neurosci.* 2, 176–185. <https://doi.org/10.1038/5739>
- Kira, S., Safaai, H., Morcos, A.S., Panzeri, S., Harvey, C.D., 2023. A distributed and efficient population code of mixed selectivity neurons for flexible navigation decisions. *Nat. Commun.* 14, 2121. <https://doi.org/10.1038/s41467-023-37804-2>
- Kolling, N., Wittmann, M.K., Behrens, T.E.J., Boorman, E.D., Mars, R.B., Rushworth, M.F.S., 2016. Value, search, persistence and model updating in anterior cingulate cortex. *Nat. Neurosci.* 19, 1280–1285. <https://doi.org/10.1038/nn.4382>
- Kolmac, C.I., Mitrofanis, J., 1998. Patterns of brainstem projection to the thalamic reticular nucleus. *J. Comp. Neurol.* 396, 531–543. [https://doi.org/10.1002/\(SICI\)1096-9861\(19980713\)396:4<531::AID-CNE9>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1096-9861(19980713)396:4<531::AID-CNE9>3.0.CO;2-2)
- Kolmac, C.I., Power, B.D., Mitrofanis, J., 1998. Patterns of connections between zona incerta and brainstem in rats. *J. Comp. Neurol.* 396, 544–555. [https://doi.org/10.1002/\(SICI\)1096-9861\(19980713\)396:4<544::AID-CNE10>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1096-9861(19980713)396:4<544::AID-CNE10>3.0.CO;2-G)
- Krout, K.E., Belzer, R.E., Loewy, A.D., 2002. Brainstem projections to midline and intralaminar thalamic nuclei of the rat. *J. Comp. Neurol.* 448, 53–101. <https://doi.org/10.1002/cne.10236>
- Lakunina, A., Socha, K.Z., Ladd, A., Bowen, A.J., Chen, S., Colonell, J., Doshi, A., Karsh, B., Krumin, M., Kulik, P., Li, A., Neutens, P., O'Callaghan, J., Olsen, M., Putzeys, J., Tilmans, H.A., Ye, Z., Welkenhuysen, M., Häusser, M., Koch, C., Ting, J.T., Consortium, N.O., Dutta, B., Harris, T.D., Steinmetz, N.A., Svoboda, K., Siegle, J.H., Carandini, M., 2025. Neuropixels Opto: Combining high-resolution electrophysiology and optogenetics. <https://doi.org/10.1101/2025.02.04.636286>
- Lam, N.H., Mukherjee, A., Wimmer, R.D., Nassar, M.R., Chen, Z.S., Halassa, M.M., 2025. Prefrontal transthalamic uncertainty processing drives flexible switching. *Nature* 637, 127–136. <https://doi.org/10.1038/s41586-024-08180-8>
- Lara, A.H., Wallis, J.D., 2015. The Role of Prefrontal Cortex in Working Memory: A Mini Review.

- Front. Syst. Neurosci. 9. <https://doi.org/10.3389/fnsys.2015.00173>
- Le, N.M., Yildirim, M., Wang, Y., Sugihara, H., Jazayeri, M., Sur, M., 2023. Mixtures of strategies underlie rodent behavior during reversal learning. *PLOS Comput. Biol.* 19, e1011430. <https://doi.org/10.1371/journal.pcbi.1011430>
- Lee, A.M., Hoy, J.L., Bonci, A., Wilbrecht, L., Stryker, M.P., Niell, C.M., 2014. Identification of a Brainstem Circuit Regulating Visual Cortical State in Parallel with Locomotion. *Neuron*. <https://doi.org/10.1016/j.neuron.2014.06.031>
- Lee, J., Sabatini, B.L., 2021. Striatal indirect pathway mediates exploration via collicular competition. *Nat.* 2021 1–5. <https://doi.org/10.1038/s41586-021-04055-4>
- Lee, J., Wang, W., Sabatini, B.L., 2020. Anatomically segregated basal ganglia pathways allow parallel behavioral modulation. *Nat. Neurosci.* 23, 1388–1398. <https://doi.org/10.1038/s41593-020-00712-5>
- Li, N., Daie, K., Svoboda, K., Druckmann, S., 2016. Robust neuronal dynamics in premotor cortex during motor planning. *Nature* 532, 459–464. <https://doi.org/10.1038/nature17643>
- Liang, H., Paxinos, G., Watson, C., 2011. Projections from the brain to the spinal cord in the mouse. *Brain Struct. Funct.* 215, 159–186. <https://doi.org/10.1007/s00429-010-0281-x>
- Lindsley, D.B., 1958. The reticular system and perceptual discrimination.
- Lindsley, D.F., Adey, W.R., 1961. Availability of peripheral input to the midbrain reticular formation. *Exp. Neurol.* 4, 358–376. [https://doi.org/10.1016/0014-4886\(61\)90062-0](https://doi.org/10.1016/0014-4886(61)90062-0)
- Liu, D., Gu, X., Zhu, J., Zhang, X., Han, Z., Yan, W., Cheng, Q., Hao, J., Fan, H., Hou, R., Chen, Z., Chen, Y., Li, C.T., 2014. Medial prefrontal activity during delay period contributes to learning of a working memory task. *Science* 346, 458–463. <https://doi.org/10.1126/science.1256573>
- Liu, L.D., Chen, S., Hou, H., West, S.J., Faulkner, M., International Brain Laboratory, Economo, M.N., Li, N., Svoboda, K., 2021. Accurate Localization of Linear Probe Electrode Arrays across Multiple Brains. *eNeuro* 8.
- Lu, X.-C.M., Slotnick, B.M., Silberberg, A.M., 1993. Odor matching and odor memory in the rat. *Physiol. Behav.* 53, 795–804. [https://doi.org/10.1016/0031-9384\(93\)90191-H](https://doi.org/10.1016/0031-9384(93)90191-H)
- Luhmann, C.C., 2009. Temporal decision-making: insights from cognitive neuroscience. *Front. Behav. Neurosci.* 3. <https://doi.org/10.3389/neuro.08.039.2009>
- Mante, V., Sussillo, D., Shenoy, K.V., Newsome, W.T., 2013. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503, 78–84. <https://doi.org/10.1038/nature12742>
- Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W., Bethge, M., 2018. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci* 21, 1281–1289.
- McAlonan, K., Cavanaugh, J.R., Wurtz, R.H., 2008. Guarding the gateway to cortex: attention in visual thalamus. *Nature* 456, 391–391. <https://doi.org/10.1038/nature07382>
- McCormick, D.A., 1989. Cholinergic and noradrenergic modulation of thalamocortical processing. *Trends Neurosci.* 12, 215–221. [https://doi.org/10.1016/0166-2236\(89\)90125-2](https://doi.org/10.1016/0166-2236(89)90125-2)
- McElvain, L.E., Chen, Y., Moore, J.D., Brigidi, G.S., Bloodgood, B.L., Lim, B.K., Costa, R.M., Kleinfeld, D., 2021. Specific populations of basal ganglia output neurons target distinct brain stem areas while collateralizing throughout the diencephalon. *Neuron* 109, 1721–1738.e4. <https://doi.org/10.1016/j.neuron.2021.03.017>
- Merre, P.L., Åhrlund-Richter, S., Carlén, M., 2021. The mouse prefrontal cortex: Unity in diversity. *Neuron* 109, 1925–1944. <https://doi.org/10.1016/j.neuron.2021.03.035>
- Miller, E.K., Cohen, J.D., 2001. An Integrative Theory of Prefrontal Cortex Function. *Annu. Rev. Neurosci.* 24, 167–202. <https://doi.org/10.1146/annurev.neuro.24.1.167>
- Miller, K.J., Botvinick, M.M., Brody, C.D., 2022. Value representations in the rodent orbitofrontal cortex drive learning, not choice. *eLife* 11, e64575. <https://doi.org/10.7554/eLife.64575>

- Miller, K.J., Eckstein, M., Botvinick, M.M., Kurth-Nelson, Z., 2023. Cognitive Model Discovery via Disentangled RNNs. <https://doi.org/10.1101/2023.06.23.546250>
- Milward, K., Busch, K.E., Murphy, R.J., de Bono, M., Olofsson, B., 2011. Neuronal and molecular substrates for optimal foraging in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U. S. A.* 108, 20672–20677. <https://doi.org/10.1073/pnas.1106134109>
- Mitchell, A.S., Chakraborty, S., 2013. What does the mediodorsal thalamus do? *Front. Syst. Neurosci.* 7. <https://doi.org/10.3389/fnsys.2013.00037>
- Mitrofanis, J., 2005. Some certainty for the “zone of uncertainty”? Exploring the function of the zona incerta. *Neuroscience* 130, 1–15. <https://doi.org/10.1016/j.neuroscience.2004.08.017>
- Montijn, J.S., Seignette, K., Howlett, M.H., Cazemier, J.L., Kamermans, M., Levelt, C.N., Heimel, J.A., 2021. A parameter-free statistical test for neuronal responsiveness. *eLife* 10, e71969. <https://doi.org/10.7554/eLife.71969>
- Morris, D.W., Davidson, D.L., 2000. Optimally Foraging Mice Match Patch Use with Habitat Differences in Fitness. *Ecology* 81, 2061–2066. [https://doi.org/10.1890/0012-9658\(2000\)081\[2061:OFMMPU\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2000)081[2061:OFMMPU]2.0.CO;2)
- Moruzzi, G., Magoun, H.W., 1949. Brain stem reticular formation and activation of the EEG. *Electroencephalogr. Clin. Neurophysiol.* 1, 455–473.
- Müller, C.M., Lewandowski, M.H., Singer, W., 1993. Structures mediating cholinergic reticular facilitation of cortical responses in the cat: effects of lesions in immunocytochemically characterized projections. *Exp. Brain Res.* 96, 8–18. <https://doi.org/10.1007/BF00230434>
- Munoz, D.P., Everling, S., 2004. Look away: the anti-saccade task and the voluntary control of eye movement. *Nat. Rev. Neurosci.* 5, 218–228. <https://doi.org/10.1038/nrn1345>
- Musall, S., Kaufman, M.T., Juavinett, A.L., Gluf, S., Churchland, A.K., 2019a. Single-trial neural dynamics are dominated by richly varied movements. *Nat. Neurosci.* 22, 1677–1686. <https://doi.org/10.1038/s41593-019-0502-4>
- Musall, S., Kaufman, M.T., Juavinett, A.L., Gluf, S., Churchland, A.K., 2019b. Single-trial neural dynamics are dominated by richly varied movements. *Nat. Neurosci.* 22, 1677–1686.
- Musall, S., Sun, X.R., Mohan, H., An, X., Gluf, S., Li, S.-J., Drewes, R., Cravo, E., Lenzi, I., Yin, C., Kampa, B.M., Churchland, A.K., 2023. Pyramidal cell types drive functionally distinct cortical activity patterns during decision-making. *Nat. Neurosci.* 26, 495–505. <https://doi.org/10.1038/s41593-022-01245-9>
- Newsome, W.T., Paré, E.B., 1988. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J. Neurosci. Off. J. Soc. Neurosci.* 8, 2201–2211. <https://doi.org/10.1523/JNEUROSCI.08-06-02201.1988>
- Nogueira, R., Rodgers, C.C., Bruno, R.M., Fusi, S., 2023. The geometry of cortical representations of touch in rodents. *Nat. Neurosci.* 26, 239–250. <https://doi.org/10.1038/s41593-022-01237-9>
- Noudoost, B., Chang, M.H., Steinmetz, N.A., Moore, T., 2010. Top-down control of visual attention. *Curr. Opin. Neurobiol.* 20, 183–190. <https://doi.org/10.1016/j.conb.2010.02.003>
- Oh, S.W., Harris, J.A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A.M., Mortrud, M.T., Ouellette, B., Nguyen, T.N., Sorensen, S.A., Slaughterbeck, C.R., Wakeman, W., Li, Y., Feng, D., Ho, A., Nicholas, E., Hirokawa, K.E., Bohn, P., Joines, K.M., Peng, H., Hawrylycz, M.J., Phillips, J.W., Hohmann, J.G., Wahnoutka, P., Gerfen, C.R., Koch, C., Bernard, A., Dang, C., Jones, A.R., Zeng, H., 2014. A mesoscale connectome of the mouse brain. *Nature* 508, 207–214. <https://doi.org/10.1038/nature13186>
- Okazawa, G., Kiani, R., 2023. Neural Mechanisms That Make Perceptual Decisions Flexible. *Annu. Rev. Physiol.* 85, 191–215. <https://doi.org/10.1146/annurev-physiol-031722-024731>

- Ottenheimer, D.J., Hjort, M.M., Bowen, A.J., Steinmetz, N.A., Stuber, G.D., 2023. A stable, distributed code for cue value in mouse cortex during reward learning. *eLife* 12, RP84604. <https://doi.org/10.7554/eLife.84604>
- Padoa-Schioppa, C., Assad, J.A., 2006. Neurons in the orbitofrontal cortex encode economic value. *Nature* 441, 223–226. <https://doi.org/10.1038/nature04676>
- Pandarinath, C., O'Shea, D.J., Collins, J., Jozefowicz, R., Stavisky, S.D., Kao, J.C., Trautmann, E.M., Kaufman, M.T., Ryu, S.I., Hochberg, L.R., Henderson, J.M., Shenoy, K.V., Abbott, L.F., Sussillo, D., 2018. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nat. Methods* 15, 805–815. <https://doi.org/10.1038/s41592-018-0109-9>
- Park, I.M., Meister, M.L.R., Huk, A.C., Pillow, J.W., 2014. Encoding and decoding in parietal cortex during sensorimotor decision-making. *Nat. Neurosci.* 2014 1710 17, 1395–1403. <https://doi.org/10.1038/nn.3800>
- Park, Y.-G., Sohn, C.H., Chen, R., McCue, M., Yun, D.H., Drummond, G.T., Ku, T., Evans, N.B., Oak, H.C., Trieu, W., Choi, H., Jin, X., Lilascharoen, V., Wang, J., Truttmann, M.C., Qi, H.W., Ploegh, H.L., Golub, T.R., Chen, S.-C., Frosch, M.P., Kulik, H.J., Lim, B.K., Chung, K., 2019. Protection of tissue physicochemical properties using polyfunctional crosslinkers. *Nat. Biotechnol.* 37, 73–83. <https://doi.org/10.1038/nbt.4281>
- Pasupathy, A., Miller, E.K., 2005. Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature* 433, 873–876. <https://doi.org/10.1038/nature03287>
- Paule, M.G., Bushnell, P.J., Maurissen, J.P., Wenger, G.R., Buccafusco, J.J., Chelonis, J.J., Elliott, R., 1998. Symposium overview: the use of delayed matching-to-sample procedures in studies of short-term memory in animals and humans. *Neurotoxicol. Teratol.* 20, 493–502. [https://doi.org/10.1016/s0892-0362\(98\)00013-0](https://doi.org/10.1016/s0892-0362(98)00013-0)
- Peng, H., Xie, P., Liu, L., Kuang, X., Wang, Yimin, Qu, L., Gong, H., Jiang, S., Li, A., Ruan, Z., Ding, L., Yao, Z., Chen, C., Chen, M., Daigle, T.L., Dalley, R., Ding, Z., Duan, Y., Feiner, A., He, P., Hill, C., Hirokawa, K.E., Hong, G., Huang, L., Kebede, S., Kuo, H.-C., Larsen, R., Lesnar, P., Li, L., Li, Q., Li, X., Li, Yaoyao, Li, Yuanyuan, Liu, A., Lu, D., Mok, S., Ng, L., Nguyen, T.N., Ouyang, Q., Pan, J., Shen, E., Song, Y., Sunkin, S.M., Tasic, B., Veldman, M.B., Wakeman, W., Wan, W., Wang, P., Wang, Q., Wang, T., Wang, Yaping, Xiong, F., Xiong, W., Xu, W., Ye, M., Yin, L., Yu, Y., Yuan, Jia, Yuan, Jing, Yun, Z., Zeng, S., Zhang, S., Zhao, S., Zhao, Z., Zhou, Z., Huang, Z.J., Esposito, L., Hawrylycz, M.J., Sorensen, S.A., Yang, X.W., Zheng, Y., Gu, Z., Xie, W., Koch, C., Luo, Q., Harris, J.A., Wang, Yun, Zeng, H., 2021. Morphological diversity of single neurons in molecularly defined cell types. *Nature* 598, 174–181. <https://doi.org/10.1038/s41586-021-03941-1>
- Perkins, E., May, P.J., Warren, S., 2014. Feed-forward and feedback projections of midbrain reticular formation neurons in the cat. *Front. Neuroanat.* 7, 55. <https://doi.org/10.3389/fnana.2013.00055>
- Peters, A.J., Fabre, J.M.J., Steinmetz, N.A., Harris, K.D., Carandini, M., 2021. Striatal activity topographically reflects cortical activity. *Nature* 591, 420–425. <https://doi.org/10.1038/s41586-020-03166-8>
- Peterson, E.W., Magoun, H.W., McCulloch, W.S., Lindsley, D.B., 1949. PRODUCTION OF POSTURAL TREMOR. *J. Neurophysiol.* 12, 371–384. <https://doi.org/10.1152/jn.1949.12.6.371>
- Peysakhovich, B., Zhu, O., Tetrick, S.M., Shirhatti, V., Silva, A.A., Li, S., Ibos, G., Rosen, M.C., Johnston, W.J., Freedman, D.J., 2024. Primate superior colliculus is causally engaged in abstract higher-order cognition. *Nat. Neurosci.* 27, 1999–2008. <https://doi.org/10.1038/s41593-024-01744-x>
- Phillips, J.M., Afrasiabi, M., Kambi, N.A., Redinbaugh, M.J., Steely, S., Johnson, E.R., Cheng, X., Fayyad, M., Mohanta, S., Carís, A., Mikell, C.B., Mofakham, S., Saalman, Y.B., 2024. Primate thalamic nuclei select abstract rules and shape prefrontal dynamics.

- <https://doi.org/10.1101/2024.03.13.584871>
- Ratcliff, R., McKoon, G., 2008. The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Comput.* 20, 873–922.
<https://doi.org/10.1162/neco.2008.12-06-420>
- Reinert, S., Hübener, M., Bonhoeffer, T., Goltstein, P.M., 2021. Mouse prefrontal cortex represents learned rules for categorization. *Nature* 1–7.
<https://doi.org/10.1038/s41586-021-03452-z>
- Rescorla, R., Wagner, A., 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement, in: *Classical Conditioning II: Current Research and Theory*.
- Rikhye, R.V., Gilra, A., Halassa, M.M., 2018. Thalamic regulation of switching between cortical representations enables cognitive flexibility. *Nat. Neurosci.* 21, 1753–1763.
<https://doi.org/10.1038/s41593-018-0269-z>
- Rodgers, C.C., DeWeese, M.R., 2014. Neural Correlates of Task Switching in Prefrontal Cortex and Primary Auditory Cortex in a Novel Stimulus Selection Task for Rodents. *Neuron* 82, 1157–1170. <https://doi.org/10.1016/j.neuron.2014.04.031>
- Roitman, J.D., Shadlen, M.N., 2002. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci. Off. J. Soc. Neurosci.* 22, 9475–9489. <https://doi.org/10.1523/JNEUROSCI.22-21-09475.2002>
- Romo, R., de Lafuente, V., 2013. Conversion of sensory signals into perceptual decisions. *Prog. Neurobiol.* 103, 41–75. <https://doi.org/10.1016/j.pneurobio.2012.03.007>
- Romo, R., Hernández, A., Zainos, A., Lemus, L., Brody, C.D., 2002. Neuronal correlates of decision-making in secondary somatosensory cortex. *Nat Neurosci* 5, 1217–1225.
<https://doi.org/10.1038/nn950>
- Roseberry, T.K., Lee, A.M., Lalive, A.L., Wilbrecht, L., Bonci, A., Kreitzer, A.C., 2016. Cell-Type-Specific Control of Brainstem Locomotor Circuits by Basal Ganglia. *Cell* 164, 526–537. <https://doi.org/10.1016/j.cell.2015.12.037>
- Sadtler, P.T., Quick, K.M., Golub, M.D., Chase, S.M., Ryu, S.I., Tyler-Kabara, E.C., Yu, B.M., Batista, A.P., 2014. Neural constraints on learning. *Nature* 512, 423–426.
<https://doi.org/10.1038/nature13665>
- Sara, S.J., 2009. The locus coeruleus and noradrenergic modulation of cognition. *Nat. Rev. Neurosci.* 10, 211–223. <https://doi.org/10.1038/nrn2573>
- Saxena, S., Cunningham, J.P., 2019. Towards the neural population doctrine. *Curr. Opin. Neurobiol., Machine Learning, Big Data, and Neuroscience* 55, 103–111.
<https://doi.org/10.1016/j.conb.2019.02.002>
- Scheibel, A.B., 1984. The Brain Stem Reticular Core and Sensory Function, in: Terjung, R. (Ed.), *Comprehensive Physiology*. Wiley, pp. 213–256.
<https://doi.org/10.1002/cphy.cp010306>
- Scheibel, M., Scheibel, A., Mollica, A., Moruzzi, G., 1955. Convergence and interaction of afferent impulses on single units of reticular formation. *J. Neurophysiol.* 18, 309–331.
<https://doi.org/10.1152/jn.1955.18.4.309>
- Scheibel, M.E., Scheibel, A.B., 1968. The Brain Stem Reticular Core—An Integrative Matrix, in: Mesarović, M.D. (Ed.), *Systems Theory and Biology*. Springer, Berlin, Heidelberg, pp. 261–285. https://doi.org/10.1007/978-3-642-88343-9_13
- Schepens, B., Drew, T., 2006. Descending signals from the pontomedullary reticular formation are bilateral, asymmetric, and gated during reaching movements in the cat. *J Neurophysiol* 96, 2229–2252. <https://doi.org/10.1152/jn.00342.2006>
- Schepens, B., Drew, T., 2004. Independent and convergent signals from the pontomedullary reticular formation contribute to the control of posture and movement during reaching in the cat. *J Neurophysiol* 92, 2217–2238. <https://doi.org/10.1152/jn.01189.2003>
- Schultz, W., Dayan, P., Montague, P.R., 1997. A Neural Substrate of Prediction and Reward.

- Science 275, 1593–1599. <https://doi.org/10.1126/science.275.5306.1593>
- Sébille, S.B., Belaid, H., Philippe, A.-C., André, A., Lau, B., François, C., Karachi, C., Bardinet, E., 2017. Anatomical evidence for functional diversity in the mesencephalic locomotor region of primates. *NeuroImage* 147, 66–78. <https://doi.org/10.1016/j.neuroimage.2016.12.011>
- Shadlen, M.N., Newsome, W.T., 2001. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol.* 86, 1916–1936. <https://doi.org/10.1152/jn.2001.86.4.1916>
- Shamonin, D.P., Bron, E.E., Lelieveldt, B.P.F., Smits, M., Klein, S., Staring, M., 2014. Fast Parallel Image Registration on CPU and GPU for Diagnostic Classification of Alzheimer's Disease. *Front. Neuroinformatics* 7. <https://doi.org/10.3389/fninf.2013.00050>
- Sherman, S.M., Usrey, W.M., 2021. Cortical control of behavior and attention from an evolutionary perspective. *Neuron* 109, 3048–3054. <https://doi.org/10.1016/j.neuron.2021.06.021>
- Shik, M.L., Severin, F.V., Orlovsky, G.N., 1969. Control of walking and running by means of electrical stimulation of the mesencephalon. *Electroencephalogr. Clin. Neurophysiol.* 26, 549.
- Smith, S.M., Krajbich, I., 2021. Mental representations distinguish value-based decisions from perceptual decisions. *Psychon. Bull. Rev.* 28, 1413–1422. <https://doi.org/10.3758/s13423-021-01911-2>
- Steinemann, N.A., Stine, G.M., Trautmann, E.M., Zylberberg, A., Wolpert, D.M., Shadlen, M.N., 2023. Direct observation of the neural computations underlying a single decision. *eLife* 12. <https://doi.org/10.7554/eLife.90859.1>
- Steinmetz, N.A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M., Bauza, M., Beau, M., Bhagat, J., Böhm, C., Broux, M., Chen, S., Colonell, J., Gardner, R.J., Karsh, B., Kloosterman, F., Kostadinov, D., Mora-Lopez, C., O'Callaghan, J., Park, J., Putzeys, J., Sauerbrei, B., Daal, R.J.J. van, Vollen, A.Z., Wang, S., Welkenhuysen, M., Ye, Z., Dudman, J.T., Dutta, B., Hantman, A.W., Harris, K.D., Lee, A.K., Moser, E.I., O'Keefe, J., Renart, A., Svoboda, K., Häusser, M., Haesler, S., Carandini, M., Harris, T.D., 2021a. Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science* 372. <https://doi.org/10.1126/science.abf4588>
- Steinmetz, N.A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M., Bauza, M., Beau, M., Bhagat, J., Böhm, C., Broux, M., Chen, S., Colonell, J., Gardner, R.J., Karsh, B., Kloosterman, F., Kostadinov, D., Mora-Lopez, C., O'Callaghan, J., Park, J., Putzeys, J., Sauerbrei, B., van Daal, R.J.J., Vollen, A.Z., Wang, S., Welkenhuysen, M., Ye, Z., Dudman, J.T., Dutta, B., Hantman, A.W., Harris, K.D., Lee, A.K., Moser, E.I., O'Keefe, J., Renart, A., Svoboda, K., Häusser, M., Haesler, S., Carandini, M., Harris, T.D., 2021b. Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science* 372, eabf4588. <https://doi.org/10.1126/science.abf4588>
- Steinmetz, N.A., Koch, C., Harris, K.D., Carandini, M., 2018. Challenges and opportunities for large-scale electrophysiology with Neuropixels probes. *Curr. Opin. Neurobiol.* 50, 92–100.
- Steinmetz, N.A., Zatka-Haas, P., Carandini, M., Harris, K.D., 2019. Distributed coding of choice, action and engagement across the mouse brain. *Nature* 576, 266–273. <https://doi.org/10.1038/s41586-019-1787-x>
- Steriade, M., Glenn, L.L., 1982. Neocortical and caudate projections of intralaminar thalamic neurons and their synaptic excitation from midbrain reticular core. *J. Neurophysiol.* 48, 352–371. <https://doi.org/10.1152/jn.1982.48.2.352>
- Sterzer, P., Voss, M., Schlagenhauf, F., Heinz, A., 2019. Decision-making in schizophrenia: A predictive-coding perspective. *Neuroimage* 190, 133–143. <https://doi.org/10.1016/j.neuroimage.2018.05.074>

- Stine, G.M., Trautmann, E.M., Jeurissen, D., Shadlen, M.N., 2023. A neural mechanism for terminating decisions. *Neuron* 111, 2601–2613.e5. <https://doi.org/10.1016/j.neuron.2023.05.028>
- Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C.B., Carandini, M., Harris, K.D., 2019. Spontaneous behaviors drive multidimensional, brainwide activity. *Science* 364. <https://doi.org/10.1126/science.aav7893>
- Stuphorn, V., Hoffmann, K.P., Miller, L.E., 1999. Correlation of primate superior colliculus and reticular formation discharge with proximal limb muscle activity. *J. Neurophysiol.* 81, 1978–1982. <https://doi.org/10.1152/JN.1999.81.4.1978/ASSET/IMAGES/LARGE/9K0490144003.JPG>
- Svoboda, K., Li, N., 2018. Neural mechanisms of movement planning: motor cortex and beyond. *Curr. Opin. Neurobiol.* 49, 33–41. <https://doi.org/10.1016/j.conb.2017.10.023>
- Thevathasan, W., Debu, B., Aziz, T., Bloem, B.R., Blahak, C., Butson, C., Czernecki, V., Foltynie, T., Fraix, V., Grabli, D., Joint, C., Lozano, A.M., Okun, M.S., Ostrem, J., Pavese, N., Schrader, C., Tai, C.-H., Krauss, J.K., Moro, E., on behalf of the Movement Disorders Society PPN DBS Working Group in collaboration with the World Society for Stereotactic and Functional Neurosurgery, 2018. Pedunculo pontine nucleus deep brain stimulation in Parkinson's disease: A clinical review. *Mov. Disord.* 33, 10–20. <https://doi.org/10.1002/mds.27098>
- Thomas, A., Yang, W., Wang, C., Tipparaju, S.L., Chen, G., Sullivan, B., Swiekatowski, K., Tatam, M., Gerfen, C., Li, N., 2023. Superior colliculus bidirectionally modulates choice activity in frontal cortex. *Nat. Commun.* 14, 7358. <https://doi.org/10.1038/s41467-023-43252-9>
- Tseng, S.-Y., Chettih, S.N., Arlt, C., Barroso-Luque, R., Harvey, C.D., 2022. Shared and specialized coding across posterior cortical areas for dynamic navigation decisions. *Neuron* 110, 2484–2502.e16. <https://doi.org/10.1016/j.neuron.2022.05.012>
- Tsunada, J., Liu, A.S.K., Gold, J.I., Cohen, Y.E., 2016. Causal contribution of primate auditory cortex to auditory perceptual decision-making. *Nat. Neurosci.* 19, 135–142. <https://doi.org/10.1038/nn.4195>
- Uddin, L.Q., 2021. Cognitive and behavioural flexibility: neural mechanisms and clinical considerations. *Nat. Rev. Neurosci.* 22, 167–179. <https://doi.org/10.1038/s41583-021-00428-w>
- Waitzman, D.M., Pathmanathan, J., Presnell, R., Ayers, A., DePalma, S., 2002. Contribution of the Superior Colliculus and the Mesencephalic Reticular Formation to Gaze Control. *Ann. N. Y. Acad. Sci.* 956, 111–129. <https://doi.org/10.1111/j.1749-6632.2002.tb02813.x>
- Wallis, J.D., Anderson, K.C., Miller, E.K., 2001. Single neurons in prefrontal cortex encode abstract rules. *Nature* 411, 953–956. <https://doi.org/10.1038/35082081>
- Wang, L., McAlonan, K., Goldstein, S., Gerfen, C.R., Krauzlis, R.J., 2020. A Causal Role for Mouse Superior Colliculus in Visual Perceptual Decision-Making. *J. Neurosci.* 40, 3768–3782. <https://doi.org/10.1523/JNEUROSCI.2642-19.2020>
- Wang, Q., Ding, S.-L., Li, Y., Royall, J., Feng, D., Lesnar, P., Graddis, N., Naeemi, M., Facer, B., Ho, A., Dolbeare, T., Blanchard, B., Dee, N., Wakeman, W., Hirokawa, K.E., Szafer, A., Sunkin, S.M., Oh, S.W., Bernard, A., Phillips, J.W., Hawrylycz, M., Koch, C., Zeng, H., Harris, J.A., Ng, L., 2020. The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell* 181, 936–953.e20. <https://doi.org/10.1016/j.cell.2020.04.007>
- Wang, T.-Y., Liu, J., Yao, H., 2020. Control of adaptive action selection by secondary motor cortex during flexible visual categorization. *Elife* 9.
- Wang, X., Chou, X., Zhang, L.I., Tao, H.W., 2020. Zona Incerta: An Integrative Node for Global Behavioral Modulation. *Trends Neurosci.* 43, 82–87. <https://doi.org/10.1016/j.tins.2019.11.007>

- Wang, X.-J., 2008. Decision Making in Recurrent Neuronal Circuits. *Neuron* 60, 215.
<https://doi.org/10.1016/J.NEURON.2008.09.034>
- Warren, S., Waitzman, D.M., May, P.J., 2008. Anatomical Evidence for Interconnections between the Central Mesencephalic Reticular Formation and Cervical Spinal Cord in the Cat and Macaque. *Anat. Rec. Hoboken NJ* 2007 291, 141–160.
<https://doi.org/10.1002/ar.20644>
- Waydo, S., Kraskov, A., Quiroga, R.Q., Fried, I., Koch, C., 2006. Sparse Representation in the Human Medial Temporal Lobe. *J. Neurosci.* 26, 10232–10234.
<https://doi.org/10.1523/JNEUROSCI.2101-06.2006>
- Weiler, S., Guggiana Nilo, D., Bonhoeffer, T., Hübener, M., Rose, T., Scheuss, V., 2023. Functional and structural features of L2/3 pyramidal cells continuously covary with pial depth in mouse visual cortex. *Cereb. Cortex* 33, 3715–3733.
<https://doi.org/10.1093/cercor/bhac303>
- Werner, W., Dannenberg, S., Hoffmann, K.P., 1997. Arm-movement-related neurons in the primate superior colliculus and underlying reticular formation: comparison of neuronal activity with EMGs of muscles of the shoulder, arm and trunk during reaching. *Exp. Brain Res.* 197 1152 115, 191–205. <https://doi.org/10.1007/PL00005690>
- Wilson, R.C., Collins, A.G., 2019. Ten simple rules for the computational modeling of behavioral data. *eLife* 8, e49547. <https://doi.org/10.7554/eLife.49547>
- Wilson, R.C., Takahashi, Y.K., Schoenbaum, G., Niv, Y., 2014. Orbitofrontal Cortex as a Cognitive Map of Task Space. *Neuron* 81, 267–279.
<https://doi.org/10.1016/j.neuron.2013.11.005>
- Winnubst, J., Bas, E., Ferreira, T.A., Wu, Z., Economo, M.N., Edson, P., Arthur, B.J., Bruns, C., Rokicki, K., Schauder, D., Olbris, D.J., Murphy, S.D., Ackerman, D.G., Arshadi, C., Baldwin, P., Blake, R., Elsayed, A., Hasan, M., Ramirez, D., Dos Santos, B., Weldon, M., Zafar, A., Dudman, J.T., Gerfen, C.R., Hantman, A.W., Korff, W., Sternson, S.M., Spruston, N., Svoboda, K., Chandrashekar, J., 2019. Reconstruction of 1,000 Projection Neurons Reveals New Cell Types and Organization of Long-Range Connectivity in the Mouse Brain. *Cell* 179, 268–281.e13. <https://doi.org/10.1016/j.cell.2019.07.042>
- Wu, Z., Litwin-Kumar, A., Shammash, P., Taylor, A., Axel, R., Shadlen, M.N., 2020. Context-Dependent Decision Making in a Premotor Circuit. *Neuron* 106, 316–328.e6.
<https://doi.org/10.1016/J.NEURON.2020.01.034>
- Xiong, Q., Znamenskiy, P., Zador, A.M., 2015. Selective corticostriatal plasticity during acquisition of an auditory discrimination task. *Nature* 521, 348–351.
<https://doi.org/10.1038/nature14225>
- Yang, E., Tadin, D., Glasser, D.M., Hong, S.W., Blake, R., Park, S., 2013. Visual context processing in schizophrenia. *Clin. Psychol. Sci. B J. Assoc. Psychol. Sci.* 1, 5–15.
<https://doi.org/10.1177/2167702612464618>
- Yao, Z., van Velthoven, C.T.J., Kunst, M., Zhang, M., McMillen, D., Lee, C., Jung, W., Goldy, J., Abdelhak, A., Aitken, M., Baker, K., Baker, P., Barkan, E., Bertagnolli, D., Bhandiwad, A., Bielstein, C., Bishwakarma, P., Campos, J., Carey, D., Casper, T., Chakka, A.B., Chakrabarty, R., Chavan, S., Chen, M., Clark, M., Close, J., Crichton, K., Daniel, S., DiValentin, P., Dolbeare, T., Ellingwood, L., Fiabane, E., Fliss, T., Gee, J., Gerstenberger, J., Glandon, A., Gloe, J., Gould, J., Gray, J., Guilford, N., Guzman, J., Hirschstein, D., Ho, W., Hooper, M., Huang, M., Hupp, M., Jin, K., Kroll, M., Lathia, K., Leon, A., Li, S., Long, B., Madigan, Z., Malloy, J., Malone, J., Maltzer, Z., Martin, N., McCue, R., McGinty, R., Mei, N., Melchor, J., Meyerdierks, E., Mollenkopf, T., Moonsman, S., Nguyen, T.N., Otto, S., Pham, T., Rimorin, C., Ruiz, A., Sanchez, R., Sawyer, L., Shapovalova, N., Shepard, N., Slaughterbeck, C., Sulc, J., Tieu, M., Torkelson, A., Tung, H., Valera Cuevas, N., Vance, S., Wadhwani, K., Ward, K., Levi, B., Farrell, C., Young, R., Staats, B., Wang, M.-Q.M., Thompson, C.L., Mufti, S., Pagan, C.M., Kruse, L., Dee, N., Sunkin,

- S.M., Esposito, L., Hawrylycz, M.J., Waters, J., Ng, L., Smith, K., Tasic, B., Zhuang, X., Zeng, H., 2023. A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature* 624, 317–332. <https://doi.org/10.1038/s41586-023-06812-z>
- Ye, Z., Bull, M.S., Li, A., Birman, D., Daigle, T.L., Tasic, B., Zeng, H., Steinmetz, N.A., 2023. Brain-wide topographic coordination of traveling spiral waves. <https://doi.org/10.1101/2023.12.07.570517>
- Ye, Z., Shelton, A.M., Shaker, J.R., Boussard, J., Colonell, J., Birman, D., Manavi, S., Chen, S., Windolf, C., Hurwitz, C., Namima, T., Pedraja, F., Weiss, S., Raducanu, B., Ness, T.V., Jia, X., Mastroberardino, G., Rossi, L.F., Carandini, M., Häusser, M., Einevoll, G.T., Laurent, G., Sawtell, N.B., Bair, W., Pasupathy, A., Lopez, C.M., Dutta, B., Paninski, L., Siegle, J.H., Koch, C., Olsen, S.R., Harris, T.D., Steinmetz, N.A., 2024. Ultra-high density electrodes improve detection, yield, and cell type identification in neuronal recordings. <https://doi.org/10.1101/2023.08.23.554527>
- Yizhar, O., Fenno, L.E., Davidson, T.J., Mogri, M., Deisseroth, K., 2011. Optogenetics in Neural Systems. *Neuron* 71, 9–34. <https://doi.org/10.1016/j.neuron.2011.06.004>
- Yule, G.U., 1926. Why do we Sometimes get Nonsense-Correlations between Time-Series?--A Study in Sampling and the Nature of Time-Series. *J. R. Stat. Soc.* 89, 1–63. <https://doi.org/10.2307/2341482>
- Zatka-Haas, P., Steinmetz, N.A., Carandini, M., Harris, K.D., 2021. Sensory coding and the causal impact of mouse cortex in a visual decision. *eLife*.
- Zeng, H., Madisen, L., 2012. Mouse transgenic approaches in optogenetics. *Prog. Brain Res.* 193–213.
- Zhong, L., Zhang, Y., Duan, C.A., Deng, J., Pan, J., Xu, N., 2019. Causal contributions of parietal cortex to perceptual decision-making during stimulus categorization. *Nat. Neurosci.* 22, 963–973. <https://doi.org/10.1038/s41593-019-0383-6>

Appendix A: Ultra-high density electrodes improve detection, yield, and cell type identification in neuronal recordings

Zhiwen Ye^{1*}, Andrew M. Shelton^{2,3*}, Jordan R. Shaker¹, Julien Boussard⁴, Jennifer Colonell⁵, Daniel Birman¹, Sahar Manavi², Susu Chen⁵, Charlie Windolf⁴, Cole Hurwitz⁴, Han Yu⁴, Tomoyuki Namima^{1,6}, Federico Pedraja⁴, Shahaf Weiss⁷, Bogdan Raducanu⁸, Torbjørn V. Ness⁹, Xiaoxuan Jia¹⁰, Giulia Mastroberardino^{11,12}, L. Federico Rossi¹³, Matteo Carandini¹¹, Michael Häusser¹², Gaute T. Einevoll^{9,14}, Gilles Laurent⁷, Nathaniel B. Sawtell⁴, Wyeth Bair^{1,6}, Anitha Pasupathy^{1,6}, Carolina Mora Lopez⁸, Barun Dutta⁸, Liam Paninski⁴, Joshua H. Siegle³, Christof Koch², Shawn R. Olsen^{2,3}, Timothy D. Harris^{5,15}, Nicholas A. Steinmetz^{1#}

¹ Department of Neurobiology and Biophysics, University of Washington, Seattle, WA, USA

² MindScope Program, Allen Institute, Seattle, WA, USA

³ Allen Institute for Neural Dynamics, Seattle, WA, USA

⁴ Department of Neuroscience, Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY, USA

⁵ Janelia Research Campus, Ashburn, VA, USA

⁶ Washington National Primate Research Center, Seattle, WA, USA

⁷ Max Planck Institute for Brain Research, Frankfurt, Germany

⁸ IMEC, Leuven, Belgium

⁹ Norwegian University of Life Sciences, Ås, Norway

¹⁰ Center for Life Sciences & IDG/McGovern Institute for Brain Research, Tsinghua University, China

¹¹ UCL Institute of Ophthalmology, University College London, London, UK

¹² Wolfson Institute for Biomedical Research, University College London, London, UK

¹³ Center for Neuroscience and Cognitive Systems, Istituto Italiano di Tecnologia, Rovereto, Italy

¹⁴ University of Oslo, Oslo, Norway

¹⁵ Biomedical Engineering Department, Johns Hopkins University, Baltimore, MD, USA

* equal contribution

Correspondence to nick.steinmetz@gmail.com

Abstract

To understand the neural basis of behavior, it is essential to sensitively and accurately measure neural activity at single neuron and single spike resolution. Extracellular electrophysiology delivers this, but it has biases in the neurons it detects and it imperfectly resolves their action potentials. To minimize these limitations, we developed a silicon probe with much smaller and denser recording sites than previous designs, called Neuropixels Ultra (*NP Ultra*). This device samples neuronal activity at ultra-high spatial density (~10 times higher than previous probes) with minimally increased noise levels, while trading off recording span. NP Ultra is effectively an implantable voltage-sensing camera that captures a planar image of a neuron's electrical field. We use a spike sorting algorithm optimized for these probes to demonstrate that the yield of

visually-responsive neurons in recordings from mouse visual cortex increases more than 2-fold. Using NP Ultra recordings, we discovered that a feature of extracellular waveforms, the spatial extent or “footprint”, distinguished axonal from somatic recordings. Small footprint waveforms were observed across multiple brain regions and four species. In addition to axons, spiking in dendritic compartments could be recorded at substantially improved signal quality relative to other probes. Using high-resolution waveform features recorded with NP Ultra, including footprint, three genetically-identified cortical cell types could be discriminated from one another with ~80% accuracy and from other neurons with ~85% accuracy. NP Ultra improves recording yield, detection of subcellular compartments, and cell type discrimination to enable more powerful dissection of neural circuit activity during behavior for a variety of use cases.

Introduction

High-density silicon electrode arrays, such as Neuropixels probes, have enabled prolonged recordings of hundreds to thousands of neurons across many brain regions in a single experiment (Jun et al., 2017; Steinmetz et al., 2018, 2021). Their versatility and robustness has enabled collection of these datasets in diverse species including fish, rodents, monkeys, reptiles, and humans (Steinmetz et al., 2019; Norimoto et al., 2020; Metzen and Chacron, 2021; Siegle et al., 2021; Chung et al., 2022; Paulk et al., 2022; Trautmann et al., 2023). The use of Neuropixels has driven discoveries and insights into the nature of decision making, perception, and the brain-wide dynamics of neuronal processing (Allen et al., 2019; Musall et al., 2019; Steinmetz et al., 2019; Stringer et al., 2019; Vesuna et al., 2020; Siegle et al., 2021; Gardner et al., 2022; Jia et al., 2022; Chen et al., 2023; International Brain Laboratory et al., 2023b).

Despite these successes, key technical challenges impede progress on a number of pressing issues in neuroscience. First, the spatial resolution of previous generations of Neuropixels probes is relatively low: the nearest contact-to-contact spacing is ~25 μm in Neuropixels 1.0 and ~15 μm in 2.0. Accordingly, previous probe generations do not have sufficient site density to optimally sample fine-scale brain structures such as small nuclei or thin cell layers. In addition, our ability to sample electrical fields of individual neurons at a high spatial resolution has been limited. While we have advanced knowledge of electrical fields at the columnar level and beyond, such as the LFP and EEG (Buzsaki, 2006; Nunez and Srinivasan, 2006; Katzner et al., 2009; Buzsáki et al., 2012; Halmes et al., 2024), we have been unable to access the microstructure of electrical fields at the micrometer scale. Moreover, the low density of previous probes may undersample extracellular action potentials with electrical “footprints” (i.e. the detectable spatial extent of the extracellular action potential) smaller than the span between electrode contacts, and therefore create a sampling bias against such small-footprint signals. If such small-footprint signals are present in neural tissue, this bias against them would limit the yield of recordings and may contribute to the 'dark matter problem' in neuroscience (Shoham et al., 2006).

A second challenge for large-scale electrophysiology is motion of the brain relative to the probe, which can cause spikes to drift between recording sites, posing problems for accurate spike sorting (Steinmetz et al., 2021; Windolf et al., 2025). Probes with insufficient recording site

density may even “lose” neurons entirely as their spikes become undetectable if the neuron’s extracellular action potential falls into the gaps between sites. Conversely, a higher density of recording sites and thus higher spatial sampling would allow for a more accurate estimate of brain motion, improved ability to record all spikes from neurons, and improved *post-hoc* ability to remove associated sorting errors from electrophysiological recordings.

In addition, the brain contains a large diversity of cell types, and extracellular electrophysiology has been limited in its ability to discriminate between these types (Gouwens et al., 2020; Yao et al., 2021). Classically, extracellularly recorded neurons in some brain regions such as cortex and striatum have been separated into coarse cell types on the basis of features such as waveform shape and firing pattern (McCormick et al., 1985; Barthó et al., 2004; Mitchell et al., 2007; Niell and Stryker, 2008; Yamin et al., 2013; Roux et al., 2014; Senzai and Buzsáki, 2017; Yu et al., 2019; Lee et al., 2021, 2024; Beau et al., 2024). NP 1.0 probes can provide additional coarse morphological information about the electrical field surrounding a neuron useful for neuron classification (Buccino et al., 2018; Jia et al., 2019), but this information may not be detailed enough to resolve the large diversity of cell types in the brain. The possibility that higher resolution electrode arrays could improve cell type classification remains largely unexplored.

We hypothesized that an electrophysiological probe with increased sampling density would address each of the above technical challenges by giving access to electrical field microstructures. Here we describe ‘Neuropixels Ultra’, a device with substantially smaller and denser recording sites than previous devices. We found that this technology increases spike sorting signal-to-noise ratio (SNR) and yield. The higher density sites reduced sampling bias, revealing more clearly a population of small-footprint units in many brain regions and in multiple species. Pharmacological experiments demonstrated that small footprint units correspond to recordings from axons. Opto-tagging experiments further confirmed that these narrow-spiking, small footprint units did not correspond to any single interneuron class. These results demonstrate that spatial footprint is a quantifiable characteristic, alongside waveform duration, that can be used in extracellular electrophysiology experiments to separate recorded unit identities. Moreover, these probes improved cell type classification, tested for three major neocortical interneuron types expressing parvalbumin (PV), somatostatin (SST), and vasoactive intestinal peptide (VIP). These could be classified with approximately 80% accuracy from each other, and with ~85% accuracy from un-tagged neurons. Classification accuracy was significantly better with NP Ultra than could be achieved with NP 1.0 recordings. The advantages in yield, detection, and cell-class specificity make NP Ultra a superior probe to NP 2.0 for experiments requiring those properties or for experiments focusing on small brain regions, while NP2.0 remains superior for studies requiring large recording span. Finally, we share a dataset of thousands of neuronal waveforms recorded with NP Ultra from over a dozen brain regions and four species as a resource for understanding the biophysical basis of diverse extracellular action potentials across the brain, along with an interactive browsing and visualization website (<https://npultra.steinmetzlab.net/>). NP Ultra probes open a window into the fine-grained spatio-temporal structure of electrically excitable cells in behaving animals.

Results

Probe design and characterization

We designed a new version of the Neuropixels probe with substantially smaller and denser recording sites, called Neuropixels Ultra (NP Ultra). The probe has 5 x 5 μm titanium nitride (TiN) recording sites on an 48 x 8 grid with 1 μm gaps (for 6 μm center-to-center spacing), densely sampling a 288 x 48 μm span of brain tissue (**Fig. 1A-D**). The site size and spacing compare favorably with Neuropixels 1.0 (NP 1.0) probes with 12 x 12 μm electrodes and 20 x 16 μm staggered spacing. The probe form-factor (shank dimensions and probe base) are identical to NP 1.0, as are mechanical characteristics. Like existing NP 1.0 and 2.0 probes (Jun et al., 2017; Steinmetz et al., 2021), NP Ultra records from 384 channels simultaneously. NP Ultra features switchable recording sites for multiple configurations including 96 x 4, 192 x 2, and 384 x 1 site arrangements for longer span at reduced density (**Fig 1E, Supp. Fig. S1**), achieving this flexibility using a grouped switching arrangement in which multiple sites share switch memory due to space constraints. The configurability also enables selecting sites in these configurations at various positions from among the 6,144 total sites spanning ~4.6 mm along the probe. NP Ultra has substantially higher site density (1.3 sites/ μm vs 0.10 and 0.13 sites/ μm for 1.0 and 2.0 probes, **Fig. 1B**), at the cost of reduced recording span when in the maximal density configuration (48 x 8). This increased recording site density provides much higher spatial resolution of extracellular voltages compared to lower-density probes (**Fig. 1C,D,E**). Recording individual neurons *in vivo* with this device revealed detailed portraits of their extracellular action potentials (**Fig. 1H**).

NP Ultra's smaller recording sites have larger electrical impedance and, accordingly, slightly larger per-channel noise levels, as predicted (López et al., 2012), but otherwise retain the uniform gain, low crosstalk, and low light sensitivity of other Neuropixels probes. In self-referenced noise measurements of NP 1.0 and NP Ultra probes in saline, we observe a small but significant difference in RMS noise (**Fig. 1G**; 0.26 μV ; 95% confidence interval from t-test = 0.19-0.34 μV). Estimating noise within tissue is complicated by the presence of true biological signals, which typically far exceed electrode noise. We obtain an approximation by selecting sites with low activity levels and estimating the noise with the median absolute deviation, which is less sensitive to outliers than RMS (Rey et al., 2015). We found that noise levels *in vivo* were $20 \pm 2\%$ higher for NP Ultra than for NP 1.0 (**Fig. 1G**; NP1.0: 10.5 μV RMS *in vivo*; NP Ultra: 12.67 μV ; difference = 2.1 μV ; 95% confidence interval from t-test = 2.0-2.3 μV). The difference in noise between NP Ultra and 1.0, and the difference between saline and *in vivo* measurements, is well-captured by a noise model incorporating the properties of TiN electrodes and probe electronics, with higher impedance *in vivo* and a contribution of background neural activity (**Supp. Fig. S2**; see Methods). Though per-site noise levels are higher for these probes, we found they nevertheless have highly uniform channel-to-channel gain (**Supp. Fig. S3**), negligible crosstalk (**Supp. Fig. S4**), and no larger light sensitivity than probes with larger sites (**Supp. Fig. S5**). Despite the small site size, recordings of local field potentials (LFPs) were

similar to probes with larger sites, revealing characteristic LFP features of the hippocampus (Supp Fig. S6).

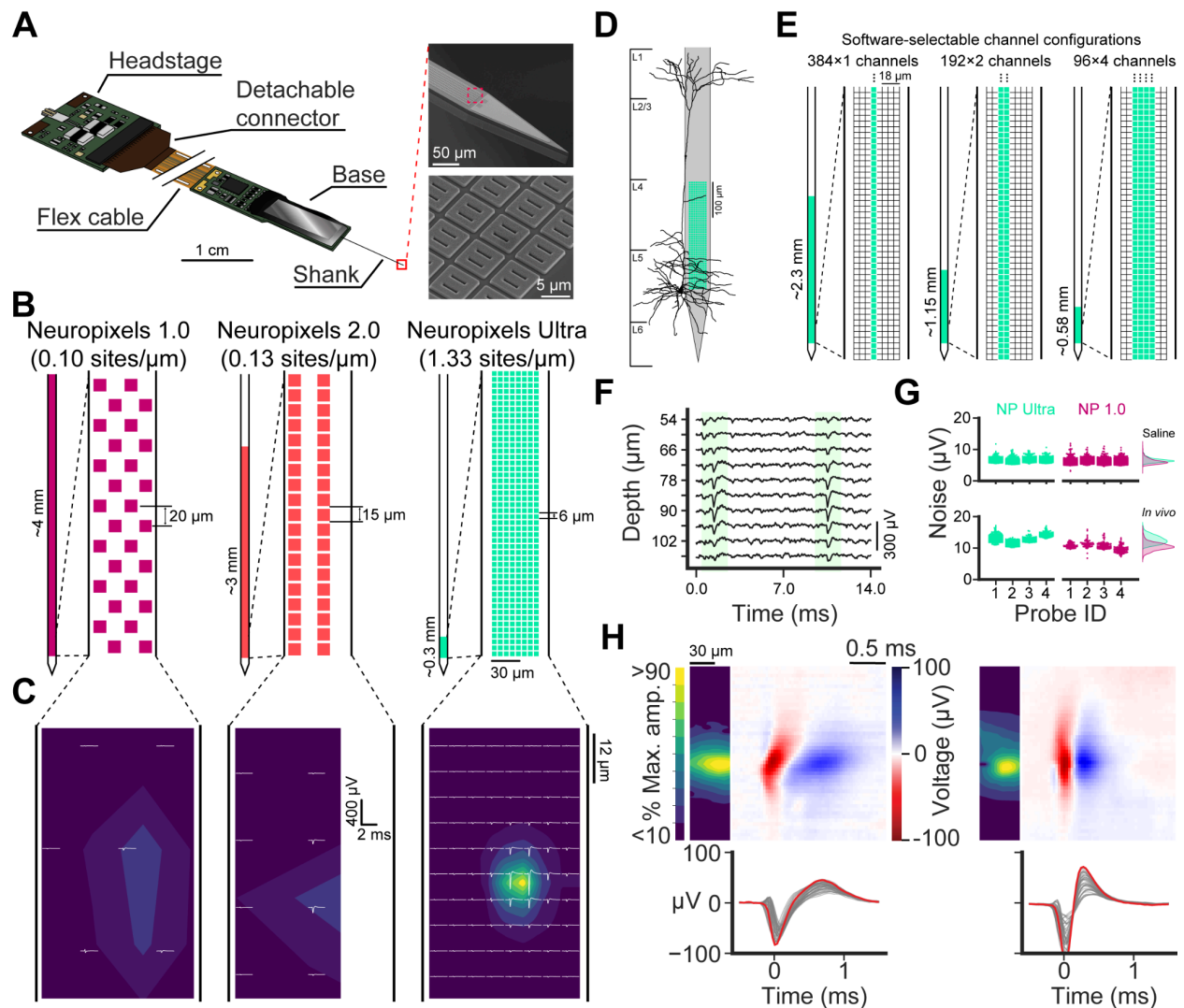


Figure 1. Neuropixels Ultra probes have a considerably denser site layout over a smaller spatial range compared to NP 1.0 and 2.0. **A** (left), Schematic of a complete NP Ultra probe, including headstage and flex cable. (right) Scanning electron microscope images of an NP Ultra probe tip (top) and individual contacts, each with two slits that increase electrode surface area (bottom, top inset). **B**, Layout of NP Ultra sites compared to previous probes. **C**, Comparison of an example waveform on all three probes. The NP Ultra panel (right) shows the mean waveform of a neuron recorded in the mouse primary visual cortex; this waveform was spatially re-sampled to approximate the signal on the 12 x 12 μm grid for NP 1.0 (left) and NP 2.0 (middle) site configurations. Heatmaps represent the interpolated voltage as proportion of the trough-to-peak amplitude recorded on the peak amplitude channel in the NP Ultra configuration. **D**, Example reconstructed morphology of a mouse L5 pyramidal neuron from the Allen Institute Cell Types Database (ID: 488679042) compared at scale to a NP Ultra probe. **E**, Schematics of span and site layout for various probe configurations using active NP Ultra probes. **F**, Example raw traces from a column of vertically adjacent recording sites in the 48 x 8 channel configuration, plotted according to their depth along on an NP Ultra probe. Shading indicates the time and location of two detected spikes.

G, Measures of noise in saline (top, root-mean-square) and during recording *in vivo* (bottom, median absolute deviation) for four NP Ultra and four NP 1.0 probes. **H**, Example spatial (top left), spatiotemporal (top right) and temporal (bottom) waveforms recorded by NP Ultra from a regular-spiking (left) and a narrow-spiking (right) unit in the visual cortex of an awake mouse. Spatial plots show the contour map of channel-wise maximum amplitudes normalized to the peak channel. The spatio-temporal plots represent voltage as a colormap across all channels in the column of sites containing the peak amplitude site. The temporal plots show the peak amplitude channel (red) with 40 nearby channels (gray).

Higher site density improves the quality and yield of spike sorting

We reasoned that a probe with ten-fold higher site density would provide improved data quality despite slightly higher noise levels, as it would sample individual action potentials at more recording sites. As detailed below, this increased channel density could lead to several advantages: 1) Higher peak waveform amplitudes; 2) Higher signal-to-noise ratio (SNR) for each waveform; 3) Improved spatial localization of individual action potential sources; and 4) Improved ability to stably record and track all spikes generated by a single neuron during relative motion of the probe and the brain. Before examining these four aspects of data quality, we first describe a spike sorting algorithm, DARTsort (Boussard et al., 2023), that is well suited to this dense data. We test the algorithm's performance in challenging validation experiments, in which we recorded from mouse visual cortex and intentionally moved the probe by adjusting the position of the micromanipulator in a triangle-wave pattern for ~500 seconds during the session (**Fig. 2**), a strategy that provides a difficult test case for the algorithm with an approximately known pattern of motion.

The algorithm, described in detail in the Methods and demonstrated on simulated data in separate work (Boussard et al., 2023), relies on estimating the 3D spatial location of each spike relative to the probe using a simple point-source model (Methods, (Boussard et al., 2021)). We use these locations to infer the motion of the probe relative to the brain over time (Windolf et al., 2025), register the spike positions, and cluster the spikes (**Fig. 2A,B**). Algorithms that infer spatial locations of spikes and use these for clustering have proven successful on data recorded *in vitro* (Hilgen et al., 2017), but here we find that this strategy can now also be used *in vivo* given the improved ability of NP Ultra to resolve spike locations relative to NP 1.0 and 2.0. Using these cluster assignments, we then estimate “super-resolved” template waveforms for each unit at different probe locations, by binning spikes localized within fine spatial grids on the probe according to the non-registered depth positions of these spikes (**Fig. 2C**). These templates capture the waveform variability within each unit due to probe drift and are used to improve template-matching accuracy further in the following steps, and are “super-resolved” in the sense that they are defined on a fine spatial grid (here chosen to be 2 μm spacing though the probe only samples at 6 μm resolution). Using computational simulations of extracellular waveforms, we estimated that this spatial sampling (2 μm) should be sufficient to capture the relevant details of the extracellular action potentials along the length of the probe (**Supp. Fig. S7**). Computing the super-resolution template, and shifting it according to a precise motion estimate, provides an accurate approximation of the neuron's spike shapes at any time in the recording, demonstrated by small residuals (**Fig. 2C**). Finally, these templates are used to deconvolve the original data, extracting spike times for each neuron and resolving

spatiotemporally overlapping spikes. During template deconvolution, the estimated super-resolution templates are shifted according to the estimated motion (rather than attempting to shift and interpolate the raw data, as in Kilosort 2.5 (Steinmetz et al., 2021)) to infer a waveform shape best matching each spike at that moment in time (**Fig. 2D**). Using this sorting algorithm, we automatically sorted about one hundred neurons from recordings with NP Ultra probes in mouse visual cortex (147 ± 37 total neurons per recording, 44 ± 22 neurons with reliable visual responses, mean $\pm$ S.D., $n=10$ recordings).

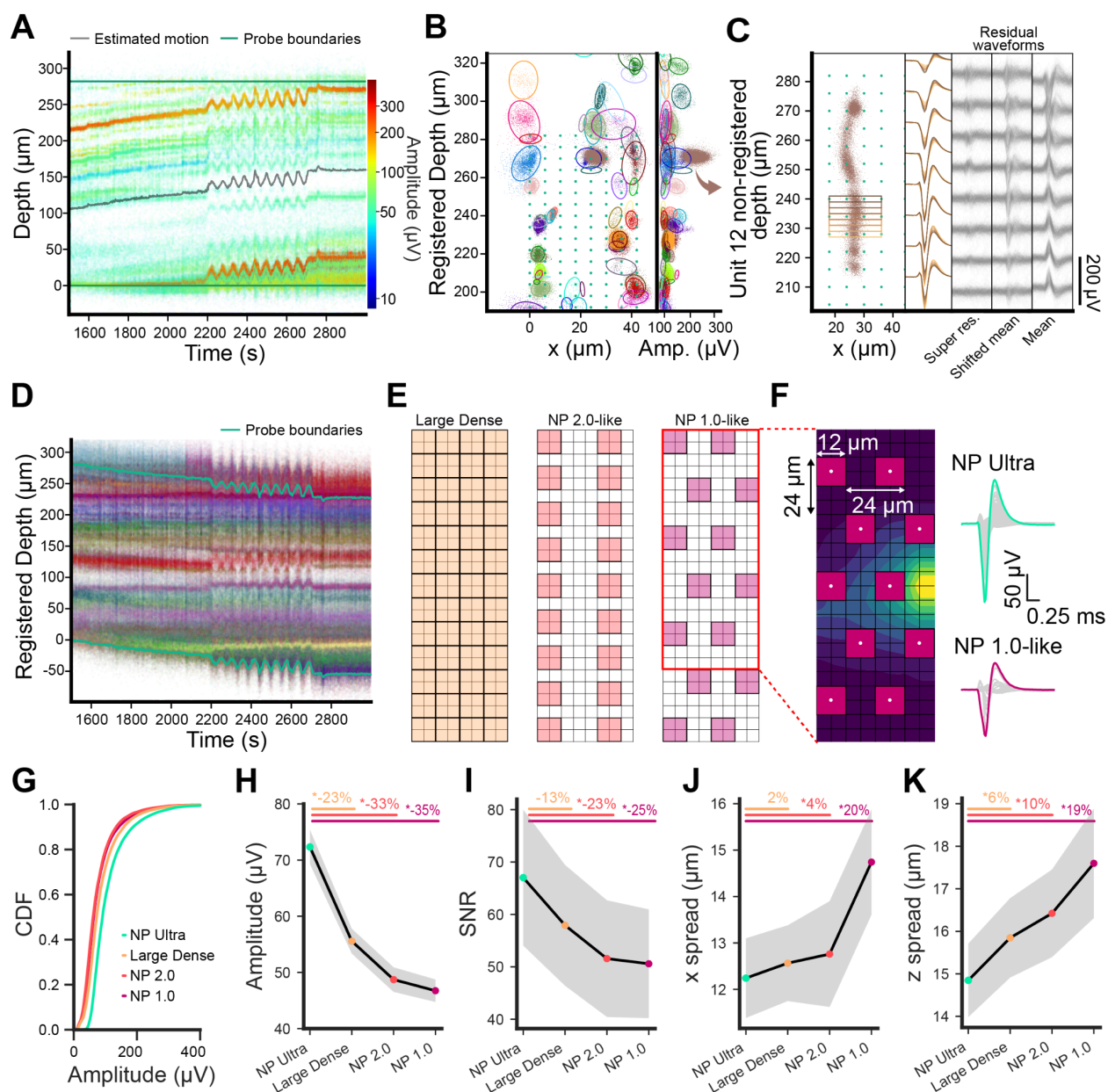


Figure 2. Localizing, clustering, and tracking spikes from Neuropixels Ultra probes. **A**, “Drift map” spatial spike raster of detected neurons in an example recording. Each point represents one spike; y-coordinate represents (unregistered) depth along the probe at which the spike was detected; x-coordinate detection time; color is log-scaled spike amplitude. Superimposed black line indicates the

estimated motion of the probe relative to the brain. Green horizontal lines indicate top and bottom boundaries of the probe. **B**, After probe motion registration, detected spikes are localized in 2D across the face of the probe. Left panel, each point represents a spike's fitted source coordinate in registered 2D space, colored by cluster identity, with ellipses representing the spatial spread of each cluster. Green markers indicate the recording site locations. Right, scatter plot of depth versus spike amplitude; colors as in left. **C**, The super-resolution template waveform and fit quality for cluster 12. Left, the non-registered positions of cluster 12's spikes. The seven colored rectangles overlaid on top of the non-registered spike positions demarcate the spikes used for estimating each component of the super-resolution template. Middle, the seven differently colored super-resolved templates for cluster 12 overlaid. Note the subtle differences in overlaid waveform shapes within each of the 8 channels (highlighted in different shades of brown color), showing how these templates capture shape variability for this unit due to probe drift. Each template is colored according to the bin where it was estimated shown in the left panel. Right, residual waveforms obtained after subtracting the super-resolution templates shifted according to the estimated motion, the mean template when shifted, and the un-shifted mean template. **D**, Drift map as in **A**, post-motion registration. Spikes colored by their cluster identity. **E**, Three spatial resampling patterns (Large Dense, NP2.0-like, and NP 1.0-like, from left to right). Superimposed grids indicate the NP Ultra site locations (thin gray grid) and the channel size and positions for the resampling patterns (bold black grid and colored sites). **F**, Left, Grid indicating the NP Ultra recording channels (black) and the NP 1.0-like resampled sites (magenta squares). Colormap indicates amplitude as in **Fig. 1C**. The corresponding NP Ultra and resampled waveforms are shown on the right, with all channels overlaid on top of each other, and the main channel colored in green for the NP Ultra and magenta for NP 1.0-like resampled pattern. **G**, Cumulative distribution function of all spike amplitudes, comparing large-dense, 2.0-like, and 1.0-like site arrangements. **H**, Peak site amplitudes of neurons in the original dataset recorded with NP Ultra versus the spatially resampled versions. Black line and the grey interval show the mean and standard error of the mean for each pattern. * indicates the statistical significance of the difference, obtained using a two-sample z-test on the unit amplitude distribution. **I**, Template SNR measured for each recording site pattern, from the original dataset and spatially resampled versions. **J**, Spatial spread, computed as the standard deviation of coordinates, in the horizontal dimension. **K**, Spatial spread in the depth dimension across the face of the probe.

Spikes recorded with NP Ultra probes have higher peak waveform amplitudes, higher signal-to-noise ratios, and more precise spatial estimates of spike locations than those recorded with previous Neuropixels probes, due to the increased spatial sampling. To measure this, we spatially resampled our NP Ultra recordings to simulate the signals that would have been recorded from the exact same position under different recording site patterns: an NP 1.0-like pattern with staggered recording sites and large spacing; an NP 2.0-like pattern with aligned recording sites and tighter spacing; and a hypothetical 'Large Dense' pattern with large 12x12 μm recording sites as in NP 1.0 and 2.0, but with dense, gap-less spacing as in NP Ultra (**Fig. 2E,F**). In each case, a single 12x12 μm simulated recording site is obtained by averaging across four 5x5 μm recording sites of NP Ultra, recapitulating the way that recording sites average signals over the extracellular space that they span (Robinson, 1968; Nelson and Pouget, 2010; Haines et al., 2024). Note that the Large Dense pattern does not reflect the geometry of any real device, but serves as a comparison to isolate the effects of gap-less coverage from the effects of the smaller site size itself. This computational resampling strategy produces datasets exactly matched in terms of the spiking events present, differing only in the sampling density, thus permitting us to directly compare the properties of the signal as observed with NP Ultra to what would have been observed with other, less-dense probes. This strategy is therefore a more

precise and tightly-controlled comparison than would be obtained by comparing true recordings from each probe type, which can vary substantially in yield from session to session (International Brain Laboratory et al., 2023a). Importantly, our goal with these analyses is to compare the differences obtained when recording neurons over a small (300 μ m) span of brain tissue, consistent with the full-density configuration of the NP Ultra probe. NP 1.0 and 2.0 probes have significant advantages in spatial coverage (spanning $\sim$ 4 and $\sim$ 3 mm, **Fig. 1B**), which may outweigh the higher yield of NP Ultra over small spans, depending on the targeted neurons and scientific objectives of a given experiment.

Peak spike amplitudes were highest in the original NP Ultra recording (**Fig. 2G,H**; Ultra vs 1.0, 35% median relative difference, $p = 1.5e-63$; Ultra vs 2.0, 33% difference, $p=5.5e-50$; Ultra vs Large Dense, 23% difference, $p=1.7e-29$; two-sample z-test on the distribution of amplitudes in each dataset). Reductions in peak site amplitude observed on the large-dense pattern reflect the effect of larger recording sites relative to NP Ultra, while reductions in amplitude under the 1.0-like and 2.0-like site patterns reflect the effect of larger sites in addition to a failure to sample the peak location due to gaps between sites. Large relative to small recording sites therefore account for about a 23% reduction in measured spike amplitude, and greater site spacing results in a further 11% reduction on average. The effect of site size and site density is a function of the spatial footprint of the signal; that is, a neuron whose extracellular action potential is visible over a large region will be relatively unaffected by sampling with a low site density. Therefore, the impact of sampling density on amplitude depends on the local population of neurons and varies across brain regions. We have characterized this for several regions (**Supp. Fig. S8, S9 and S10**). Together, these results demonstrate that NP Ultra records larger amplitude signals from individual neurons.

The higher amplitude of spikes measured with NP Ultra (**Fig. 2G,H**) overcomes the higher noise of its small sites (**Fig. 1G**), resulting in higher SNR when considering the combined signal recorded across channels. We computed a “Template SNR” for each neuron by first convolving the identified template waveform with the spikes identified as belonging to the neuron, thus capturing the amplitude of the spike across multiple channels. Then, we computed the SNR between this distribution and a noise distribution obtained by convolving the template waveform with samples from random other time points (see Methods; **Fig. 2I**). Neurons recorded with NP Ultra had the greatest template SNR, followed by the Large Dense, 2.0-like, and 1.0-like site patterns in order (Ultra vs 1.0, 25% difference, $p = 0.007$; Ultra vs 2.0, 23% difference, $p = 0.03$; Ultra vs Large Dense, 13% difference, $p = 0.16$). As with amplitude, the magnitude of the effect varies across brain regions, depending on the distribution of waveform footprint (**Supp. Fig. S9**). Using a spatial resampling strategy with a range of site size and spacing, our data suggest that an intermediate site size with minimal gaps between sites may comprise an optimal site layout for most neurons in most mouse brain regions (**Supp. Fig. S10**). Thus, despite higher noise levels at individual channels, NP Ultra ultimately yields higher SNR for identifying spikes from recorded neurons.

The increased recording site density of NP Ultra also produced substantially greater precision in the estimated spatial position of each spike relative to previous probes (**Fig. 2J-K**; **Supp. Fig.**

S8, S9). The success of the spike sorting strategy described above depends on the ability to spatially segregate spikes from individual neurons relative to nearby neurons. To understand how much this strategy will gain from the high sampling density of NP Ultra, we compared the spatial scatter in estimated spike positions from sorted neurons in the original data to the scatter obtained when estimating the spatial positions of the same spikes in the spatially resampled data (**Fig. 2J-K; Supp. Fig. S11**). Spikes were estimated with substantially higher precision in the NP Ultra dataset (**Fig. 2J**; Along horizontal dimension: Ultra vs 1.0, 20% difference, $p = 3e-8$; Ultra vs 2.0, 4% difference, $p = 0.005$; Ultra vs large-dense, 2% difference, $p = 0.99$. **Fig. 2K**; Along vertical dimension: Ultra vs 1.0, 19% difference, $p = 5e-13$; Ultra vs 2.0, 10% difference, $p = 5e-6$; Ultra vs large-dense, 6% difference, $p = 0.01$.), confirming that these probes are well-suited to employ this spike localization-based sorting strategy.

The improved SNR and spatial precision translated to improved yield of sortable neurons. We evaluated yield by focusing on neurons that were present throughout the recording and had a biologically meaningful response, here defined to be reliable and selective responses to visual stimuli. We collected new datasets from the primary visual cortex in which awake, head-fixed mice were presented with a battery of 118 natural images. Neurons were included in the final yield calculation if they had a significant visual “fingerprint”, defined here as visual responses that are reliable (i.e., reproducible across repetitions of the same image) and selective (i.e. differ in a consistent manner across different images) (see Methods; **Fig. 3A-B, Supp. Fig. S12**). We ran the DARTsort algorithm on these datasets and on the spatially resampled versions of the datasets and found that the yield of neurons with reliable and selective visual responses was highest for NP Ultra, which yielded over 10 extra neurons on average than the 1.0-like, 2.0-like, and large-dense patterns (**Fig. 3C-D**, mean $\pm$ standard error of the mean (SEM) difference in neurons: Ultra vs large-dense, 10.1 ± 3.7 , $p = 0.016$; Ultra vs 2.0, 14.4 ± 3.3 , $p = 0.002$; Ultra vs 1.0, 14.4 ± 3.4 , $p = 0.002$). The large difference in yield between Ultra and 1.0 amounted to a greater than 2-fold improvement (2.2x greater yield of NP Ultra relative to 1.0-like on average). This yield difference did not arise from over-splitting of neurons in NP Ultra relative to other patterns (**Supp. Fig. S13**). NP Ultra recorded from more neurons partly because it captured more neurons with low amplitude than the other probe geometries (**Fig. 3E**, median amplitude, bootstrapped 95% CI in μV : NP Ultra = 44.6, 42.1 - 56.0; large-dense = 56.6, 48.5 - 66.0; 2.0-like = 68.8, 54.7 - 79.1; 1.0-like = 65.1, 53.3 - 72.6). When further filtering visually reliable neurons for those with a median amplitude greater than 50 μV and passing a quality metric based on refractory period contamination (International Brain Laboratory et al., 2023), the yield differences between NP Ultra and resampled patterns are relatively smaller, though NP Ultra still improved the yield of these well-isolated single neurons relative to NP 1.0 (1.7x greater yield of NP Ultra relative to 1.0-like on average; mean $\pm$ SEM difference in neurons = 2.8 ± 1.1 ; $p = 0.039$; **Fig. 3C, D**). In sum, the higher site density of NP Ultra greatly increases the yield of neurons with biologically meaningful responses compared to lower density probes over the spatial range sampled by NP Ultra.

The increased yield of visually responsive neurons from NP Ultra translated into a higher performance decoding visual stimuli than the other site patterns. Linear discriminant analysis with dimensionality reduction was used to decode each possible pairing of the battery of 118

natural images from all neurons with significant visual responses in each pattern (as identified by the fingerprint reliability metric described above). Decoding accuracy was significantly higher for NP Ultra than large-dense, 2.0-like, and 1.0-like patterns (**Fig. 3F**, 1.8%, 3.8%, 3.0% increase in median accuracy across sessions, $p = 0.016, 0.008, 0.008$ signed-rank test). The increase in decoding accuracy for NP Ultra resulted from the increased number of neurons rather than any difference in properties of recovered neurons, as decoding accuracies were equal when the number of included neurons was matched between patterns (**Fig. 3G**). This test case demonstrates that the ~2x extra yield of NP Ultra enables the decoding to capture more subtle details of the population response pattern than captured with earlier NP versions, improving any analysis for which doubled yield is beneficial.

The higher site density of NP Ultra also afforded a moderately improved ability to stably track drifting neuron locations across time. To measure this, we imposed slow probe motion as a ground truth pattern of unstable probe location relative to the brain during the visual fingerprint sessions, following a method we developed in past work (Steinmetz et al., 2021). We presented each image an average of 30 times before moving the probe, and 15 times afterwards. To assess tracking stability, we computed a stability ratio for each neuron exhibiting reliable visual responses by dividing the mean pre- vs. post-motion fingerprint correlation by the mean pre- vs. pre-motion correlation. These mean correlations are calculated by randomly splitting the 30 pre-motion repetitions of all images into two non-overlapping 15-repetition subsets 1000 times, and correlating the fingerprints obtained from one pre-motion trial subset and from the other pre-motion subset (pre- vs. pre-motion) or from the post-motion repetitions (pre- vs. post-motion) (**Supp. Fig. S12B-C**). A stability ratio of 1 indicates a stably tracked neuron: the visual responses before and after motion are just as similar as those from two subsets of the trials before motion. Pooling the stability ratios of neurons within patterns across sessions, we found that there was a significant difference in the ratios between NP Ultra vs. 1.0-like and 2.0-like patterns, but not between NP Ultra and the large-dense pattern (% difference between Ultra and large-dense, 2.0-like, and 1.0-like = 11.4%, 18.3%, 15.1%; $p = 0.062, 0.0030, 0.0064$ Mann-Whitney U test). (**Fig. 3H**). These findings are corroborated by computational simulations of extracellular action potentials (EAP) sampled at different spatial densities, which reveal that significant information for reconstructing the EAP begins to be lost around the NP 1.0 spatial frequency (**Supp. Fig. S7**). Thus, the extra spatial resolution available from NP Ultra provides additional improvements in correction of slow probe instabilities compared to both Neuropixels 2.0 and 1.0 probes.

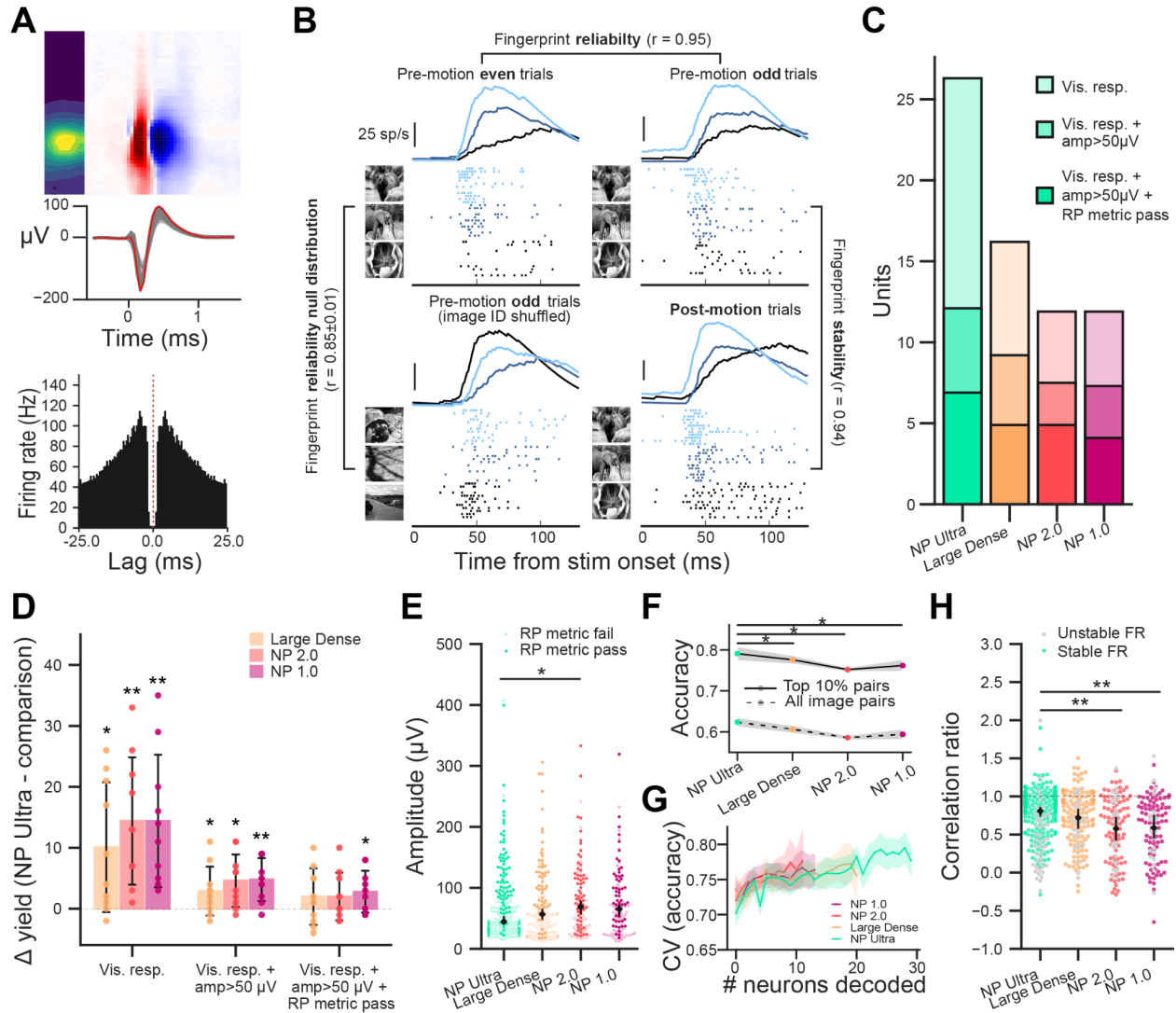


Figure 3. Tracking neurons with NP Ultra using visually-evoked fingerprint responses. **A**, Top: Spatial (left), spatio-temporal (right) and temporal (bottom) waveforms recorded by NP Ultra of an exemplar neuron shown in B. Bottom: Autocorrelogram of exemplar neuron shown in B. **B**, Peri-stimulus spike rasters and average rate (smoothed with causal half-gaussian filter with standard deviation = 45 ms) for an example neuron with reliable and stable visual response, aligned to the onsets of 3 of the 118 visual stimuli (indicated in 3 different bluish hues). Top two plots show odd and even trials from the pre-motion period, revealing a close match. Bottom left plot shows shuffled image ID control, with a poor match. Bottom right shows the post-motion responses, which again match the pre-stimulus un-shuffled responses. **C**, Mean yield of neurons across sessions, each pattern type progressively filtered by the following three metrics: reliable visual response ("vis. resp."), median spike amplitude > 50 μ V ("amp>50 μ V"), and 70% confidence of $\leq 10\%$ contamination in the refractory period contamination metric ("RP metric pass"). **D**, Differences in neuron yield between NP Ultra and each spatially resampled pattern calculated per session, using the same quality metrics as in E. Bars and error bars are mean $\pm$ SEM. Stars indicate significant p -values from the Wilcoxon signed-rank test. **E**, Amplitudes of stable neurons with reliable visual responses across all sessions. Black diamonds and error bars are median and 95% confidence intervals. Stars indicate significant p -values from Mann-Whitney U test. **F**, Median of decoding accuracies across $n=10$ sessions from NP Ultra and spatially resampled datasets. Each session and

probe has one value for mean decoding accuracy across all image pairs or top 10% of most decodable pairs. Stars indicate significant p -values from the Wilcoxon signed-rank test between decoding accuracies (amongst all image pairs) of each probe type within a session. **G**, Median decoding accuracies for top 10% of most decodable pairs as a function of number of neurons included in the decoder, split by pattern type. Each plotted point has ≥ 3 sessions with the requisite number of visually responsive neurons. **H**, Stability ratios (ratio of correlations between mean pre- vs. post-motion and mean pre- vs. pre-motion visual fingerprints; see **Supp. Fig. S12B,C**) for neurons with reliable visual responses across all sessions. Colored points indicate neurons with a stable firing rate from the pre- to post-motion period; gray dots indicate neurons with unstable firing rates (see Methods: Yield and stability analysis). Black diamonds and error bars are median and 95% confidence interval only from neurons with stable firing rates. Stars indicate significant p -values from Mann-Whitney U test; only stable firing rate neurons included in the test. For all panels, error bars for medians are calculated using $n=1000$ bootstrapped 95% confidence interval of the median, error bars for mean are SEM, and * indicates $p<0.05$; ** indicates $p<0.01$, *** indicates $p<0.001$.

Subcellular recordings from axons and dendrites in mouse isocortex

NP Ultra probes detect extracellular action potentials from neuronal structures typically under-sampled by other probes. Neurons or subcellular structures with waveforms that decay over distances smaller than 20 μm would not be detected by NP 1.0 probes if an action potential was initiated between electrode contacts. Indeed, in recordings from the visual cortex of awake mice, we noted a substantial number of extracellular action potentials that had large amplitude but were present only on a few recording channels of NP Ultra (**Fig 4A**). To quantify this observation, for each waveform we calculated its “spatial footprint” as the radius from the peak channel at which the average spike amplitude fell below 30 μV , a value sufficiently above the noise level (**Supp. Fig. S14A-C**). We observed a bimodal distribution of this quantity, with a substantial fraction of footprints less than 20 μm , which we define as having a “small footprint” (**Fig. 4B**).

Recordings of axons had small footprint waveforms, providing a means to identify them. To test this, we applied the GABA agonist muscimol to the cortex to suppress somatic spiking while leaving the activity of non-local axons unaffected. All recorded waveforms that survived muscimol application (firing rate > 1 spike/s in muscimol; 23/175, $n = 6$ sessions, $n = 3$ mice) had a small footprint (**Fig. 4C-G**). Not all small footprint units survived muscimol application, which is consistent with recordings from axons of local origin. Aside from their small spatial footprint, the muscimol-resistant axonal waveforms were inconsistent in appearance: some had narrow, negative peaks with little or no early positive component while others had prominent early positive peaks (**Fig. 4G, Supp. Fig. S15**). Similarly, spike waveforms histologically localized to the corpus callosum could have either or both of these waveform characteristics (**Supp. Fig. S16A-C**). Biophysical simulations confirmed that small-footprint, narrow, negative spikes could be observed at nodes of Ranvier (**Supp. Fig. S16D**; (Haines et al., 2024)), that positive peaks may arise in unmyelinated segments of otherwise myelinated axons (**Supp. Fig. S16E**), and that unmyelinated axons are unlikely to be detected given their small amplitude except when they have substantial branching and termination (**Supp. Fig. S16E**; (McColgan et

al., 2017)). Therefore, the characteristics of axonal spikes vary dramatically depending on the exact properties of the axon recorded, but are unified by their small spatial footprint rather than any particular waveform-peak direction, number, or width.

NP Ultra also records from dendritic compartments of neurons with higher resolution and SNR than previous technologies. Extracellular recordings of back-propagating action potentials in apical dendrites of cortical pyramidal neurons have been reported with NP 1.0 (Jia et al., 2019) and with other probes (Buzsáki and Kandel, 1998). We made recordings with NP Ultra and NP 1.0 using an insertion strategy to obtain recordings in mouse V1 approximately aligned with apical dendrites. First, we opto-tagged layer 5 pyramidal neurons (**Fig. 4H-I**, Sim1-Cre;Ai32), and observed propagation in waveforms consistent with known back-propagation waveform characteristics (i.e. initial positive peak at the time of the somatic spike; (Gold et al., 2006)) and speed (Stuart et al., 1997; **Fig. 4J**). Then, to record dendritic signals across a larger extent of the apical tree (**Fig. 4L**), we switched to the linear 192 x 2 configuration (**Fig. 4K**). With this configuration, we recorded back-propagation dynamics spanning more than 400 μm from the putative somatic channel, at $\sim 7\times$ higher spatial resolution than with NP 1.0 (6 μm linear spacing versus 40 μm ; **Fig. 4L**). This enhanced resolution led to significantly greater SNR in resolving back-propagating signals across the apical dendrite (**Fig. 4M**). These improvements over NP 1.0 recordings did not depend on probe alignment or recording quality, as they persisted when comparing dendritic signals of similar amplitudes (**Fig. 4N**). The variance in amplitude of back-propagating action potentials exceeded the variance expected from background noise, suggesting that recordings of back-propagation with NP Ultra detect meaningful variations in amplitude of individual events not explained by noise, likely attributable to sensory or state-dependent factors (**Supp. Fig. S14D-F**). In addition to back-propagation, our dataset revealed a number of distinctive waveforms with spatiotemporal characteristics not previously reported, which appear similar to biophysical simulations of close dendritic apposition (**Supp. Fig. S17**). Taken together, NP Ultra enables subcellular recordings from dendrites with improved resolution and SNR relative to prior devices.

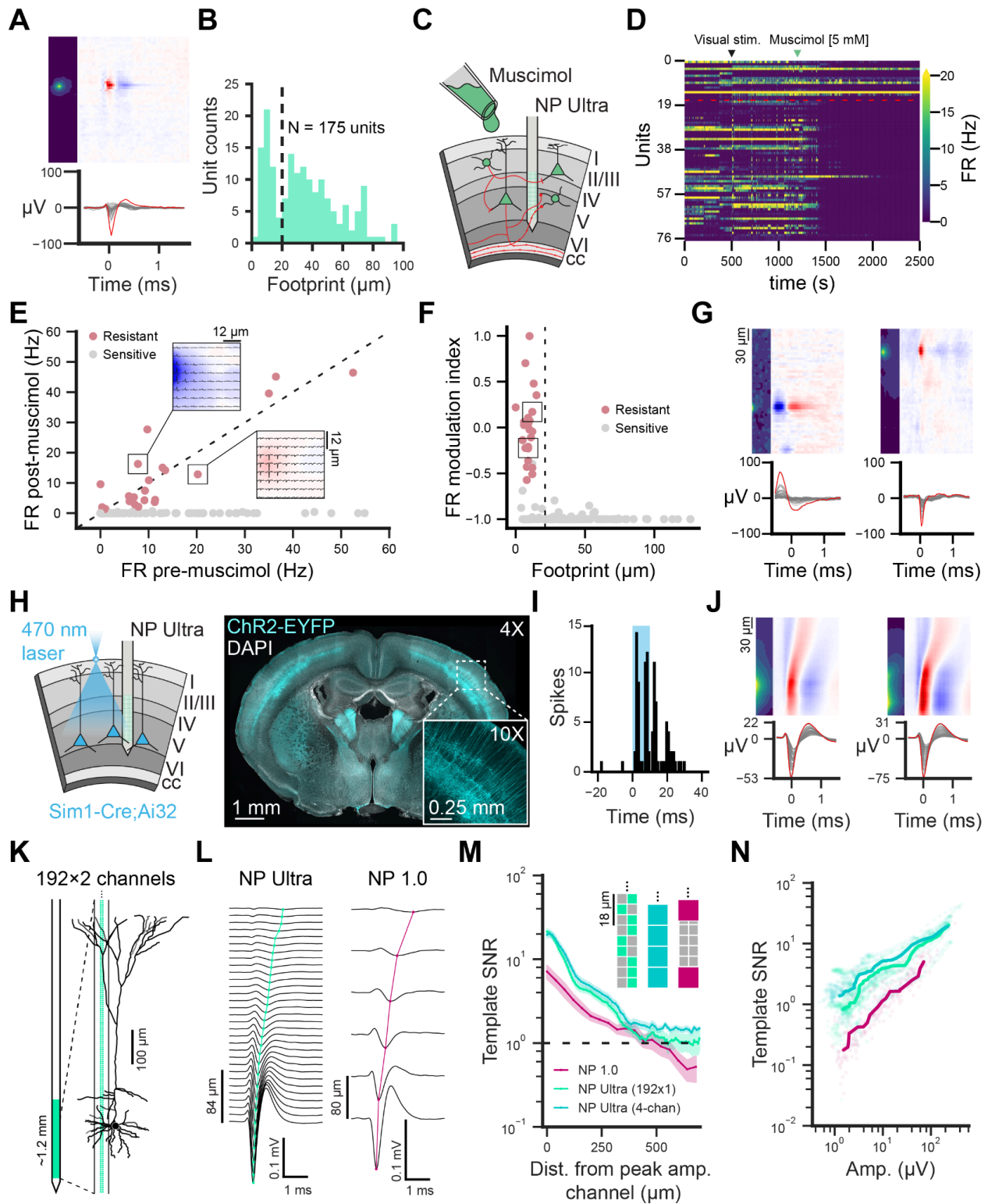


Figure 4. Recordings from subcellular compartments. **A**, Example small-footprint waveform recorded from primary visual cortex. **B**, Histogram of measured spatial footprints across a total of 175 units recorded in mouse visual cortex with muscimol application (mice = 3; sessions = 6). Dashed line denotes the boundary (20 μm) between what we define as large and small footprints. **C**, Schematic of NP Ultra

recordings in cortex with surface application of muscimol [$5\ \mu\text{M}$]. Red lines indicate axonal segments of intra- and extra-cortical neurons. **D**, Raster plot showing the firing rate of 77 units during an exemplar NP Ultra recording. Black arrow = time of visual stimulation, green arrow = time of muscimol application. Units are sorted by ascending footprint size, the boundary between large and small footprints is denoted by the red dashed line. **E**, Scatter plot of firing rate across all units measured within a 100 ms window pre- and post-muscimol application, colored by where a given unit was resistant or sensitive to muscimol. Dashed line indicates identity between pre- and post-FR. Insets display the waveforms and heat maps of two example muscimol-resistant units. **F**, Relationship between the spatial footprint and firing rate modulation index for all recorded units. Units from **E** marked by black boxes. **G**, Full waveform plots of the two units in **E**, respectively. **H** (left), Schematic of optotagging experiment in *Sim1-Cre;Ai32+* L5b pyramidal neurons. (Right) histology from a *Sim1-Cre;Ai32* animal showing labeled L5 pyramidal neurons in S1. **I**, PSTH of an example *Sim1+* unit in response to 10 ms 470 nm photostimulation. **J**, Two example *Sim1+* units with visible potentials propagating along the NP Ultra probe. **K**, Schematic of a L5 cell recorded with a switchable NP Ultra probe in the linearized (192×2 channels) configuration. **L**, Example bAPs recorded with NP Ultra (left) and NP 1.0 (right). Traces show spike-triggered average waveforms ($n = 2,000$ spikes) recorded from the column of channels that includes the peak amplitude channel. Colored traces track the minimum voltage recorded on each channel with a recorded amplitude above 5% of the maximum. **M**, Dendritic channel template SNR measured as a function of vertical distance from the peak amplitude channel for NP Ultra (192×2 configuration, best column selected, green), and NP Ultra (192×2 configuration, 4-channel average, cerulean) or NP 1.0 probes (magenta). Channel configuration diagrams are shown on the top right. **N**, Template SNR plotted against channel waveform amplitude for each channel in every probe configuration from **M**.

Small footprint extracellular action potentials across mouse brain regions

To assess the prevalence of small footprint, axon-like extracellular action potentials across the brain, we used NP Ultra probes to make recordings from 4,666 single units in 18 brain regions (>50 single units/region; $n = 4$ mice; $n = 12$ sessions; **Fig. 5A**). In each experiment, we inserted probes and recorded for a short time (~ 5 minutes) at a given recording site before advancing the probe $250\ \mu\text{m}$, then repeating this process several times over many recording locations. Targeted areas included several regions of the isocortex, hippocampus, basal ganglia, thalamus, midbrain, and corpus callosum. We sorted spikes from each recording position separately (see Methods) and compiled the waveforms of all single units, which exhibited a range of diverse spatial features (**Supp. Fig. S18, S19**).

NP Ultra recorded many small footprint units in each brain region, reflecting otherwise unobservable or under-sampled units. As in NP 1.0 recordings (Jia et al., 2019), we observed a wide distribution of single channel spike waveform features, such as amplitude, peak-to-trough ratio (PTR), and spike duration, across brain regions (**Supp. Fig. S19, S20**). In each brain region we recorded, a significant fraction of extracellular action potentials ($>10\%$ in most regions) had a spatial footprint of less than $20\ \mu\text{m}$ (**Fig. 5B,C**). This fraction of small footprint units was significantly greater in recordings localized to the corpus callosum (57.5% , $31.4 \pm 26.1\ \mu\text{m}$, mean $\pm$ SD, 113 units) and the dentate gyrus (75.3% , $20.0 \pm 16.2\ \mu\text{m}$, 77 units), compared to primary visual cortex (VISp), for example (10.0% , $50.0 \pm 25.6\ \mu\text{m}$, 359 units; **Supp. Fig. S21**,

S22). These small-footprint units may reflect recordings from axons or from neurons with exceptionally small somata, such as granule cells in DG, or both.

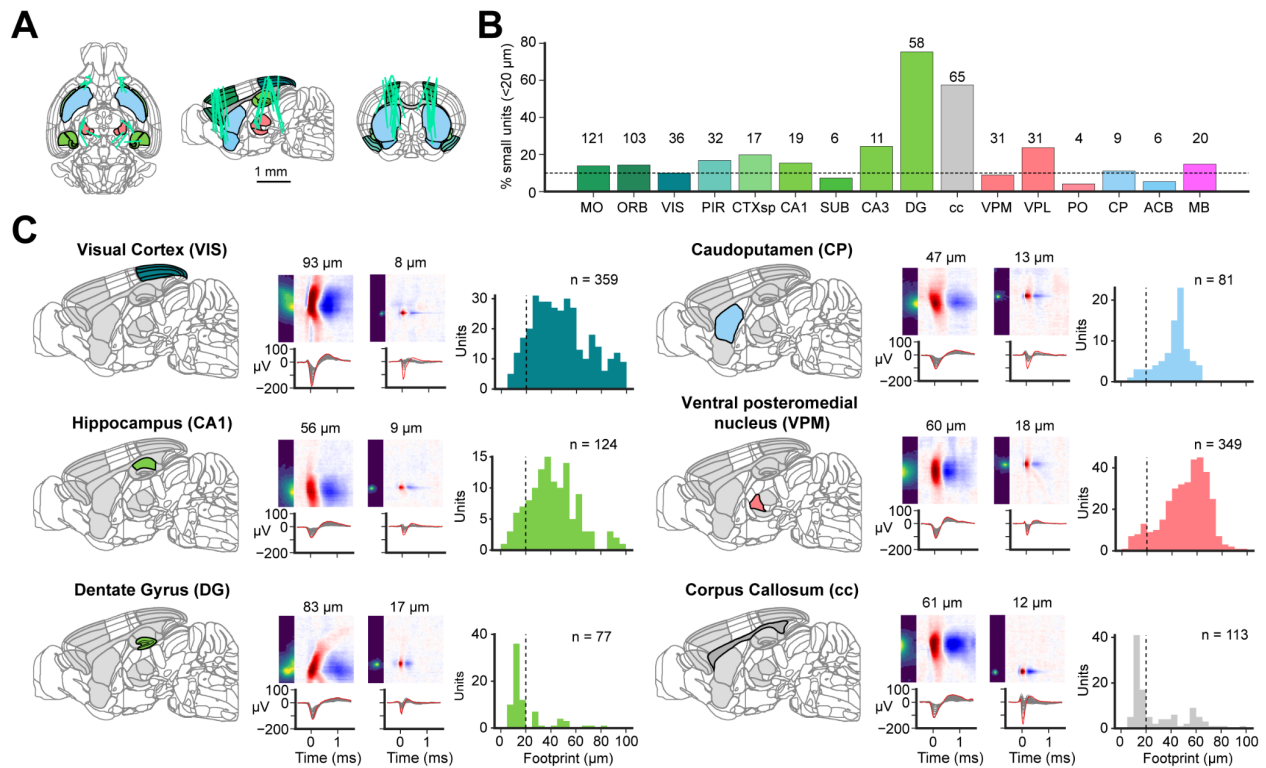


Figure 5. NP Ultra recordings reveal small electrical footprint units in many brain regions of the mouse. **A**, NP Ultra insertion trajectories from different mice and sessions, registered to the CCF atlas coordinates. **B**, Percentage of units with small footprints less than 20 μm in regions with the most recorded single units. Number of small units per region indicated. Dashed line corresponds to 10%. **C**, Two example waveforms with one large and one small spatial footprint from each of six different brain regions, and spatial footprint histogram of all units from these regions. See <https://npultra.steinmetzlab.net/> for a browsable interface to explore the neurons in this dataset. MO: Somatomotor areas; ORB: orbital area; VIS: visual areas; PIR: Piriform area; CTXsp: cortical subplate; CA1: hippocampal CA1; SUB: Subiculum; CA3: hippocampal CA3; DG: Dentate gyrus; cc: corpus callosum; VPM: ventral posteromedial nucleus of the thalamus; VPL: Ventral posterolateral nucleus of the thalamus; PO: Posterior complex of the thalamus; CP: Caudoputamen; ACB: Nucleus accumbens; MB: midbrain.

Small footprint extracellular action potentials across species

Species with larger brains tend to have larger neurons in most regions (Herculano-Houzel et al., 2014; Beaulieu-Laroche et al., 2021). Different species have specialized brain regions with varying neuronal densities, tailored to behavioral demands through evolution. For instance, whereas the cerebral cortex constitutes the majority of brain volume in primates, weakly electric mormyrid fish possess a greatly enlarged cerebellum that covers their entire dorsal brain surface (Sukhum et al., 2018). To understand the generality of observations about waveform spatial footprints in mice, we therefore collected high resolution waveforms using NP Ultra from

three additional species: monkey (*Macaca nemestrina*) visual cortex, electric mormyrid fish (*Gnathonemus petersii*) cerebellum, and bearded dragon (*Pogona vitticeps*) medial cortex (an area homologous to mammalian dentate gyrus (Tosches et al., 2018)).

We observed a substantial number of units with a footprint less than 20 μm in each species (**Fig. 6A,B**). We collected 124 units from the monkey visual cortex, and found 13 units with footprints smaller than 20 μm (10.5%). This closely matches the observation in the mouse visual cortex (36/359, 10.0%). The similar distribution of footprints between mouse and monkey visual cortex is consistent with the similar distribution of somatic sizes in this area across these two species, despite differences in soma size in other cortical regions (Gilman et al., 2017). A similar proportion of units in the lizard medial cortex had small footprints (2/19, 10.5%). Close to half of the units detected in the cerebellum of the electric fish had footprints below 20 μm (18/37, 48.7%). Together these observations demonstrate that units with small spatial footprints (<20 μm), which are difficult to detect with lower density probes, are consistently detected with NP Ultra across species.

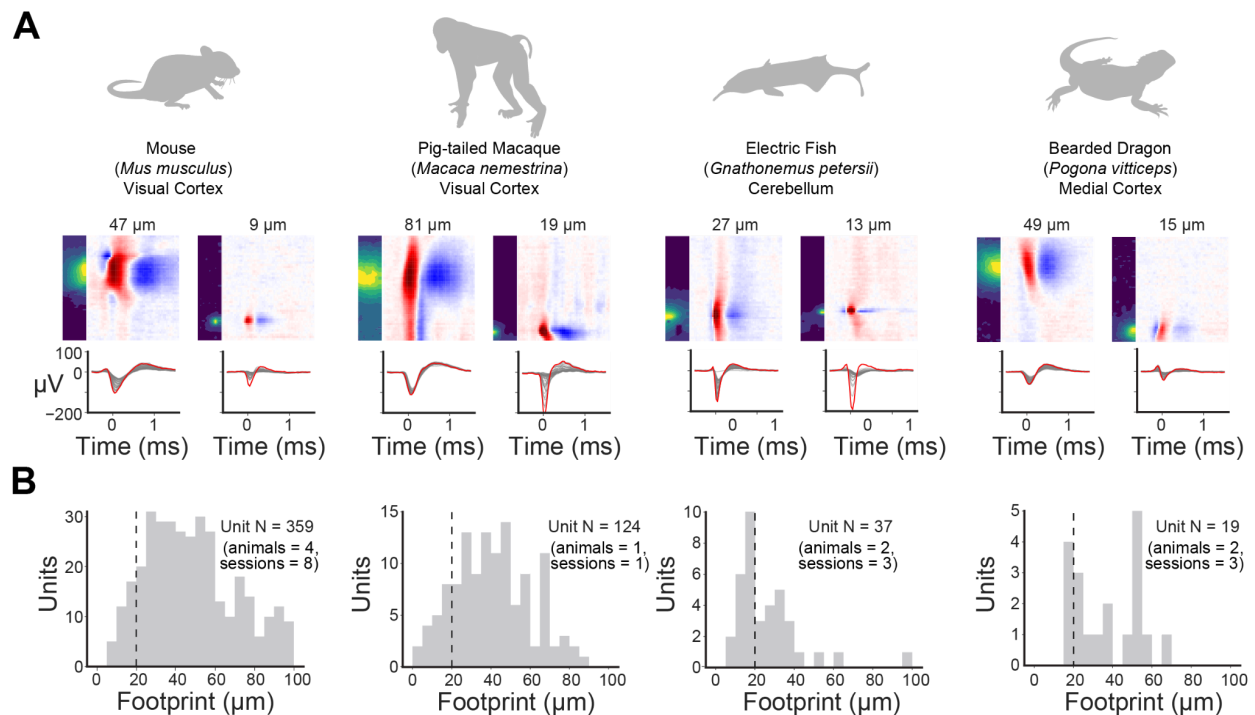


Figure 6. Small footprint units are detected across different species. **A**, Two example waveforms from four species (mouse visual cortex, monkey visual cortex, electric fish cerebellum, and bearded dragon medial cortex), with one large and one small spatial footprint. Two waveforms with similar peak amplitude in the same species were selected for comparison. **B**, Spatial footprint histogram of all units from each species.

Spatial footprints of genetically identified cell types

In addition to dependence on subcellular morphology and brain region as described above, the spatial footprint of recorded neurons differed between genetic cell types. To measure and compare the spike waveforms of genetically identified cells, we performed optotagging of three inhibitory neuron types in the visual cortex. We used three transgenic mouse lines to express, in separate subjects, channelrhodopsin (ChR2) in each of three major inhibitory neuron subclasses: parvalbumin- (PV), somatostatin- (SST), and vasointestinal polypeptide-expressing (VIP) cells. During the recording session, we photostimulated the cortex causing these neurons to spike and thereby 'optotagging' them (**Fig. 7A**; (Lima et al., 2009)). Optotagged units had short latency and stereotyped spiking responses to photostimulation (**Fig. 7B**). 243 PV, 116 SST, and 126 VIP interneurons were identified as optotagged based on both unsupervised density-based clustering methods and qualitative assessment (243/3344 = 7.3% of units tagged in PV-targeting mice; 116/2641 = 4.4% for SST; 126/3492 = 3.6% for VIP; **Fig. 7B-D**, **Supp. Fig. S24**; Methods). Opto-tagged neurons had diverse spatiotemporal waveforms (**Fig. 7C**; **Supp. Fig. S25**; **Supp. Movie S1**). We computed the spatial footprint of each optotagged unit and found that each of these three inhibitory types consisted of predominantly large footprint waveforms (**Fig. 7D**; >20 μm ; 91% SST, 92% VIP, 94% PV). Footprints significantly differed across all three cell types (**Fig. 7D**; $p < 0.001$, Kruskal-Wallis; median footprint = 40.0 μm for PV, 35.0 μm for SST, 42.0 μm for VIP; $p = 0.01$, PV vs. SST, ranksum; $p = 0.01$, PV vs. VIP; $p < 0.001$, SST vs. VIP). Thus, spatial footprints depended on genetic cell type, and all examined types had predominantly large footprint waveforms.

PV+ neurons correspond to the subset of narrow-spiking waveforms with large footprint. Narrow spike duration is a standard feature used to identify putative PV+ interneurons in past literature (Barthó et al., 2004; Niell and Stryker, 2008; Kvitsiani et al., 2013), since only a minority of SST (Kvitsiani et al., 2013), VIP (Petersen et al., 2021), and excitatory neurons (Nowak et al., 2003; Lemon et al., 2021) have narrow spike waveforms. We separated all untagged units into narrow-spiking (NS, duration < 0.4 ms; $n = 2,058$) and regular-spiking (RS, duration ≥ 0.4 ms; $n = 6,838$) based on spike duration and examined footprint distribution within these two groups (**Fig. 7E, F**). Surprisingly, we found that NS waveforms displayed a distinctly bimodal footprint distribution compared to RS. Based on this observation, we subdivided the untagged NS group into two categories based on their footprint (greater or less than 20 μm), which we call large narrow-spiking (NS_L , $n = 952$) and small narrow-spiking (NS_S , $n = 1,133$) units. PV+ optotagged neurons had an overlapping distribution of footprint with NS_L but not with NS_S (**Fig. 7G**). In addition, PV+ neurons and NS_L units were not well-discriminable with multiple waveform features (pre-trough peak-to-trough ratio and peak-to-trough ratio; see Methods) using Linear Discriminant Analysis (0.56 ± 0.042 , mean $\pm$ SEM), while PV+ neurons and NS_S units were highly discriminable (PV: NS_S , 0.87 ± 0.022 ; **Fig. 7H**).

NS_S provided either excitatory or inhibitory monosynaptic input to putative downstream neurons, further supporting the interpretation that, despite their narrow spike waveform, they arise from axonal recordings rather than PV+ interneurons. To investigate this, we performed an analysis of cross-correlograms to identify putatively monosynaptically connected pairs. Such pairs were

extremely rare ($<0.02\%$ of all observed interactions, $>125,000$ pair interactions; **Fig. 7I, Supp. Fig. S26**). We found instances of both putative excitatory and inhibitory interactions at short latencies ($<3\text{ms}$) from the time of an NS_s unit. Thus, NS units are not a uniform population and NS_s units, which comprised about half of the NS units recorded with NP Ultra in mouse visual cortex, mostly do not correspond to PV+ interneurons. In recordings with NP1.0 probes, small footprint units were detected at a lower rate, but NS_s units still comprised about one third of all NS units overall (**Supp. Fig. S26C**).

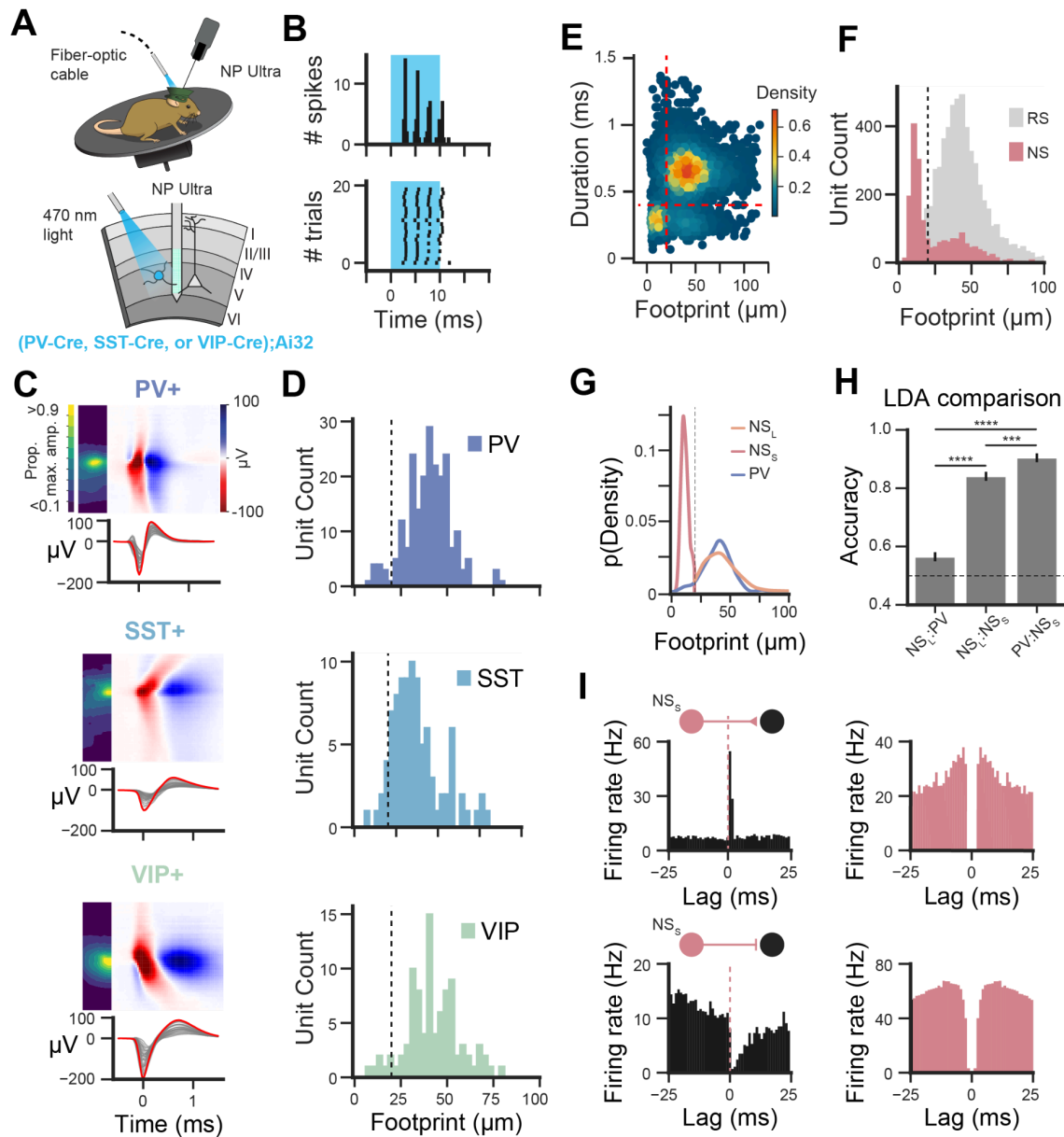


Figure 7. Optotagging three inhibitory neuron classes measured with NP Ultra. **A**, Schematic of experimental setup in which neurons in visual cortex transgenically expressing channelrhodopsin were recorded with NP Ultra probes and optotagged using 470 nm light. **B**, Trial-wise optogenetically-driven

responses of an example PV-Cre;Ai32-expressing unit during a 10 ms 470 nm light pulse. **C**, Example spatial, spatio-temporal, and temporal waveforms from optotagged units in each of the examined transgenic mouse lines (PV-Cre;Ai32, SST-Cre;Ai32, VIP-Cre;Ai32). **D**, Distributions of spatial footprints for each optotagged group of units. **E**, Scatter plot of waveform duration and footprint size, colored by the density of units within a given region of the graph. Note the existence of two maxima, one with relatively long waveform durations and large footprint, and another with short-duration waveforms and small footprint. **F**, Distributions of footprints for all units, segregated by waveform duration (RS = regular-spiking, NS = narrow-spiking). NS units displayed a bimodal distribution split at 20 μ m. Black dashed line indicates the demarcation between “large” and “small” units. **G**, Density histograms displaying the distribution of footprint between PV and NS units. NS units are subcategorized into NS_L (narrow-spiking, large) and NS_S (narrow-spiking, small) based on a footprint size threshold (20 μ m). **H**, Accuracy of Linear Discriminant Analysis (LDA) in separating different types of NS and PV units using pre-peak-to-trough ratio (prePTR) and peak-to-trough ratio (PTR) as features. **I**, Left, Cross-correlograms displaying the firing rate of example units as a function of spiking in two NS_S units (top/bottom rows). Dashed line indicates the time of an NS_S unit spike. Right, Autocorrelograms of the example NS_S units.

Classification of cortical inhibitory neurons

The dense and small recording sites of NP Ultra acquire greater resolution information about the electrical features of recorded neurons, improving cell type classification from extracellular signatures. To test this, we took a decoding approach to assess classification performance in three situations: between opto-tagged interneurons and untagged units; amongst opto-tagged interneurons and untagged units separated by key features (RS, NS_L and NS_S units; **Fig. 7E, F**); and amongst the three types of opto-tagged interneurons only. In addition, we compared the classification performance of NP Ultra against NP 1.0-like resampled waveforms, and assessed the utility of footprint as a discriminative metric.

First, we trained a simple logistic regression model between single opto-tagged types (PV, SST, or VIP) against untagged units, using all recorded waveform features (Methods; see below) as predictors. This model performed well, especially compared to a model using waveform duration alone, both when using balanced sampling (89% correct for PV, 82% for SST, 83% for VIP; **Fig. 8A, B, Supp. Fig. S29A**) and sampling that maintained interneuron proportions approximately equal to those found in cortex (**Supp. Fig. S29B-D**). Importantly, these are underestimates of the true performance since opto-tagging can fail for multiple reasons (for example, insufficient light reaching a recorded neuron), which would result in neurons from the targeted class being categorized as untagged. These failure rates are not possible to estimate precisely. Nevertheless, high accuracy was obtained, so these datasets could be used to assess the relative advantages of NP Ultra for cell type classification and the features that contribute to this performance, including spatial footprint.

We extracted single-channel waveform features using the peak amplitude channel for each unit, which revealed waveform characteristics unique to each type of neuron discussed above (**Fig. 8C, D, Supp. Fig. S26, Table 4**). PV+ neurons displayed large amplitude, short-duration spike waveforms that were highly similar to the NS_L units (**Fig. 7H**), especially in terms of

peak-to-trough ratio (PTR) and repolarization slope. SST+ neurons and NS_S units tended to have the lowest amplitude waveforms on average but were notably different in prePTR and PTR (**Fig. 8C**; Methods), with the former having both smaller prePTR and larger PTR ($p_{\text{prePTR}} = 7.7 \times 10^{-6}$, $p_{\text{PTR}} = 5.43 \times 10^{-30}$, Mann-Whitney U test, multiple comparisons corrected via Benjamini-Hochberg method, **Supp. Fig. S27**). VIP neurons were generally heterogenous, forming a bimodal distribution along some metrics, especially spike duration. However, VIP units as a population closely resembled untagged RS units in many respects with the notable exception of baseline firing rate, which was higher in VIP cells, and the average inter-spike interval (avg. ISI), which was lower ($p_{\text{bFR}} = 4.6 \times 10^{-8}$, $p_{\text{ISI}} = 5.3 \times 10^{-22}$, **Supp. Fig. S27**). Reducing the dimensionality of the features revealed substantial, but incomplete, segregation of the cell classes according to these features (**Fig. 8E**, **Supp. Fig. S28**). We therefore undertook a quantitative assessment of cell type classification.

Interneuron classification was improved by the dense sampling of NP Ultra compared to NP 1.0. To analyze this, we trained a non-linear supervised classifier (random forest) to predict the identity of each neuron among RS, NS_L, NS_S, PV, SST, and VIP types, with balanced groups (chance performance = 0.17; see Methods). We assessed model performance when training on three different feature sets: single-channel waveform features (1-ch features; **Fig. 8C**), the single-channel waveform itself, or all 1-ch features along with the footprint radius (1-ch wf, 1-ch + footprint; **Fig. 8D, F**, **Supp. Fig. S28**). Because classifier accuracy was not improved using the single-channel waveform over single-channel features using all units, we focus here only on analyses using the single-channel features and footprint. Overall, classification performance on both of these feature sets was significantly above chance for these six classes (mean $\pm$ S.D. accuracy: 1-ch feats. = 0.63 ± 0.03 , 1-ch feats. + footprint = 0.67 ± 0.04). To compare NP Ultra waveforms with lower resolution NP 1.0 waveforms, we spatially resampled data from NP Ultra-recorded units to provide a direct unit-for-unit comparison (**Fig. 8F**, **Fig. 2E**, see Methods). NP Ultra-recorded units had better performance than NP 1.0, regardless of the features used in classification (mean accuracy for NP 1.0-like: 1-ch feats. = 0.44 ± 0.16 , 1-ch feats. + footprint = 0.46 ± 0.15 , $p \ll 0.001$ for each comparison, independent t-test, multiple comparisons corrected via Benjamini Hochberg method). The improvement in performance of NP Ultra over NP 1.0-like data extended to classification using all six unit class categories as well: RS units were classified with the highest accuracy (mean $\pm$ SD = 0.83 ± 0.10) followed by NS_S (0.71 ± 0.08), SST (0.67 ± 0.10), PV (0.59 ± 0.09), VIP (0.49 ± 0.09), and NS_L (0.48 ± 0.11 , **Fig. 8G,I**). Note that errors between PV and NS_L, which represent the largest source of error (**Fig. 8G**) presumably reflect the fact that these two classes are largely overlapping as discussed above.

Classification performance was even higher when considering only the three unit classes for which we have ground truth identification, and still superior for NP Ultra relative to 1.0. We performed classification using only those neurons that had been optotagged (PV, SST, and VIP; **Fig. 8H**). To achieve maximum accuracy, we used the NEMO algorithm (Neuronal Embeddings via Multimodal contrastive learning) for classification (Yu et al., 2024), an approach based on contrastive learning from waveform and autocorrelogram features (see Methods). The accuracy among only these three types was high, 80.33% on average (**Fig. 8H**). While accuracy was similar for NP Ultra and NP 1.0-like data using 1-ch features alone (NP Ultra = 0.75 ± 0.05 , NP

1.0-like = 0.76 ± 0.11 , $p = 0.57$, **Supp. Fig. S28D**), NP Ultra outperformed NP 1.0 when including spatial footprint (1-ch waveform: NP Ultra = 0.77 ± 0.05 , NP 1.0-like = 0.68 ± 0.05 ; 1-ch features + footprint: NP Ultra = 0.74 ± 0.05 , NP 1.0-like = 0.56 ± 0.07 , $p = 3.6 \times 10^{-42}$), likely because NP Ultra was better able to estimate spatial footprint due to its increased site density.

The importance of spatial footprint as a feature useful for discriminating unit class was reinforced by further analyses. Spatial footprint varied significantly with cell type (**Fig. 7D**). Mean classification accuracy using NP Ultra data, but not NP 1.0-like data, increased significantly when including footprint as a feature relative to performance without footprint (**Fig. 8F**; $p = 0.0003$). Classification of PV, SST, VIP, and NS_L improved when footprint was included and by a greater margin when compared to NP 1.0-like classification (**Fig. 8I**). Next, we iteratively removed each feature and assessed the difference in overall performance, a measure of feature importance (**Fig. 8J**). Doing so revealed that removing footprint impacted classification performance as much as spike duration (mean accuracy $\pm$ S.D. = 0.61 ± 0.04 for each) and more than the majority of other single-channel features, with the exception of average inter-spike interval and baseline firing rate. Classifier accuracy was still above chance levels even when only decoding from footprint, average ISI, and baseline FR (**Fig. 8J**, inset; mean accuracy = 0.38 ± 0.03). In summary, we demonstrate that NP Ultra probes offer an enhanced ability to discriminate unit classes recorded from the visual cortex compared to classifiers using NP 1.0-like data, and that they do so in particular because of the high resolution spatial information captured by NP Ultra.

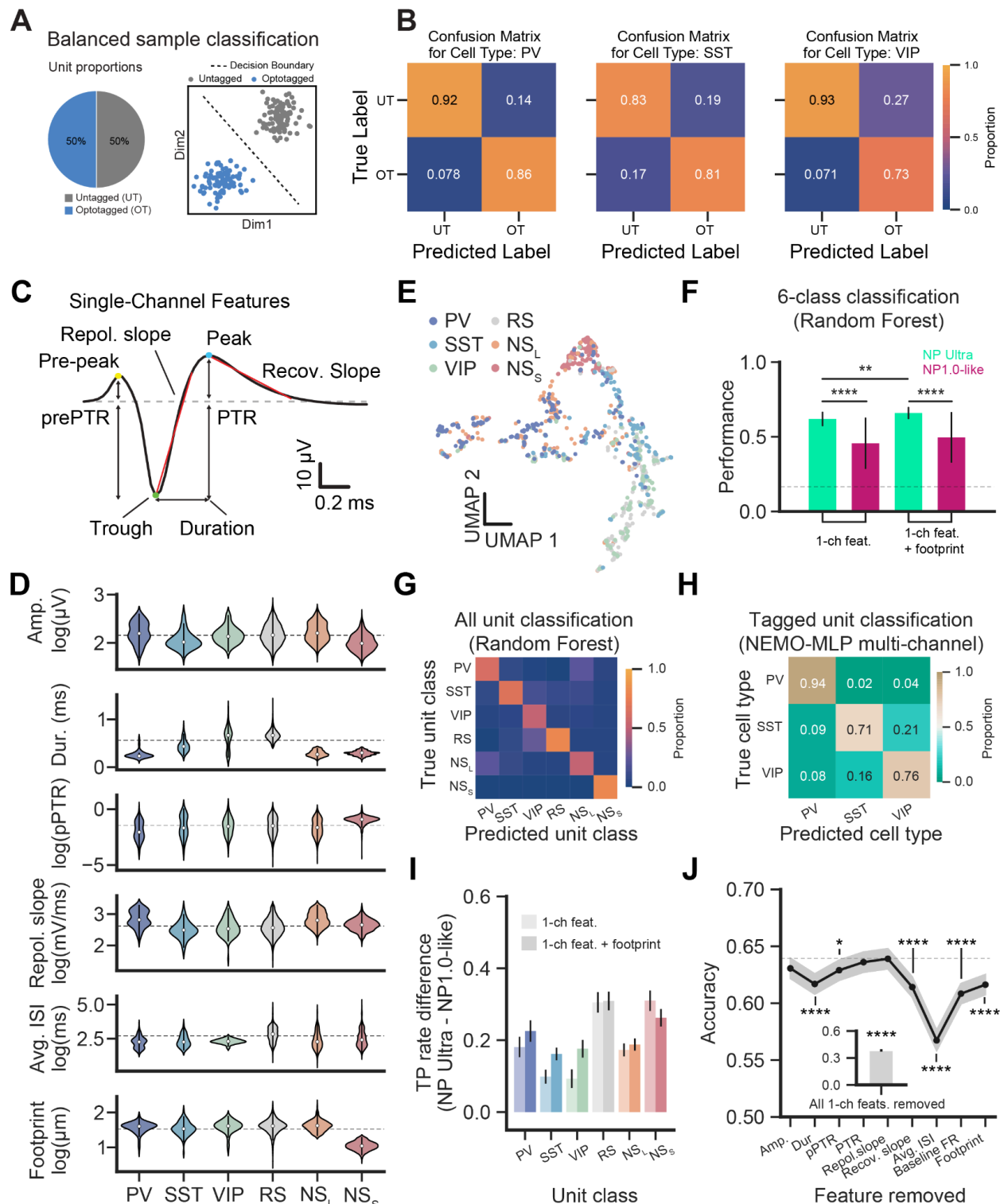


Figure 8. Enhanced classification of inhibitory neuron types with NP Ultra. **A**, Schematic of the classification scheme using balanced unit proportions and a logistic regression classifier with all features included. **B**, Confusion matrices for each interneuron cell type classified via logistic regression using balanced samples and all features. UT = Untagged, OT = Optotagged. **C**, Schematic of features extracted from single-channel waveform ("1-ch features"). Amplitude is the absolute difference between trough and

peak. Duration is the time between trough and peak. Peak-to-trough ratio (PTR) is the ratio between absolute amplitudes of peak and trough, prePTR is the ratio of the absolute amplitudes of the pre-peak and trough. Red lines show slopes for repolarization and recovery. **D**, Comparison of waveform features extracted from the peak amplitude channel for each unit within a transgenic mouse line (all metrics except duration are log-scaled). All untagged units were split into three groups based on spike duration (RS = regular-spiking, NS = narrow-spiking) and footprint size (NS_L and NS_S). Dots indicate the median and black vertical lines indicate the interquartile range. Dashed horizontal lines indicate the population median for each metric. **E**, UMAP projection of an equally sampled subset of NP Ultra-recorded units from each unit class. **F**, Classifier performance (out-of-bag score, mean $\pm$ S.D.) on NP Ultra (green) and NP 1.0 (magenta) data using single and multi-channel features, combined across all six classes. **G**, Confusion matrices for all unit class classification using 1ch features + footprint. **H**, Confusion matrices for classification amongst the three types of opto-tagged interneurons only, using NEMO (see Methods). **I**, Difference in True Positive (TP) rate between NP Ultra and NP 1.0-like data for each unit class. **J**, Classifier accuracy (mean $\pm$ 95% confidence interval) across all unit classes when one feature is iteratively excluded from the dataset. Asterisks indicate significant changes in accuracy as a result of feature removal. Dashed line is the mean classifier accuracy with all 1-ch features + footprint. Inset shows the accuracy of the classifier when no 1-ch features are included in the dataset.

Discussion

Here, we introduce Neuropixels Ultra, a device capable of recording extracellular neural activity with an unprecedented site density (1.3 sites/ μm). These new probes have small site size and spacing, resulting in a trade-off in recording span compared to NP 1.0 probes (288 μm vs. 3840 μm vertical span), but allowing the detailed spatial structure of electrical fields to be sampled with unprecedented resolution. By harnessing this improved spatial resolution for recording extracellular data and employing tailored spike sorting approaches, we demonstrate significant improvements in extracellular data quality, including higher spike amplitudes and improved signal-to-noise ratios, resulting in increased yield of functionally responsive neurons. Recordings made with NP Ultra probes capture subcellular features – axonal and dendritic signals – and provide a new window into the biophysics of action potential initiation and propagation. Finally, recordings at this density offer enhanced discrimination between cell types, specifically demonstrated for interneuron cell types within the visual cortex. These devices therefore enable high-resolution measurements across multiple brain regions and species. Moreover, our shared large-scale datasets provide a useful resource for spike sorting algorithm development, for electrode array design, and for biophysical modeling.

The NP Ultra device may be optimal for multiple experimental objectives, but has tradeoffs in recording span versus density relative to other electrophysiological options, and tradeoffs in temporal resolution and cell-type specificity relative to optical recording methods. NP Ultra probes are particularly suitable, relative to NP 1.0 and 2.0, for maximizing yield from thin layers or small structures, such as individual cortical layers, the CA1 pyramidal layer, or subcortical nuclei including the claustrum and parts of thalamus, basal ganglia, midbrain, and hindbrain. The ability to select sites for recording along the probe at different depths and configurations, out of 6,144 available sites, moreover, improves the ease and specificity of targeting such structures relative to other technologies that target small regions such as tetrodes or silicon

arrays with limited site count. Relative to optical methods for recording neuronal activity such as 2-photon calcium imaging, NP Ultra has millisecond temporal resolution (versus typically 30 Hz or less for imaging), the ability to reach deep brain regions with minimal tissue damage (versus inserting large lenses), and does not require genetic manipulations. However, NP Ultra has reduced spatial resolution (unable to record from small dendrites) and reduced ability to target specific neuronal types (~85% accuracy for PV, SST, and VIP cortical interneurons), compared to imaging methods. Therefore, NP Ultra will be useful for studies of dendritic back-propagation, axonal signaling, and cell type-specific coding when the required temporal resolution, brain region, or species is incompatible with imaging-based approaches.

We observed higher spike amplitudes and improved neuron yield with NP Ultra relative to previous probes. Larger amplitudes can be attributed to the interaction of two factors: 1) higher spatial sampling, which increases the likelihood of recording closer to the site of action potential initiation (Fiáth et al., 2018), and 2) smaller recording sites average signals over a smaller area, avoiding underestimation of peak amplitudes as much as large sites do (Canakci et al., 2017; Viswam et al., 2017; Hill et al., 2018). Additionally, NP Ultra probes exhibited a significantly higher signal-to-noise ratio compared to lower density probes. Although smaller sites incur a penalty in noise level, the increased site density of NP Ultra allows for the combination of information from multiple sites, effectively reducing the averaged noise level. Finally, NP Ultra has more precise localization of spikes for improved discrimination of nearby units. Combined, these benefits in spike sorting resulted in NP Ultra having a consistently higher yield of neurons with reliable visual responses recorded from mouse visual cortex compared to NP 2.0 and NP 1.0. Much of this improvement in yield results from an increased ability with NP Ultra probes to capture low amplitude spikes, which would have undetectably low SNR on probes with lower spatial densities. The improvement in the yield of well-isolated neurons passing rigorous quality metrics (International Brain Laboratory et al., 2023a) was somewhat more modest (Ultra vs. 1.0, 170%) than the improvement in yield of all visually responsive neurons (Ultra vs 1.0, 220%). Such neurons that are less well-isolated would nevertheless be useful for analyses such as dimensionality reduction or decoding (Trautmann et al., 2019; Zhang et al., 2024). We expect that whether users require well-isolated single neurons or multi-unit clusters, the ~2x improvement from NP Ultra will be beneficial for many scientific applications in which the recording span of NP Ultra (~300 μm in the full density configuration) is suitable for the recording target.

Recordings obtained from various brain regions in four species unveiled a remarkable diversity of extracellular neural waveforms. Particularly noteworthy was the presence of a population of "small footprint" units detected using NP Ultra probes across brain regions in different species, operationally defined here as units with an electrical waveform extent of less than 20 μm (smaller than the smallest distance between contacts on an NP 1.0 probe). In the mouse visual cortex, we observed that recordings from identified axons all exhibit small footprints. Despite the commonly accepted wisdom that axonal spikes have early positive peaks (Cooper et al., 1969; Deligkaris et al., 2016; Someck et al., 2023), we observed a fraction of axonal waveforms lacking such positive peaks, consistent with simulations of recordings from nodes of Ranvier in myelinated axons (Haines et al., 2024). Our high resolution recordings therefore reveal that the

spatial footprint, rather than a particular peak direction, is a more reliable signature of axonal recordings. Moreover, while recordings of axonal spikes have been reported with Neuropixels probes (Schröder et al., 2020; Sibille et al., 2022) and with other devices (Merrill et al., 1978; Goldberg and Fee, 2012; Robbins et al., 2013; Barthó et al., 2014; Sun et al., 2021), it has been assumed that this is only possible in limited situations when axons have particularly favorable properties (McColgan et al., 2017; Barthó, 2021). Here, we instead demonstrate these putative axonal recordings in all 15 brain regions and 4 species that we tested, establishing that NP Ultra probes can reliably achieve these recordings. Careful confirmation of the identity of these units will be required in each brain area in the future.

In mouse dentate gyrus (DG) and fish cerebellum (CB), we observed a significant fraction of small footprint units that might represent action potentials from small neuron types, axons, or both. Both DG and CB are notable for their populations of granule cells with very small somata. Considering DG granule cells, however, previous studies have indicated that they are relatively quiescent (Senzai and Buzsáki, 2017; Neubrandt et al., 2018; McHugh et al., 2022), firing at frequencies of about 0.15 spikes per second on average *in vivo* during wakefulness. The short duration of our DG recordings (approximately 5 minutes) would make it difficult to isolate such sparsely firing neurons and might have resulted in failure to record this cell type in general, regardless of the size of its extracellular action potential. Furthermore, we note that the majority of units recorded from the dentate gyrus were narrow-spiking, an attribute not typically associated with mature granule cells (Pedroni et al., 2014), but which we observed in the small proportion of small footprint units in regions such as visual cortex and in small footprint units of corpus callosum. Therefore, while it appears plausible that a subset of the units in our recordings may represent extracellular measurements of granule cells in DG and CB, this cannot be decisively assessed without further ground-truth experiments, such as recording genetically identified granule cells via opto-tagging.

In the visual cortex, the narrow action potential waveform (≤ 0.4 ms) spikes has traditionally been associated with putatively PV-expressing narrow-spiking (NS) inhibitory interneurons (McCormick et al., 1985; Barthó et al., 2004; Kvitsiani et al., 2013). We discovered that the majority of optotagged neurons from genetically identified inhibitory subclasses exhibited large waveform footprints, consistent with their soma sizes being comparable to those of pyramidal neurons (Beebe et al., 2016; Elabbady et al., 2024). Our results indicate that PV+ optotagged neurons correspond to large footprint narrow-spiking units (NS_L), while NS_S units represent axon segments of both excitatory and inhibitory neurons. NS_S units constitute a small, yet significant portion of the total units (~10%) and almost one-third of all narrow-spiking units in NP 1.0 data. Reliable identification of putative PV+ interneurons therefore requires considering both waveform duration and footprint, without which approximately one third of those units identified by waveform duration alone may be misidentified.

Our ability to classify cell types of recorded neurons was generally high within each cell type, and classification between cell types was significantly improved using NP Ultra compared to 1.0 probes. Nevertheless, our findings are applicable to existing recordings made with previous probes, enabling the identification of recorded units as PV+, VIP+, SST+, or other based on

single-channel waveforms with a reasonable level of accuracy. In fact, our category of 'untagged' units includes neurons from the PV, VIP, and SST cell types, which likely leads to an underestimation of classification accuracy (i.e., some untagged units labeled by the classifier as VIP might in fact be VIP+, though this is counted as an error). In the future, unit classification using a variety of additional metrics, such as non-linear waveform features (Lee et al., 2021), autocorrelograms (Barthó et al., 2004; Yamin et al., 2013; Senzai and Buzsáki, 2017), and oscillatory phase locking (Reifenstein et al., 2016), as well as alternative machine learning approaches (Crawford and Pineau, 2019; Tymochko et al., 2020; Kingma and Welling, 2022; Beau et al., 2024; Lee et al., 2024), may yield even greater performance.

Taken together, our findings collectively highlight the advantages of electrophysiological probes with increased site density for a wide range of important neuroscience applications. While sacrificing recording span, Neuropixels Ultra probes offer improved signal quality, making them ideal for recording from small neuron types and subcellular compartments, maximizing yield in dense brain regions, and accurately distinguishing cell types of recorded neurons. Moreover, our results on the axonal nature of small footprint waveforms and on the differences in footprint between cell types suggest that spatial footprint will be a useful feature for identification of recorded waveforms in many electrophysiological experiments.

Methods

Probe noise model

When using neural probes with planar microelectrodes, the quality of the neural recording (e.g. signal to noise ratio) will depend on several factors, including the electrode size; the electrode impedance and the input impedance of the recording amplifier; the electrode noise and the noise of the recording amplifier; and finally the distance and alignment between the neuron and the electrode. These factors contribute to different signal-degradation effects that are explained as follows.

Signal attenuation due to electrode impedance

The electrode impedance is inversely proportional to the electrode size. Small electrodes can exhibit larger signal attenuation due to the impedance ratio (i.e. voltage divider) at the input of the amplifier. This attenuation is given by: Z_{elec}/Z_{amp} . In the Neuropixels 1.0 design, Z_{amp} corresponds to 27.2 MΩ (at 1 kHz). Thus, the signal attenuation can be calculated as:

	Neuropixels 1.0 & 2.0	Neuropixels Ultra	Notes
Electrode area	144 μm ²	25 μm ²	
Z_{elec}	~100 kΩ	500 kΩ*	* Estimated from test structures

Z_{amp}	27.2 M Ω	27.2 M Ω	
$Z_{elec}/Z_{amp} * 100\%$	0.4%	1.8%	

Table 1. Estimated signal attenuation of Neuropixels probes.

Although the signal attenuation is increased by almost 5 times, it continues to be negligible due to the much higher input impedance of the recording amplifier.

Noise performance

The total noise affecting neural recording has two components: i) the thermal noise generated by the electrode-tissue interface (V_{n-elec}) and ii) the noise of the readout electronics (V_{n-amp}). The total noise can be calculated as:

$$V_{n-total} = \sqrt{V_{n-elec}^2 + V_{n-amp}^2}$$

The noise generated by the electrode-tissue (or electrode-electrolyte) mostly depends on the electrode area, the double-layer capacitance formed at the electrode-electrolyte interface (i.e. electrode impedance) and the resistivity of the saline solution, medium or tissue (Yang et al., 2009; Lopez et al., 2012; Viswam et al., 2017).

In the Neuropixels 1.0 design, V_{n-amp} corresponds to 5.4 μV_{rms} (for the action potential band: 300 Hz to 10 kHz). For the low-impedance TiN electrodes, the electrode noise was modeled and measured in saline as 2.7 μV_{rms} . Based on estimated electrode impedance values, it is possible to predict the noise performance of smaller electrodes for different solution/tissue resistivities (Lopez et al., 2012). In saline, the total noise can be calculated as:

	Neuropixels 1.0	Neuropixels Ultra	Notes
Electrode area	144 μm^2	25 μm^2	
Z_{elec}	~100 k Ω	500 k Ω^*	* Estimated from test structures
V_{n-elec}	2.7 μV_{rms}	5 μV_{rms}^{**}	** Modeled based on estimated impedance
V_{n-amp}	5.4 μV_{rms}	5.4 μV_{rms}	
$V_{n-total}$	6 μV_{rms}	7.4 μV_{rms}	

Table 2. Estimated noise of Neuropixels probes in saline.

If we consider the resistivity of the brain tissue, which is at least 4 times higher than the resistivity of saline, the total noise can be calculated as:

	Neuropixels 1.0	Neuropixels Ultra	Notes
Electrode area	144 μm^2	25 μm^2	

Z_{elec}	~100 k Ω	500 k Ω^*	* Estimated from test structures
V_{n-elec}	4.6 μV_{rms}	7.6 μV_{rms}^{**}	** Modeled based on estimated impedance
V_{n-amp}	5.4 μV_{rms}	5.4 μV_{rms}	
$V_{n-total}$	7.1 μV_{rms}	9.4 μV_{rms}	

Table 3. Estimated noise of Neuropixels probes in the brain.

Note that in Neuropixels 1.0, the noise is dominated by the readout-electronics noise, while in Neuropixels-Ultra the electrode noise has a bigger impact on the total noise.

Although the electrode noise increases by a factor of ~2, the increase of the total recording noise is only 23% in saline or 32% in brain tissue. This increase is modest relative to the amplitude of ‘biological’ noise (i.e. spiking activity of distant neurons). This indicates that, thanks to the low-impedance of the TiN electrode material, the reduction in electrode size does not significantly impact the quality of the recordings.

Noise and gain characterization

In vitro noise measurements are performed in the standard self-referenced configuration as described in the Neuropixels manual, with the reference and ground connected together and to a Pt wire electrode in the saline bath. Noise on each channel is measured by averaging Fourier power spectra from 5x 3-second-long sections of data, and estimating rms from the integral over 300-10000 Hz:

$$rms = \sqrt{P * binwidth} = \sqrt{P * \frac{fs}{N_{FFT}}}$$

Where P is the sum over the frequency range in the power spectrum, F_s is the sampling frequency, and N_{FFT} is the number of points in the FFT calculation.

Noise is estimated from *in vivo* data for “low activity” channels with event rates < 0.01 Hz using (1.4826*median absolute deviation). For the gain measurement, the probe ground and reference are connected to the Faraday cage ground, and a 2 mV, 3 kHz sine wave is applied to the saline bath through a Pt wire electrode. The amplitude of the sine wave on each channel is measured by averaging over a 15 Hz window about the 3 kHz peak in the Fourier power spectrum.

Crosstalk estimation

To estimate the crosstalk between channels, we took advantage of our *in vivo* recordings which provide a situation in which a small voltage source is located immediately adjacent to the probe. We reasoned that for neurons with spatially restricted extracellular action potentials, distant sites

on the probe should not reflect the waveform of the neuron unless by crosstalk. To the extent that this assumption might not be true (i.e. distant sites really do detect the actual extracellular action potential), we would measure signal at the distant site and incorrectly attribute it to crosstalk; therefore, this estimation strategy produces an upper bound on the true amount of crosstalk.

Specifically, we selected units for which at most 20 sites had significant amplitude of the average waveform (defined as peak-trough amplitude > 25% of the amplitude on the peak site), and averaged the amplitude at the 10 most physically distant sites relative to the peak site. The amplitude on the distant sites was calculated using the same peak and trough timepoints as the peak channel, since any crosstalk should be temporally instantaneous. We report the ratio between this average amplitude on the distant 10 sites and the amplitude on the peak site. This measurement was subject to biological noise, but across an average of $n=519$ small neurons, we measured $0.0007\% \pm 0.0736\%$ (mean $\pm$ s.e.m.; **Supp. Fig. S4**).

Light artifact tests

We tested the sensitivity of NP Ultra probes to 470 nm light (the wavelength used for channelrhodopsin activation). We used a fiber-coupled laser (200 μm core fiber) to shine light directly on probes submerged in phosphate-buffered saline. Light was incident on the front side of the probes. During the tests, we simultaneously recorded from two NP Ultra probes (passive, non-switchable version) and one NP 1.0 probe to compare light artifacts across probe variants. Two different stimulation waveforms were tested (10 ms pulse and 1 s raised cosine stimulus) at three light intensities (0.5, 4.0, 10.0 mW/mm^2). We quantified photo-artifacts in both the spike (0.3-10 kHz) and LFP bands (0.5-500 Hz) by computing trial-averaged photo-evoked responses (100 trials) time-locked to the start of photo-stimulation (**Supp. Fig. S5**). Artifact magnitudes were computed as the absolute value of the light-evoked voltage deflection peak during the photo-stimulation period.

Recordings from diverse brain regions in the mouse

At UW

All experimental protocols were conducted according to US National Institutes of Health guidelines for animal research and approved by the Institutional Animal Care and Use Committee at the University of Washington.

For awake, head-fixed acute recordings, wild type C57BL/6 mice of both sexes, between 2 and 8 months of age, were first implanted with custom-made steel headplates and 3D-printed plastic recording chambers. Following recovery, mice were acclimated to head-fixation for at least two sessions before recording. Head-fixation mice were seated on a plastic body restraint tube with forepaws on a rotating rubber wheel. Three iPad screens were positioned around the mouse at right angles. On or before the first day of recording, a 1-2 mm diameter craniotomy with intact dura was prepared with a dental drill over the target brain area under anesthesia. The

craniotomy was either protected with removable silicone sealant (Kwik-Cast, World Precision Instrument) or covered with transparent dura-gel (Dowsil 3-4680) with a further protective plastic cap sitting on top of the recording chamber, after the surgery and before recording procedure starts. After several hours of recovery, mice were head-fixed in the recording setup. We used single-ended configuration for all recordings, with ground and reference pins on the probe tied together. We used either external or internal reference methods. For recordings with external referencing, an Ag wire was connected to the probe ground/reference wire and positioned above the skull. After peeling off the silicone sealant, the craniotomy and Ag wire were then submerged in a bath of Ringer's lactate solution. For recordings with internal reference, a probe was directly inserted through the dura gel without solution bath and the reference site on the probe tip was used instead. Prior to each insertion, the electrode shank was coated with CM-Dil (Invitrogen), a red lipophilic dye, for later histological reconstruction of probe tracks.

To sample waveforms from diverse brain areas in a single penetration, we automated the advancement of probes through manipulators (Sensapex Inc., uMP-4) using the Sensapex API via Matlab, following every 5 minutes of stable recordings. For each step, we advanced the probe by 300 μm at 5 $\mu\text{m}/\text{sec}$ speed, followed by 50 μm retraction, to help reduce the tissue compression caused by the probe. Recordings normally stabilized within one minute after probe motion, as visualized by post-processing drift maps. We typically repeated this for 8-16 steps from the surface of the penetration, depending on the final depth of target brain area. For most sessions, we recorded with two probes simultaneously on two hemispheres in a single session.

After recording, mice were perfused with 4% paraformaldehyde. The brain was extracted and then fixed in 4% paraformaldehyde for a further 24 hours at least at 4 °C. We cleared the brain tissue with iDISCO before 3D imaging with lightsheet microscopy (UltraMicroscope II, LaVision BioTec). We used 561 nm channel for imaging probe track with Dil and 488 nm channel for brain autofluorescence, with 10 μm isotropic resolution. We registered the imaging volume to Allen Mouse Brain Common Coordinate Frameworks (CCF) Atlas with ARA tools (https://github.com/SainsburyWellcomeCentre/ara_tools), which calls Elastix via a Matlab wrapper. Probe tracks were traced in the atlas-registered brain volume with open-source package Lasagna (<https://github.com/SainsburyWellcomeCentre/lasagna>). While the Dil track provides a good estimate of the probe trajectory, the depth information between Dil track and electrophysiological recordings usually do not match with high precision, due to the warping and shrinkage of the brain sample during perfusion and tissue clearing steps, and due to uncertainty about the final probe tip location along the track. Therefore, we used multiple electrophysiological landmarks (e.g., cortical layer 5 neuron layer, CA1 pyramidal layer, white matter boundaries, etc) to match the recording sites with the probe trajectory by linear interpolation (Steinmetz et al., 2019). After this manual curation process, we assigned each probe site and therefore each spike cluster to designated CCF coordinates.

The NP Ultra brain-wide data was sorted with Kilosort 2.0, using a 96 site template. The timing of all recording epochs were extracted from a manipulator motion start/stop channel synchronized to the electrophysiology data. The first minute of the recording after manipulator motion was discarded due to probe-tissue motion, leaving 4 minutes of stable recording data for

each epoch. Kilosort was then batch run based on epoch timing, and units were selected with an automated quality metric within Kilosort 2.0 after spike sorting.

At Janelia

Acute recordings were conducted at the HHMI Janelia Research Campus, following guidelines set by the Institutional Animal Care and Use Committee. Three male VGAT-ChR2-EYFP mice, aged 6 months, underwent a brief stereotaxic surgery to implant a titanium headpost and build a recording chamber using dental cement for acute head-fixed recordings. Postoperative analgesia was provided using Buprenorphine (0.1 mg/kg, intraperitoneal injection), and Ketoprofen (5 mg/kg, subcutaneous injection) was used at the time of surgery and postoperatively for two days. Following surgery, mice were habituated under head restraint in the recording setup for two days with incremental acclimatization duration (30 minutes to one hour).

On the first day of recording, two to four craniotomies of diameter 1-1.5mm were made with a dental drill over the target areas of interest under anesthesia, while leaving the dura intact. The craniotomies were protected with removable silicone sealant (Kwik-Cast, World Precision Instrument) sitting on top of the recording chamber between recording sessions. After a three-hour recovery period, the animal was head-fixed in the recording setup, and two or four NP Ultra probes were inserted into separate craniotomies and to the target depths, inferred from manipulator readings. The insertions covered brain regions such as the cortex, striatum, thalamus, midbrain, cerebellum, and medulla. The tip of the electrode was coated with CM-Dil (Invitrogen), a red fixable lipophilic dye, before each insertion to enable histological reconstruction of probe tracks. Probes were slowly lowered at a speed of 10 $\mu\text{m/s}$ to the first intended depth using micromanipulators (uMP-4, Sensapex, Inc), and after reaching the desired depth, the inserted probes were allowed to settle for approximately 5 minutes before recording commenced. Multiple probes were recorded simultaneously for 20-30 minutes. Since the NP Ultra sites span a length of 288 μm only, after each recording, probes were further inserted by 300 μm to make a successive recording. As such, over the course of a session, probes were advanced 2-6 times to collect data at multiple depths from each penetration, and brain tissue was allowed to relax for a few minutes after each probe advancement. All probes had a wire soldered onto the reference pad, which was shorted to the ground pad. This wire was then connected to an Ag/AgCl wire positioned above the skull. During recordings, the craniotomies and reference wire were submerged in cortex buffer (NaCl 125 mM, KCl 5 mM, Glucose 10 mM, HEPES 10 mM, CaCl_2 2 mM, MgSO_4 2 mM, pH 7.4). Daily recording sessions lasted approximately two hours and were repeated for multiple consecutive days in a craniotomy, with each penetration separated by at least 200 μm at the point of entry. All recordings were made with open-source software SpikeGLX (<http://billkarsh.github.io/SpikeGLX/>) in external or tip reference mode, and the animal was awake and sitting quietly in a body restraint tube throughout the recording sessions. The dataset was sorted offline with Kilosort 2.0 using a 96 site template.

To visualize the recording locations, the mice were perfused transcardially with PBS followed by 4% PFA, and their brains were fixed overnight and then cleared with SPiB solution using an aqueous-based clearing method called 'Uniclear' to facilitate whole-brain dilapidation and refractive index matching (Chen and Svoboda, 2020). The cleared brains were then imaged with a light sheet microscope (Zeiss Z1) using a 5x objective with a voxel size of 1.2 x 1.2 x 6 μm . Whole-brain autofluorescence was captured with a 488 nm channel, while Dil tracks were imaged with a 561 nm channel. The resulting tiled images were stitched together using IMARIS software, and the entire brain stacks were registered to the Allen Institute Common Coordinate Framework (CCFv3) of the mouse brain based on anatomical landmarks, allowing each spike cluster to be assigned to a specific CCF coordinate.

Data was preprocessed using CatGT (<https://billkarsh.github.io/SpikeGLX/#catgt>) for multiplex correction, filtering, artifact removal, and common average subtraction. Spike sorting was performed with Kilosort 2.0, followed by manual curation. We classify units as either 'unimodal' or 'overlapped'. Unimodal units have distributions in feature space visually that appear to come from a single distribution that does not overlap other units, and refractory period violations are low. 'Overlapped' units have distributions in feature space that overlap other units, and generally have higher refractory period violations. In the analysis of signal and physical distance, all units with > 500 spikes are included, to provide as complete a picture of the activity as possible. For the waveform shape analysis, only 'unimodal' units are included, so that mean waveforms are as close as possible to single units.

At University College London

All experimental procedures were carried out under license from the UK Home Office in accordance with the UK Animals (Scientific Procedures) Act (1986). Acute recordings were performed on adult (>P60) Rbp4-Cre male and female mice kept on a 12 hours dark/light cycle. At the time of surgery, mice were subcutaneously injected with Rimadyl (5mg/ml) and Lidocaine (5mg/ml) to numb the area above the skull prior to headplate implantation. A 0.5 g customized metal headplate was attached to the skull using super glue and dental acrylic (Super bond C&B polymer mixed with Super bond monomer and catalyst). After headplate implantation, mice were allowed to recover at 37 °C inside an incubator and were closely monitored for 5 days post-surgery. On the first day of recording, mice were administered dexamethasone (Dexadreson, 2 mg/ml, intramuscular injection) to reduce brain swelling. On the first day of recording, one 2 mm-diameter craniotomy was made over the region of interest under anesthesia using a surgical motorized drill. The skull flap was carefully removed to leave the dura mater intact and the exposed brain surface was cleaned using IVE and subsequently protected with transparent Duragel and removable silicone sealant (Kwik-Cast, World Precision Instrument) placed on top of the recording chamber. Finally, the animal was allowed to recover for 3-hours before starting any further experimental procedure.

Before recording sessions, mice were acclimated to the experimental rig for at least 3 days: animals were head-fixed above a static 3D-printed holder using a head bar clamp, and habituated to head restraint of increasing duration (starting from 10 minutes and up to 1h). On

recording days, the shank of Neuropixels probes was covered with red fluorescent dye Dil (Invitrogen) to allow for histological probe tracing. Then the probe was inserted perpendicularly to the cortical surface using a micromanipulator (Sensapex uMP-4). Probe insertion speed was kept at 20 $\mu\text{m/s}$ to minimize cortical damage, terminating at 2000-2800 μm below the brain surface for NP 1.0 and 1200 μm for NP Ultra. Probes were allowed to settle for about 10 minutes before the start of every recording. All recordings were carried out using the acquisition software SpikeGLX (<http://billkarsh.github.io/SpikeGLX/>) using tip reference mode, and lasting about 1.5-2 hours. Every animal was recorded for a maximum of three consecutive days, with recordings taking place at different locations within the craniotomy.

All recordings were processed to align channel sampling using the common average referencing software CatGT (<https://billkarsh.github.io/SpikeGLX/#catgt>); putative neuronal units were spike sorted with Kilosort 2.0 (<https://github.com/MouseLand/Kilosort2>) and Phy2 (<https://github.com/cortex-lab/phy>), and manually curated in Phy2. For every unit, we computed spike-triggered voltage maps showing the extracellular signature of each action potential across time and channels. We then averaged 2000 of these maps to compute a template spike voltage map for each unit. We located the putative soma as the channel with the largest sink recorded at the time of spike in the template maps.

To correct for vertical probe motion during the recording we used a two-step strategy. First, we used Kilosort 3.0 drift estimation to correct for slow probe movements at the second scale. Second, to correct for fast movements on a spike-by-spike basis, we used 1D cross correlation to compute the vertical shifts needed to register spike-triggered voltage maps to a target reference map obtained from 1000 long-ISI spikes selected randomly throughout the recording.

To identify units with significant dendritic back-propagation, e.g. those for which the apical dendrite was best aligned with the probe, we used a permutation test. Starting from the soma, we compared the average spike amplitude of every channel in the template spike map to a null distribution of amplitudes obtained from 500 shuffled template maps, each computed from a random linear shift of the spike times. Moving from towards the cortical surface, we considered as significant dendritic recordings those channels bearing signals larger than 95th percentile of the null distribution. We included for further analysis neurons which displayed significant dendritic back-propagating signals extending beyond 250 μm from the soma.

Muscimol experiments

To verify the identities of small footprint units in the mouse cortex, we topically administered muscimol to the surface of the dura through the craniotomy, while acutely recording cortical activity with NP ultra probes in head-fixed mice. Acting as a GABA agonist, muscimol will silence neural activity of local cortical neurons, but leaving activity of long range axons from elsewhere intact.

Headplate implant surgery, craniotomy surgery and acute NP ultra recordings were performed mostly as described in the Methods “Recording from diverse brain regions in the mouse (At UW)” section, with additional details as below. During the craniotomy surgery procedure, a cement well was built around the craniotomy after the craniotomy was made and covered with removable silicone sealant, to constrain the spatial spread of drug application within the craniotomy during recording. For acute NP ultra recordings, the craniotomy was submerged in a bath of Ringer’s lactate solution during control conditions. After a period of baseline spontaneous activity recording for around 10 minutes, a set of 5-minute sparse noise stimuli were presented on the iPad screens for receptive field mapping. Thereafter, 5mM muscimol in Ringer’s solution was applied to the craniotomy and “bathed” the surface of the brain for the remainder duration of the recording. The majority of unit activity were silenced, 3-10 minutes after the muscimol application. A second set of 5-minute sparse noise stimuli were then presented for receptive field mapping for units that survived muscimol.

Spike sorting methods

To take advantage of the high spatial resolution of NP Ultra sites, we utilize a spike sorting pipeline, DARTsort (Boussard et al., 2023), that takes advantage of recent progress in spatial localization of spikes. After an initial detection step, we denoise (Lee et al., 2020) and localize the spikes (Boussard et al., 2021). Localization is obtained in 3D by modeling the source of spikes as a point-source of brightness a and location (x,y,z) and solving for (a,x,y,z) by minimizing the L2 loss between the observed spike amplitudes on all channels and the point-source reconstructed amplitude. We then estimate the motion of the probe (i.e., drift) relative to the recorded neurons. For this registration step we use the decentralized approach from (Windolf et al., 2025), which is more robust to boundary effects (i.e., neurons moving on or off of the short NP Ultra probe) than the template-based registration method used in Kilosort2.5 (see (Windolf et al., 2025) for further analysis). After registration, we have a robust estimate of each spike’s amplitude and position relative to the probe. Then, by subtracting the estimated drift from the position of each spike, we obtain a set of 3D features for each unit: registered 2D position (taking only the ‘x’ and ‘z’ coordinates across the face of the probe from the full fit 3D position) and spike amplitude (Hilgen et al., 2017). We cluster in this 3D feature space using HDBSCAN (Malzer and Baum, 2019), and then recursively re-split these clusters using additional features computed by principal components analysis applied within each cluster (Swindale and Spacek, 2014). Even after registration, we find that some units have some remaining drift (changes in mean registered location and amplitude) over time. Therefore we divide the recording into 5-minute segments that we separately cluster and then merge, to make the clustering step more robust to these individual-unit drift effects.

Next we run a template-matching deconvolution step to resolve “collided” spikes that overlap temporally and spatially. To enable drift-aware template matching, we construct a set of denoised “super-resolution” templates for each unit by dividing the depth of the probe into 2 μm bins and then averaging the spikes localized to each bin separately. During template matching, we can then shift these super-resolved templates according to the drift to efficiently detect spikes for each unit. One major benefit of this approach is that we do not need to apply a

shift/interpolation step to the raw data (as in Kilosort 2.5); in practice, we found that this raw-data interpolation removed signal near the edge of the probe due to boundary effects, which are particularly problematic for the small dense NP Ultra probes. In our approach, instead of shifting the raw data, we shift the templates, allowing us to robustly track neurons moving on or off the probe.

As the initial clustering is usually imperfect, we perform the above steps iteratively: cluster, then deconvolve, then re-cluster on the deconvolution output. We use the post-deconvolution collision-subtraction method from (Lee et al., 2020) to denoise the spikes obtained in the deconvolution step, to enable improved iterative clustering.

Methods for site pattern comparison using spatial resampling

We resampled data recorded with NP Ultra to predict what signals would have been recorded from the same brain location and time period with a lower density probe. To do this, we simply average together the four 5x5 μm NP Ultra sites that are co-localized with a single 12x12 μm site in the resampled pattern (**Fig. 2E**; due to the 1 μm gaps between sites in NP Ultra, the four 5x5 sites fill the 12x12 μm area almost exactly, with small missing gaps). We used this method to measure the expected changes in amplitude and waveform shape for the Neuropixels 1.0 and 2.0 patterns, and a “large dense” pattern of 12x12 μm sites with 12 μm pitch (**Fig. 2G-K, Fig. 3D-G, Fig. 8D,F**).

For measurements of the site size and pattern dependence that include the effects of noise - specifically, for computing the ‘Template SNR’ (**Fig. 2I**) and for analyses involving spike sorting the resampled raw data (**Fig. 3D-G**) - we needed to create data with realistic noise and waveform variation, which in real data combines electronic noise with signals from distant units. Averaging the four raw traces together as described above reduces the noise by $1/\sqrt{2}$; to boost the noise back to normal levels, simulated noise with a matched frequency spectrum is added to the averaged traces (code at: https://github.com/jenniferColonell/NP_Ultra_downsample).

An interpolation method was used for the analyses on waveform shape and peak amplitude in **Supp. Figs 8, 9**. The mean waveform from the Ultra data was used to create a gridded interpolant (MATLAB, using ‘makima’ interpolation). The waveforms for an arbitrary pattern with any site size and position can then be estimated by averaging together interpolated waveforms at sets of points that span the model sites. This method has the advantage of allowing a test site pattern that is slightly better matched to the true NP 1.0 and NP 2.0 geometry than in **Fig. 2E**.

Analyses of sorting quality

Spike amplitude was calculated as peak-to-peak voltage, i.e. the maximum voltage of the waveform minus the minimum (**Fig. 2G-H; Supp. Fig. S8A, S9A**).

In template matching methods, a multi-site template is convolved with the filtered data to detect spikes. To quantify the impact of site density on detectability in template matching, we calculate the ‘Template Matching Signal to Noise Ratio’ for each unit (**Fig. 2I**; **Supp. Fig. S8B, S9B**) defined as:

$$\text{TM-SNR} = \frac{\sum_S (T \cdot V_S)}{N \cdot \text{std}(T \cdot V_R)}$$

Where T is the template vector of 40 time samples * number of sites in the template, and V is measured voltage on those sites centered either at a spike time (to measure signal) or a randomly chosen time (to measure noise). S indexes spike times of the unit, R indexes random times, and N is the spike count of the unit. This is an estimate of the signal to noise ratio in data filtered by the template. Spike times for each unit are drawn from the sorted Ultra data, and the TM-SNR is calculated for the Ultra data and spatially resampled data obtained from averaging sites (see exact patterns in **Fig. 2E**).

The spatial spread of spikes was calculated as the standard deviation of the x-coordinates (lateral dimension, across the face of the probe) or of the z-coordinates (axial dimension, down the length of the probe) across all spikes in a given unit (**Fig. 2J-K**). For the resampled patterns, these coordinates were computed by resampling each spike individually and then computing that spike’s estimated coordinates using the same method employed in the main DARTsort pipeline.

To understand the possible effects of this spatial scatter in more detail, we computed the average distances between neurons, as well as the minimum distance that each site pattern should be able to resolve, in a simple simulation (**Supp. Fig. S8C, S9C**). We use a Monte Carlo model to generate sets of peak-to-peak voltages over sites for a point source with 1/R amplitude falloff, placed 15 μm from the probe, with peak amplitude at the probe of 50 μV and rms noise of 14 μV . For each set, all sites within 60 μm of the peak site are fit to obtain an estimated XZ position (Boussard et al., 2021). Modeled noise in the position estimate is the standard deviation of the Monte Carlo trials, and is 5 μm for a NP Ultra probe, and 25 μm for the NP 2.0 pattern. The increased accuracy is due to the inclusion of more sites in the 60 μm radius of fit points, which decreases the impact of noise. These values are the upper limit of position resolution, and can be compared to the measured distribution of nearest neighbor unit distances from the Ultra data. Units with nearest neighbor distances below the spatial resolution of the probe will be more difficult to sort. The experimental nearest neighbor distributions are calculated including both ‘unimodal’ and ‘overlapped’ units to build as complete a set as possible to assess the potential impact of this difference in spatial resolution on sorting. Note that standard deviations of positions in experimental data can be larger (see **Fig. 2J-K**); this is likely due to variability in the spikes (background firing, incomplete drift correction) not captured in the simple model.

To measure how more detailed spatio-temporal waveforms could affect sorting, we created a waveform distance metric and compared the waveform distances measured at varying site density (**Supp. Fig. S8D, S9D**). Higher site density will always provide more detail, and better

confirmation of small differences, but the engineering question is whether the waveforms are ‘mostly’ smoothly varying and distinguishable with sparse sampling. To isolate the waveform distance from difference in z position and amplitude, a neighborhood of rows centered about the peak row is taken for each template, and the amplitude is normalized. For the spatially resampled patterns, the expected waveforms are calculated using interpolation of the full density waveform, with the sparser pattern centered at the position of the peak row for each template. Centering the sparse pattern on the peak row ensures the two templates look as similar as possible, by ensuring that the same part of the footprint is sampled.

A simple metric to compare the two normalized templates [A,B] is the L1 norm of the difference between the two divided by the sum of the L1 norm of both waveforms, summed over sites, that is:

$$\text{multi site wave L1} = \frac{\sum_{\text{sites}} L1(A-B)}{\sum_{\text{sites}} (L1(A) + L1(B))}$$

This quantity is essentially the fraction of detected signal that is different between the two templates. It is a characterization of the waveform shape that does not account for noise in individual spikes.

Yield, decoding, and stability testing with imposed probe-brain motion

Recordings

To make recordings with imposed probe motion in head-fixed mice (n=2 mice, n=6 sessions), we automated the probe motion through manipulators (Sensapex Inc., uMP-4) using the Sensapex API via Matlab. Manipulators were programmed to alternate between moving forward and backward along the probe axis, for a total of 10 steps forward and 9 alternating steps backwards with 25 µm travel in each direction at 1 µm/sec speed, which results in 25 µm net total forward displacement disposition of the probe by the end of imposed motion. A digital synchronization signal was issued at the start and end of each motion step for later alignment with neural data. To match neuron identities before and after imposed motion, a battery of 118 natural images with 60 repetitions per image were presented to ‘fingerprint’ visual responses. Images were presented in a random order. An average of 30 repetitions of the image set were presented before imposed motion, and an average of 15 repetitions were presented after imposed motion. The duration of each image follows a randomized exponential distribution (mean: 0.5 seconds; minimum: 0.2 seconds; maximum: 3 seconds). NP Ultra probes were inserted into the visual cortex at 700-900 µm depth to sample deep layer neurons.

The NP Ultra imposed motion dataset was spike-sorted with DARTsort (see “Spike Sorting Methods”) (Boussard et al., 2023). The original full pattern data, as well as the same data resampled to NP 1.0-like, NP 2.0-like, and 12 µm large dense site patterns (see “Methods for pattern comparison using spatial resampling”, **Fig. 2E**), were sorted. The timing of imposed

motion epochs was extracted from a manipulator motion start/stop channel synchronized to the electrophysiology data.

Yield and stability analysis

To assess and compare the yield and tracking stability of recorded neurons across different channel densities (full density NP ultra, NP 1.0-like, NP 2.0-like, and 12 μm large dense), visual response fingerprints were created for each unit by concatenating average PSTHs (window 0 to +250 ms aligned to image onset, bin size 10 ms) for all 118 images into a single vector. PSTHs were normalized for a given unit by subtracting its baseline activity (-100 to 0 ms) averaged across trials prior to image onset. PSTHs were denoised by reconstructing each PSTH from the top two principal components taken on the average PSTH per image x time points matrix for each unit (dimensions 118x35).

To determine whether a unit had a reliable visual fingerprint, a null distribution for each unit was created by randomly permuting the image identities before concatenating the PSTHs into a vector (**Fig. 3B and Supp. Fig. S12A**). This distribution preserves the noise and average PSTH time-course of a given unit while disrupting its selective visual response fingerprint. Average PSTHs were calculated from odd-numbered pre-motion trials and even-numbered pre-motion trials. Image identities were randomly permuted 1000 times, and odd-numbered pre-motion trials were used to create the null distribution. Units were considered to have a reliable visual response if their original correlation score (i.e. pre-motion odd vs. pre-motion even trials) was greater than all scores in their null distribution (i.e. 1000 pre-motion odd shuffles vs. pre-motion even trials). This ensures that included neurons have both a reliable fingerprint (i.e. responses are consistent between trials of the same image) and a selective fingerprint (i.e. responses to each image significantly vary).

To create a distribution of pre- vs. pre- and pre- vs. post- imposed motion fingerprint correlations (Pearson correlation coefficient), 1000 unique sets of visual fingerprint vectors were constructed for each neuron by randomly partitioning the image repetitions for each image before imposed motion into 2 non-overlapping subsets (15/30 trials in each subset, for all 118 images). In each of these 1000 sets, fingerprints from each pre-motion subset of trials (i.e. the average across trials) were correlated, and the fingerprint from one pre-motion subset was correlated with the fingerprint from the post-motion trials (**Supp. Fig. S12B-C**). The fingerprint stability ratio for a given neuron was calculated by dividing the mean of the 1000 pre- vs. post-motion correlations by the mean of the 1000 corresponding pre- vs. pre- correlations. To compare fingerprint stability tracking across the 4 site patterns, stability ratios for every neuron sorted from a given pattern were pooled across all sessions (**Fig. 3H**). Neurons were only included in this analysis if they had a reliable visual response and had a stable firing rate between the pre- and post-motion periods (neurons were dropped if the mean firing rate across all PSTHs in pre-motion period was >4x different than that of the post-motion period).

Image decoding analysis

To assess whether the increased yield of visually responsive neurons captured by NP Ultra corresponded to increased information about visual stimuli, we decoded natural image identities

from NP Ultra and compared the performance to decoding from LD, 2.0, and 1.0 patterns. Only visually responsive neurons, i.e. those with a reliable visual fingerprint, were included in the analysis. A minimum of 5 visually responsive neurons per probe was required for inclusion in analysis. Only images displayed before imposed probe motion were included in the analysis.

For each visually responsive neuron, we constructed PSTHs for each trial across all images, using 10 ms bins from -100 to 250 ms relative to stimulus onset. We then performed singular value decomposition (SVD) on the binned PSTH arrays of each neuron (size trials x time bins), retaining the first two principal components to obtain a low-dimensional representation of each neuron's activity pattern. These per-neuron components were concatenated to form a population representation of image PSTH responses. A second SVD was then applied to this population representation, and a maximum of 15 dimensions were retained as predictors for decoding (fewer than 15 dimensions if there were 7 or fewer visually responsive neurons available).

Rather than attempting to classify all 118 images simultaneously, we employed a pairwise decoding strategy. For each possible pair of images (6,903 unique pairs from 118 images), we trained a separate binary classifier. Unregularized Linear Discriminant Analysis (LDA) with diagonal covariance estimation was used as the classification algorithm ("diaglinear" option in MATLAB's `fitcdiscr`). To address potential class imbalance, observation weights were applied inversely proportional to class frequencies. Decoding accuracy was assessed using 5-fold cross-validation.

The decoding pipeline (SVD dimensionality reduction → LDA on all image pairs) was run multiple times for each probe with increasing numbers of neurons, starting from the minimum of 5 and incrementally adding neurons until reaching the maximum available in each population. When using less than the maximum number, neurons were randomly selected from the pool of visually responsive cells.

We analyzed decoding performance per session and probe type by calculating mean decoding accuracy across all image pairs and across the top 10% most decodable pairs per session and probe (**Fig. 3F**). When analyzing decoding performance as a function of increasing neuron count, a minimum of 3 sessions was required for a given neuron count on a specific probe type to be included in the analysis (**Fig. 3G**). Statistical comparisons between probe types were conducted using paired comparisons (Wilcoxon signed-rank test) between probe types on the same session.

Biophysical simulations

All neural simulations (**Supp. Figs. S7, S16, S17**) were done through LFPy 2.3 (Hagen et al., 2018) running on NEURON 8.1 (Carnevale and Hines, 2006). Action potentials were evoked by a step current injection (POINT_PROCESS in NEURON), where the amplitude of the current was adjusted for each cell model to evoke an action potential. Calculations of extracellular potentials were done with LFPy, assuming an electrode diameter of 5 μm , incorporated through

the disc-electrode approximation. Simulations in **Supp. Fig. S7** used the rat layer 5 pyramidal cell model from (Hay et al., 2011). Simulations in **Supp. Fig. S17** used the rat layer 5 pyramidal cell from (Hallermann et al., 2012) (available from <https://modeldb.science/144526>), and the parameters for the passive and active conductances for the axon models used in **Supp. Fig. S16** were extracted from the same neuron model.

Recordings from species other than mouse

Recordings from monkey

One macaca nemestrina weighing 6.6 kg (female, 9 yr) participated in the experiments which were conducted in the anesthetized, paralyzed preparation. At the start of the experiment the animal was placed in the stereotaxic instrument and then a craniotomy and a durotomy were performed to target the primary visual cortex (V1). Anesthetic and paralytic regimens are as described elsewhere (McLelland et al., 2015). All animal procedures conformed to National Institutes of Health guidelines and were approved by the Institutional Animal Care and Use Committee at the University of Washington.

The NP Ultra probe was advanced into the cortex using a hydraulic Microdrive (MO-97A, Narishige) which was mounted onto the stereotaxic arm. Once spike waveforms first appeared on the spikeGLX activity map, we retracted the probe 50 μm and mapped the receptive field (RF) of recorded neurons using achromatic bars under experimenter control. Then data collection on the main experiment began, during which we presented images of naturalistic objects (200 images from (Kiani et al., 2007)) repeatedly (presentation duration: 200 ms). Images were sized to cover mapped RFs.

As the data were continuously collected we completed the following sequence of operations 11 times: (1) probe at rest for 10 minutes, (2) advance probe 300 μm deeper over a 2 minute duration, (3) retract probe for 50 μm . This procedure allowed us to record spike waveforms at 11 different depths where every depth was 250 μm deeper than the previous recording site. Following completion of the recording session, we retracted the probe and confirmed that spikes completely disappeared at a depth within 200 μm from where spikes had first appeared during probe insertion.

Stimuli were presented on a liquid crystal display monitor (24 inches; 100-Hz frame rate; 1920x1080 pixel size, XL2430-B, BENQ) using custom experimental control software, Pype2 (Mazer, Jamie, 2013). Neural data were acquired using spikeGLX with internal tip reference.

The data were batch sorted based on epoch timing, with Kilosort 2.0 using a 96 site template, and units were selected with an automated quality metric within Kilosort 2.0 like recordings from other species.

Recordings from lizard

All experimental procedures were performed in accordance with German animal welfare guidelines: permit no: V54-19c20/15-F126/2006 delivered by the Regierungspraesidium Darmstadt (E. Simon).

A day before surgery, the lizard was administered analgesics (butorphanol, 0.5 mg kg⁻¹ subcutaneously; meloxicam, 0.2 mg kg⁻¹ subcutaneously) and antibiotics (marbofloxacin, marbocyl, 2 mg kg⁻¹). On the day of surgery, anesthesia was initiated with isoflurane in an induction box and maintained with 1–4% isoflurane after intubation (Hallowell EMC anesthesia Workstation AWS). The lizard was placed in a stereotactic apparatus (Kopf 963) after ensuring deep anesthesia (absence of corneal reflex). Body temperature during surgery was maintained at 30 °C using a heating pad and esophageal temperature probe (Harvard apparatus part number 50-7213). Heart rate was monitored using an Ultrasonic Doppler flow detector (PARKS medical electronics, INC. model 811-B). The skin covering the skull was disinfected using a 10% povidone-iodine solution before removal with a scalpel. A fine layer of UV-cured glue (Oxford Flow Light Cure Flowable micro-hybrid composite A2) was applied around the exposed surfaces of the skull. A 3×2 mm craniotomy was then drilled around the parietal eye and the pericranium was removed. The parietal eye was retracted and fixed to the posterior edge of the craniotomy using histo-acryl tissue glue. The dura and arachnoid layers covering the forebrain were removed with fine forceps. A silver wire was inserted into the CSF fluid next to the olfactory tract, acting as both reference and ground. Pia was removed over the area of probe insertion (medial cortex). Prior to insertion, the probe was mounted on a movable drive (R2Drive, 3DNeuro), and secured to a stereotactic holder. The probe was stained with Dil (Invitrogen™, Vybrant™ Dil cell labeling solution) and implanted 500 µm deep in the medial cortex (MC) at a speed of around 100 µm/second.

After insertion of the probe, the brain was covered with Duragel (Cambridge Neurotech), and the craniotomy sealed with sterile vaseline. The microdrive base was fixed to the skull with UV glue and the microdrive was released from the stereotactic holder. Finally, a 3D-printed cap was mounted around the skull. The cap provided mechanical protection, carried the head-stage, as well as colored strips of tape for head-direction and position tracking. A silicone sealant (Kwik-Cast, World Precision Instruments) was applied around the edges of the cap and skin. After surgery, the lizard was released from the stereotactic apparatus and left on a heating pad set to 30 °C until full recovery from anesthesia. On the days after the surgery, probes were slowly lowered into the tissue (up to 280 µm a day).

During recording sessions, lizards were freely moving in a 150 cm diameter circular arena, with two salient opposing visual cues in a brightly lit room. The room was heated to 29 °C, allowing lizards to stay active for an extended duration. Live prey items (mealworms) were dropped into the arena at random times and locations, motivating lizards to explore and maximize arena coverage.

After the end of the experiments, the lizard received intramuscular injections of ketamine and Dormicom, followed by induction and intubation with Isoflurane as described in the surgical method. The 3D printed cap, head-stage, and microdrive carrying the probe were retrieved.

Lizard was then decapitated and immediately perfused through the carotid arteries with PBS, followed by 4% formaldehyde. The brain was removed and stored in 4% formaldehyde for one day, followed by an additional day in a 30% sucrose solution. The brain was frozen with dry ice and cut into 70 μm coronal slices. Slices were mounted on glass slides, and stained with DAPI. The final position of the tip of the probe was identified based on tissue damage and a Dil signal, imaged using a slide-scanner (Zeiss, Axio scan Z.7). The data were batch sorted with Kilosort 2.0 using a 96 site template, and units were selected with an automated quality metric within Kilosort 2.0 like recordings from other species.

Recordings from electric fish

All experiments adhered to the American Physiological Society's Guiding Principles in the Care and Use of Animals and were approved by the Institutional Animal Care and Use Committee of Columbia University. Weakly electric mormyrid fish (7-12 cm in length) of the species *Gnathonemus petersii* were used for recordings. Fish were housed in 60 gallon tanks in groups of 5-20. Water conductivity was maintained between 65-100 microsiemens both in the fish's home tanks and during experiments. For surgery to expose the brain for recording, fish were anesthetized (MS:222, 1:25,000) and held against a foam pad. Skin on the dorsal surface of the head was removed and a long-lasting local anesthetic (0.75% Bupivacaine) was applied to the wound margins. A plastic rod was attached to the skull with Metabond (Parkell) to secure the head and a craniotomy was performed over the C1 region of the cerebellum. Gallamine triethiodide (Flaxedil) was given at the end of the surgery ($\sim 20 \mu\text{g}/\text{cm}$ of body length) to immobilize the fish and fresh aerated water was passed over the fish's gills for respiration. The rate of the electric organ discharge motor command was monitored continuously by electrodes positioned near the electric organ in the tail. Electrosensory stimuli were delivered (0.2 ms duration square pulses) between an electrode in the stomach and another positioned near the tail. Probes were inserted vertically into the cerebellum along the midline and lowered slowly ($\sim 10 \mu\text{m}/\text{s}$) to a final depth of $\sim 3.5 \text{ mm}$. The data were batch sorted with Kilosort 2.0 using a 96 site template, and units were selected with an automated quality metric within Kilosort 2.0 like recordings from other species.

Optotagging

Optotagging experiments

For optotagging experiments, four different transgenic mouse lines were used: 1.) Sst-IRES-Cre;Ai32, 2) Vip-IRES-Cre;Ai32, 3) Pvalb-IRES-Cre;Ai32, and 4) Sim1-Cre;Ai32. Ai32 is a reporter line that expresses channelrhodopsin in Cre⁺ cells. In each experiment we made recordings in parallel from 2-3 NP Ultra probes mounted on separate NewScale micromanipulators. Electrophysiological signals were acquired with OpenEphys software as described previously (Siegle et al., 2021). During the experiment, we recorded sequentially at 3-4 depths within the cortex. Probes were first inserted superficially in the cortex ($\sim 200 \mu\text{m}$ deep) and allowed to settle for 10 minutes. We then ran the optogenetic protocol (described

below). Next, probes were inserted deeper into the cortex (200-300 μm), allowed an additional 10 minutes to settle, and the optogenetic protocol was repeated. This process was repeated for up to 4 probe insertion depths. Recordings were made on a single day or on consecutive days from the same mouse.

We used a 470 nm laser to stimulate and optotag neurons expressing Cre and channelrhodopsin in each transgenic mouse line and experiment. For interneuron photostimulation, lasers were coupled to a 200 μm optic fiber. For Sim1-Cre;Ai32 L5 pyramidal neuron photostimulation, the laser path was guided by a pair of galvo-coupled mirrors to focus light to a precise point on the brain surface to minimize recurrent excitation. We used two different photostimulation patterns: a 10 ms square-wave pulse and a 1 s raised cosine ramp. On each trial, light was delivered at one of three light intensities levels (0.2, 4.1, and 10.0 mW/mm^2). These six photostimulation inputs were randomly interleaved, and each was repeated for 100 trials. The average inter-trial interval was 2 s. In total, the photostimulation protocol lasted ~20 minutes and recordings were conducted over a total of ~30 minutes.

Identification of opto-tagged units

First-pass identification of opto-tagged units used an algorithmic approach to find units of similarly shaped peri-stimulus time histograms (PSTH) during a 1 second raised cosine ramp stimulation at the highest light intensity (see above). This method did not disambiguate between units that were directly stimulated and those that showed an indirect increase in firing rate. Identified units were used as a baseline comparator for further classification.

Unsupervised identification of opto-tagged units in each transgenic mouse line was performed similarly to (Jia et al., 2019). First, the PSTHs for all stimulation patterns within a given photostimulation intensity were averaged across all trials, then the average PSTH for each pattern and intensity was concatenated, forming a neurons x response-vector matrix. This matrix was then normalized and PCA was applied to reduce the dimensionality of the dataset. All units were then visualized in a low-dimensional space via Uniform Manifold Approximation and Projection (UMAP) using all timeseries PCs. Doing so allowed us to determine whether units with a similar PSTH structure aggregated to the same regions (**Supp. Fig. S24**).

Clustering of units was performed using Density-based Spatial Clustering of Applications with Noise (DBSCAN). We chose this method of clustering units as it does not require an *a priori* knowledge of the number of clusters expected and is robust against clusters of varying shapes and sizes, like those produced by UMAP. Hyperparameters for DBSCAN clustering were specific to experiments performed within each transgenic line and were obtained using K-Nearest Neighbors (KNN) to estimate the distance of each data point to the next closest point, then expanding this distance to include the 10th NN. The inflection point in the distribution of these distance values was used as the cluster radius (ϵ) for DBSCAN. The resulting clusters of units were then compared against those found in our first-pass sorting approach and validated by ensuring a short latency response during a 10 ms square wave pulse photostimulation.

Supervised waveform classification of inhibitory neurons

Feature	Abbreviation	Definition	Units
Amplitude	Amp.	The absolute difference in voltage between the spike minimum (trough) and the post-trough maximum (peak).	μV
Duration	Dur.	The time delay between the trough and post-trough peak.	ms
Pre-peak-to-trough ratio	prePTR or pPTR	The ratio between the maximum voltage prior to the trough and the trough.	ratio
Peak-to-trough ratio	PTR	The ratio between the maximum voltage after the trough and the trough.	ratio
Repolarization slope	Repol. slope	The slope of the linear fit between the trough and 0.3 ms after the trough.	mV/ms
Recovery slope	Recov. slope	The slope of the linear fit between the post-trough peak and 0.3 ms after the peak.	mV/ms
Average inter-spike-interval	Avg. ISI	The average time between spikes calculated across the recording, excluding epochs of photostimulation.	s
Baseline firing rate	Baseline FR	The average firing rate 1 s prior to photostimulation epochs.	spks/s
Spatial footprint	Footprint	The radial distance about the peak amplitude channel within which the average amplitude across channels $\leq 30 \mu\text{V}$.	μm

Table 4. Features used for cell type classification.

Linear Discriminant Analysis

Linear Discriminant Analysis (LDA, **Fig. 7H**) was performed via its implementation in SciKit-Learn (Python version 3.9.12). Briefly, pairwise comparisons between unit classes proceeded first by randomly subsampling each dataset, without replacement, to the lowest number of units among the unit classes to be compared (PV, $n = 238$). Only prePTR and PTR single-channel 1D features were used in this analysis. Bootstrapping of the LDA model to subsamples of each unit class over 100 iterations with 100 random initial states and each iteration cross-validated five-fold. Performance of the model was quantified as how accurately the model classified each unit, given a label matrix. Accuracy scores were averaged across all iterations for each comparison to give a mean classification accuracy score and compared against chance levels of classifying a unit correctly (two classes per comparison, chance = 0.5).

Cross-correlogram analysis for putatively monosynaptically connected neurons

Cross-correlograms were computed between each pair of neurons that included one narrow-spiking small footprint (NS_s) and one other ($n > 125,000$ pairs, $n = 40$ recordings) with a 1 ms bin size using all spikes from the recording. Excitatory interactions were identified as those in which the post-reference spike firing rate period increased to greater than five times the pre-spike standard deviation, whereas inhibitory interactions were those in which the firing rate decreased by at least the same amount.

Random Forest Classification

Validated opto-tagged units and untagged units formed the basis of a label matrix that was then used to classify neuronal types using random forest classification. Because significant differences could be found between RS, NS_L, and NS_s groups (**Fig. 7E-G, 8D, Supp. Fig. S27**), these units were included as distinct categories during classification.

Random forest classification was performed via its implementation in SciKit-Learn (Python version 3.9.12) using three different feature sets computed independently on mean electrophysiological waveforms gathered from NP Ultra probes and interpolated, “NP 1.0-like” unit-matched data. These feature sets were 1) 1-dimensional scalar values computed from the peak amplitude channel (**Fig. 8C, Table 4**) 2) the peak amplitude waveform and 3) All 1D scalar features and spatial footprint.

The process of random forest classification proceeded as the following: NP Ultra data from the entire population of recorded units was spatially resampled to a NP 1.0-like geometry (**Fig. 2E**). Units for which the peak channel amplitude dropped below 50 μ V were then removed from both datasets for the purposes of this analysis. Data (NP Ultra and NP 1.0-like) from each unit class category was then randomly subsampled, without replacement, to the lowest number of units among optotagged cell types (SST, 116 units) and pooled. This process was repeated through 100 iterations using 100 initial states. Principal components analysis was then performed on the pooled data for each probe type prior to classification and classifier hyperparameters were then optimized via grid search using five-fold cross validation. Classification proceeded via five-fold cross-validation where the classifier was trained on 80% of the input data and performance was evaluated using held-out test data (20%). Classification performance was evaluated as the prediction accuracy of the classifier on left-out data over each bootstrapped iteration. Confusion matrices were computed as the comparison of predicted classes to true classes for each subsampled dataset under 100 random initial states.

Multimodal classification of inhibitory neurons

To improve cell-type classification of the optotagged units, a recently developed multimodal cell-type classification method NEMO (Neuronal Embeddings via Multimodal contrastive

learning) was utilized (Yu et al., 2024). NEMO is a multimodal contrastive method based on CLIP (Radford et al., 2021) that jointly embeds the activity autocorrelations and average extracellular waveforms (templates) of each neuron. By finding a joint representation of different modalities, NEMO is able to extract intrinsic properties of a neuron that are predictive of its cell-type.

The activity autocorrelations were constructed by creating an autocorrelogram (ACG) image for each neuron. This involved smoothing the spiking activity using a 250-ms boxcar filter, segmenting the firing rate distribution into 10 deciles, and constructing ACGs for each decile. The resulting ACG image effectively represents how a neuron's autocorrelations vary with its firing rate. By generating ACGs for each decile, this approach captures firing rate-dependent changes in autocorrelations (Beau et al., 2024). To eliminate stimulus-induced effects, we excluded spikes occurring within 1100 ms of each stimulus onset and computed ACG images using only the remaining data. The average waveforms were constructed from approximately 500 recorded waveforms for each neuron. We utilize multi-channel templates which take advantage of the detailed spatial structure enabled by the small channel spacing. We use nine channels with the highest peak-to-peak (ptp) amplitude, re-ordered from highest to lowest amplitude.

NEMO was trained on the activity autocorrelations and average waveforms for all the *untagged* units (8491 neurons) in a fully self-supervised manner. NEMO was then fine-tuned to classify the *optotagged* units. NEMO was evaluated using 5-fold cross-validation with 10 repeats. By pre-training on untagged units and fine-tuning on the optotagged units, NEMO was able to reach a .80 F1 score and balanced accuracy on classifying inhibitory cell-types (the confusion matrix is shown in **Fig. 8H**).

References

- Allen WE, Chen MZ, Pichamoorthy N, Tien RH, Pachitariu M, Luo L, Deisseroth K (2019) Thirst regulates motivated behavior through modulation of brainwide neural population dynamics. *Science* 364.
- Barthó P (2021) Extracellular recording of axonal spikes in the visual cortex. *J Physiol* 599:2131–2131.
- Barthó P, Hirase H, Monconduit L, Zugaro M, Harris KD, Buzsáki G (2004) Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. *J Neurophysiol* 92:600–608.
- Barthó P, Slézia A, Mátyás F, Faradzs-Zade L, Ulbert I, Harris KD, Acsády L (2014) Ongoing Network State Controls the Length of Sleep Spindles via Inhibitory Activity. *Neuron* 82:1367–1379.
- Beau M et al. (2024) A deep-learning strategy to identify cell types across species from high-density extracellular recordings. Available at:

- <http://biorxiv.org/lookup/doi/10.1101/2024.01.30.577845> [Accessed March 26, 2024].
- Beaulieu-Laroche L, Brown NJ, Hansen M, Toloza EHS, Sharma J, Williams ZM, Frosch MP, Cosgrove GR, Cash SS, Harnett MT (2021) Allometric rules for mammalian cortical layer 5 neuron biophysics. *Nature* 600:274–278.
- Beebe NL, Young JW, Mellott JG, Schofield BR (2016) Extracellular Molecular Markers and Soma Size of Inhibitory Neurons: Evidence for Four Subtypes of GABAergic Cells in the Inferior Colliculus. *J Neurosci* 36:3988–3999.
- Boussard J, Varol E, Lee HD, Dethe N, Paninski L (2021) Three-dimensional spike localization and improved motion correction for Neuropixels recordings. *Neuroscience*. Available at: <http://biorxiv.org/lookup/doi/10.1101/2021.11.05.467503> [Accessed May 18, 2023].
- Boussard J, Windolf C, Hurwitz C, Lee HD, Yu H, Winter O, Paninski L (2023) DARTsort: A modular drift tracking spike sorter for high-density multi-electrode probes. :2023.08.11.553023 Available at: <https://www.biorxiv.org/content/10.1101/2023.08.11.553023v1> [Accessed August 15, 2023].
- Buccino AP, Kordovan M, Ness TV, Merkt B, Häfliger PD, Fyhn M, Cauwenberghs G, Rotter S, Einevoll GT (2018) Combining biophysical modeling and deep learning for multielectrode array neuron localization and classification. *J Neurophysiol* 120:1212–1232.
- Buzsáki G (2006) *Rhythms of the Brain*. Oxford university press.
- Buzsáki G, Anastassiou CA, Koch C (2012) The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nat Rev Neurosci* 13:407–420.
- Buzsáki G, Kandel A (1998) Somadendritic Backpropagation of Action Potentials in Cortical Pyramidal Cells of the Awake Rat. *J Neurophysiol* 79:1587–1591.
- Carnevale NT, Hines ML (2006) *The NEURON Book*. Cambridge: Cambridge University Press. Available at: <https://www.cambridge.org/core/books/neuron-book/7C8D9BD861D288E658BEB652F593F273> [Accessed April 2, 2024].
- Chen S, Liu Y, Wang Z, Colonell J, Liu LD, Hou H, Tien N-W, Wang T, Harris T, Druckmann S (2023) Brain-wide neural activity underlying memory-guided movement. *BioRxiv*:2023–03.
- Chen S, Svoboda K (2020) “Uniclear” water-based brain clearing for light sheet imaging of electrode tracks. Available at: <https://www.protocols.io/view/39-unclear-39-water-based-brain-clearing-for-lig-zndf5a6> [Accessed August 22, 2023].
- Chung JE, Sellers KK, Leonard MK, Gwilliams L, Xu D, Dougherty ME, Kharazia V, Metzger SL, Welkenhuysen M, Dutta B, Chang EF (2022) High-density single-unit human cortical recordings using the Neuropixels probe. *Neuron* 110:2409–2421.e3.
- Cooper GF, Robson JG, Waldron I (1969) The action potentials recorded from undamaged

- nerve fibres with micro-electrodes. *J Physiol* 200:9P-11P.
- Crawford E, Pineau J (2019) Spatially Invariant Unsupervised Object Detection with Convolutional Neural Networks. *Proc AAAI Conf Artif Intell* 33:3412–3420.
- Deligkaris K, Bullmann T, Frey U (2016) Extracellularly Recorded Somatic and Neuritic Signal Shapes and Classification Algorithms for High-Density Microelectrode Array Electrophysiology. *Front Neurosci* 10 Available at: <https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2016.00421/full> [Accessed March 27, 2024].
- Elabbady L et al. (2024) Perisomatic Features Enable Efficient and Dataset Wide Cell-Type Classifications Across Large-Scale Electron Microscopy Volumes. :2022.07.20.499976 Available at: <https://www.biorxiv.org/content/10.1101/2022.07.20.499976v2> [Accessed April 7, 2025].
- Fiáth R, Raducanu BC, Musa S, Andrei A, Lopez CM, Welkenhuysen M, Ruther P, Aarts A, Ulbert I (2018) Fine-scale mapping of cortical laminar activity during sleep slow oscillations using high-density linear silicon probes. *J Neurosci Methods*:0–1.
- Gardner RJ, Hermansen E, Pachitariu M, Burak Y, Baas NA, Dunn BA, Moser M-B, Moser EI (2022) Toroidal topology of population activity in grid cells. *Nature* 602:123–128.
- Gilman JP, Medalla M, Luebke JI (2017) Area-Specific Features of Pyramidal Neurons—a Comparative Study in Mouse and Rhesus Monkey. *Cereb Cortex* 27:2078–2094.
- Gold C, Henze DA, Koch C, Buzsáki G (2006) On the origin of the extracellular action potential waveform: A modeling study. *J Neurophysiol* 95:3113–3128.
- Goldberg JH, Fee MS (2012) A cortical motor nucleus drives the basal ganglia-recipient thalamus in singing birds. *Nat Neurosci* 15:620–627.
- Gouwens NW et al. (2020) Integrated Morphoelectric and Transcriptomic Classification of Cortical GABAergic Cells. *Cell* 183:935-953.e19.
- Hagen E, Næss S, Ness TV, Einevoll GT (2018) Multimodal Modeling of Neural Network Activity: Computing LFP, ECoG, EEG, and MEG Signals With LFPy 2.0. *Front Neuroinformatics* 12 Available at: <https://www.frontiersin.org/articles/10.3389/fninf.2018.00092> [Accessed April 2, 2024].
- Hallermann S, de Kock CPJ, Stuart GJ, Kole MHP (2012) State and location dependence of action potential metabolic cost in cortical pyramidal neurons. *Nat Neurosci* 15:1007–1014.
- Halnes G, Ness T, Naess S, Hagen E, Pettersen K, Einevoll G (2024) Electric brain signals. Cambridge University Press.
- Hay E, Hill S, Schürmann F, Markram H, Segev I (2011) Models of Neocortical Layer 5b Pyramidal Cells Capturing a Wide Range of Dendritic and Perisomatic Active Properties. *PLOS Comput Biol* 7:e1002107.

- Herculano-Houzel S, Manger PR, Kaas JH (2014) Brain scaling in mammalian evolution as a consequence of concerted and mosaic changes in numbers of neurons and average neuronal cell size. *Front Neuroanat* 8 Available at: <https://www.frontiersin.org/articles/10.3389/fnana.2014.00077> [Accessed May 17, 2023].
- Hilgen G, Sorbaro M, Pirmoradian S, Muthmann J-O, Kepiro IE, Ullo S, Ramirez CJ, Encinas AP, Maccione A, Berdondini L (2017) Unsupervised spike sorting for large-scale, high-density multielectrode arrays. *Cell Rep* 18:2521–2532.
- International Brain Laboratory et al. (2023a) Reproducibility of in-vivo electrophysiological measurements in mice. :2022.05.09.491042 Available at: <https://www.biorxiv.org/content/10.1101/2022.05.09.491042v4> [Accessed August 18, 2023].
- International Brain Laboratory, Benson B, Benson J, Birman D, Bonacchi N, Carandini M, Catarino JA, Chapuis GA, Churchland AK, Dan Y (2023b) A Brain-Wide Map of Neural Activity during Complex Behaviour. *bioRxiv*:2023–07.
- Jia X, Siegle J, Bennett C, Gale S, Denman D, Koch C, Olsen S (2019) High-density extracellular probes reveal dendritic backpropagation and facilitate neuron classification. *J Neurophysiol*.
- Jia X, Siegle JH, Durand S, Heller G, Ramirez TK, Koch C, Olsen SR (2022) Multi-regional module-based signal transmission in mouse visual cortex. *Neuron* 110:1585-1598.e9.
- Jun JJ et al. (2017) Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551:232–236.
- Katzner S, Nauhaus I, Benucci A, Bonin V, Ringach DL, Carandini M (2009) Local Origin of Field Potentials in Visual Cortex. *Neuron* 61:35–41.
- Kiani R, Esteky H, Mirpour K, Tanaka K (2007) Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *J Neurophysiol* 97:4296–4309.
- Kingma DP, Welling M (2022) Auto-Encoding Variational Bayes. Available at: <http://arxiv.org/abs/1312.6114> [Accessed August 18, 2023].
- Kvitsiani D, Ranade S, Hangya B, Taniguchi H, Huang JZ, Kepecs A (2013) Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature* 498:363–366.
- Lee EK, Balasubramanian H, Tsolias A, Anakwe SU, Medalla M, Shenoy KV, Chandrasekaran C (2021) Non-linear dimensionality reduction on extracellular waveforms reveals cell type diversity in premotor cortex. *Elife* 10:e67490.
- Lee EK, Gül AE, Heller G, Lakunina A, Jaramillo S, Przytycki PF, Chandrasekaran C (2024) PhysMAP - interpretable *in vivo* neuronal cell type identification using multi-modal analysis of electrophysiological data. Available at: <http://biorxiv.org/lookup/doi/10.1101/2024.02.28.582461> [Accessed March 26, 2024].
- Lee J, Mitelut C, Shokri H, Kinsella I, Dethe N, Wu S, Li K, Reyes EB, Turcu D, Batty E, Kim YJ,

- Brackbill N, Kling A, Goetz G, Chichilnisky EJ, Carlson D, Paninski L (2020) YASS: Yet Another Spike Sorter applied to large-scale multi-electrode array recordings in primate retina. *Neuroscience*. Available at: <http://biorxiv.org/lookup/doi/10.1101/2020.03.18.997924> [Accessed May 17, 2023].
- Lemon RN, Baker SN, Kraskov A (2021) Classification of Cortical Neurons by Spike Shape and the Identification of Pyramidal Neurons. *Cereb Cortex* 31:5131–5138.
- Lima SQ, Hromádka T, Znamenskiy P, Zador AM (2009) PINP: A New Method of Tagging Neuronal Populations for Identification during In Vivo Electrophysiological Recording. *PLOS ONE* 4:e6099.
- López CM, Welkenhuysen M, Musa S, Eberle W, Bartic C, Puers R, Gielen G (2012) Towards a noise prediction model for in vivo neural recording. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp 759–762. IEEE.
- Lopez CM, Welkenhuysen M, Musa S, Eberle W, Bartic C, Puers R, Gielen G (2012) Towards a noise prediction model for in vivo neural recording. *Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS*:759–762.
- Malzer C, Baum M (2019) A Hybrid Approach To Hierarchical Density-based Cluster Selection. Available at: <https://arxiv.org/abs/1911.02282> [Accessed May 17, 2023].
- Mazer, Jamie (2013) Pype2. Available at: <https://github.com/mazerj/pype3>.
- McColgan T, Liu J, Kuokkanen PT, Carr CE, Wagner H, Kempter R (2017) Dipolar extracellular potentials generated by axonal projections van Rossum MC, ed. *eLife* 6:e26106.
- McCormick DA, Connors BW, Lighthall JW, Prince DA (1985) Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *J Neurophysiol* 54:782–806.
- McHugh SB, Lopes-dos-Santos V, Gava GP, Hartwich K, Tam SKE, Bannerman DM, Dupret D (2022) Adult-born dentate granule cells promote hippocampal population sparsity. *Nat Neurosci* 25:1481–1491.
- McLelland D, Baker PM, Ahmed B, Kohn A, Bair W (2015) Mechanisms for Rapid Adaptive Control of Motion Processing in Macaque Visual Cortex. *J Neurosci* 35:10268–10280.
- Merrill EG, Wall PD, Yaksh TL (1978) Properties of two unmyelinated fibre tracts of the central nervous system: lateral Lissauer tract, and parallel fibres of the cerebellum. *J Physiol* 284:127–145.
- Metzen MG, Chacron MJ (2021) Population Coding of Natural Electrosensory Stimuli by Midbrain Neurons. *J Neurosci* 41:3822–3841.
- Mitchell JF, Sundberg KA, Reynolds JH (2007) Differential Attention-Dependent Response Modulation across Cell Classes in Macaque Visual Area V4. *Neuron* 55:131–141.
- Musall S, Kaufman MT, Juavinett AL, Gluf S, Churchland AK (2019) Single-trial neural dynamics

- are dominated by richly varied movements. *Nat Neurosci* 22:1677–1686.
- Nelson MJ, Pouget P (2010) Do Electrode Properties Create a Problem in Interpreting Local Field Potential Recordings? *J Neurophysiol* 103:2315–2317.
- Neubrandt M, Oláh VJ, Brunner J, Marosi EL, Soltesz I, Szabadics J (2018) Single Bursts of Individual Granule Cells Functionally Rearrange Feedforward Inhibition. *J Neurosci* 38:1711–1724.
- Niell CM, Stryker MP (2008) Highly Selective Receptive Fields in Mouse Visual Cortex. *J Neurosci* 28:7520–7536.
- Norimoto H, Fenk LA, Li H-H, Tosches MA, Gallego-Flores T, Hain D, Reiter S, Kobayashi R, Macias A, Arends A, Klinkmann M, Laurent G (2020) A claustrum in reptiles and its role in slow-wave sleep. *Nature* 578:413–418.
- Nowak LG, Azouz R, Sanchez-Vives MV, Gray CM, McCormick DA (2003) Electrophysiological Classes of Cat Primary Visual Cortical Neurons In Vivo as Revealed by Quantitative Analyses. *J Neurophysiol* 89:1541–1566.
- Nunez PL, Srinivasan R (2006) *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA.
- Paulk AC, Kfir Y, Khanna AR, Mustroph ML, Trautmann EM, Soper DJ, Stavisky SD, Welkenhuysen M, Dutta B, Shenoy KV, Hochberg LR, Richardson RM, Williams ZM, Cash SS (2022) Large-scale neural recordings with single neuron resolution using Neuropixels probes in human cortex. *Nat Neurosci* 25:252–263.
- Pedroni A, Minh DD, Mallamaci A, Cherubini E (2014) Electrophysiological characterization of granule cells in the dentate gyrus immediately after birth. *Front Cell Neurosci* 8 Available at: <http://journal.frontiersin.org/article/10.3389/fncel.2014.00044/abstract> [Accessed July 17, 2023].
- Petersen PC, Siegle JH, Steinmetz NA, Mahallati S, Buzsáki G (2021) CellExplorer: A framework for visualizing and characterizing single neurons. *Neuron* 109:3594–3608.e2.
- Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Aspell A, Mishkin P, Clark J, Krueger G, Sutskever I (2021) Learning Transferable Visual Models From Natural Language Supervision. Available at: <http://arxiv.org/abs/2103.00020> [Accessed March 25, 2025].
- Reifenstein ET, Ebbesen CL, Tang Q, Brecht M, Schreiber S, Kempter R (2016) Cell-Type Specific Phase Precession in Layer II of the Medial Entorhinal Cortex. *J Neurosci* 36:2283–2288.
- Rey HG, Pedreira C, Quiroga R (2015) Past, present and future of spike sorting techniques. *Brain Res Bull* 119:106–117.
- Robbins A, Fox S, Holmes G, Scott R, Barry J (2013) Short duration waveforms recorded extracellularly from freely moving rats are representative of axonal activity. *Front Neural Circuits* 7 Available at: <https://www.frontiersin.org/articles/10.3389/fncir.2013.00181>

[Accessed August 17, 2023].

- Robinson DA (1968) The electrical properties of metal microelectrodes. *Proc IEEE* 56:1065–1071.
- Roux L, Stark E, Sjulson L, Buzsáki G (2014) In vivo optogenetic identification and manipulation of GABAergic interneuron subtypes. *Curr Opin Neurobiol* 26:88–95.
- Schröder S, Steinmetz NA, Krumin M, Pachitariu M, Rizzi M, Lagnado L, Harris KD, Carandini M (2020) Arousal Modulates Retinal Output. *Neuron* 107:487–495.e9.
- Senzai Y, Buzsáki G (2017) Physiological Properties and Behavioral Correlates of Hippocampal Granule Cells and Mossy Cells. *Neuron* 93:691–704.e5.
- Shoham S, O'Connor DH, Segev R (2006) How silent is the brain: is there a “dark matter” problem in neuroscience? *J Comp Physiol A* 192:777–784.
- Sibille J, Gehr C, Benichov JI, Balasubramanian H, Teh KL, Lupashina T, Vallentin D, Kremkow J (2022) High-density electrode recordings reveal strong and specific connections between retinal ganglion cells and midbrain neurons. *Nat Commun* 13:5218.
- Siegle JH, Jia X, Durand S, Gale S, Bennett C, Graddis N, Heller G, Ramirez TK, Choi H, Luviano JA (2021) Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* 592:86–92.
- Someck S, Levi A, Sloin HE, Spivak L, Gattegno R, Stark E (2023) Positive and biphasic extracellular waveforms correspond to return currents and axonal spikes. *Commun Biol* 6:1–15.
- Steinmetz NA et al. (2021) Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science* 372 Available at: <https://science.sciencemag.org/content/372/6539/eabf4588> [Accessed July 20, 2021].
- Steinmetz NA, Koch C, Harris KD, Carandini M (2018) Challenges and opportunities for large-scale electrophysiology with Neuropixels probes. *Curr Opin Neurobiol* 50:92–100.
- Steinmetz NA, Zatka-Haas P, Carandini M, Harris KD (2019) Distributed coding of choice, action and engagement across the mouse brain. *Nature* 576:266–273.
- Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD (2019) Spontaneous behaviors drive multidimensional, brainwide activity. *Science* 364.
- Stuart G, Schiller J, Sakmann B (1997) Action potential initiation and propagation in rat neocortical pyramidal neurons. *J Physiol* 505:617–632.
- Stuart GJ, Spruston N (2015) Dendritic integration: 60 years of progress. *Nat Neurosci* 18:1713–1721.
- Sukhum KV, Shen J, Carlson BA (2018) Extreme Enlargement of the Cerebellum in a Clade of Teleost Fishes that Evolved a Novel Active Sensory System. *Curr Biol* 28:3857–3863.e3.
- Sun SH, Almasi A, Yunzab M, Zehra S, Hicks DG, Kameneva T, Ibbotson MR, Meffin H (2021)

- Analysis of extracellular spike waveforms and associated receptive fields of neurons in cat primary visual cortex. *J Physiol* 599:2211–2238.
- Swindale NV, Spacek MA (2014) Spike sorting for polytrodes: a divide and conquer approach. *Front Syst Neurosci* 8 Available at: <http://journal.frontiersin.org/article/10.3389/fnsys.2014.00006/abstract> [Accessed August 8, 2023].
- Tosches MA, Yamawaki TM, Naumann RK, Jacobi AA, Tushev G, Laurent G (2018) Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. *Science* 360:881–888.
- Trautmann EM et al. (2023) Large-scale brain-wide neural recording in nonhuman primates. :2023.02.01.526664 Available at: <https://www.biorxiv.org/content/10.1101/2023.02.01.526664v2> [Accessed February 8, 2023].
- Trautmann EM, Stavisky SD, Lahiri S, Ames KC, Kaufman MT, O'Shea DJ, Vyas S, Sun X, Ryu SI, Ganguli S (2019) Accurate estimation of neural population dynamics without spike sorting. *Neuron* 103:292–308.
- Tremblay R, Lee S, Rudy B (2016) GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits. *Neuron* 91:260–292.
- Tymochko S, Munch E, Dunion J, Corbosiero K, Torn R (2020) Using persistent homology to quantify a diurnal cycle in hurricanes. *Pattern Recognit Lett* 133:137–143.
- Vesuna S, Kauvar IV, Richman E, Gore F, Oskotsky T, Sava-Segal C, Luo L, Malenka RC, Henderson JM, Nuyujukian P (2020) Deep posteromedial cortical rhythm in dissociation. *Nature* 586:87–94.
- Viswam V, Obien M, Frey U, Franke F, Hierlemann A (2017) Acquisition of bioelectrical signals with small electrodes. In: 2017 IEEE Biomedical Circuits and Systems Conference (BioCAS), pp 1–4. IEEE.
- Windolf C et al. (2025) DREDge: robust motion correction for high-density extracellular recordings across species. *Nat Methods*:1–13.
- Yamin HG, Stern EA, Cohen D (2013) Parallel Processing of Environmental Recognition and Locomotion in the Mouse Striatum. *J Neurosci* 33:473–484.
- Yang Z, Zhao Q, Keefer E, Liu W (2009) Noise characterization, modeling, and reduction for in vivo neural recording. *Adv Neural Inf Process Syst* 22.
- Yao Z et al. (2021) A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* 184:3222–3241.e26.
- Yu H, Lyu H, Xu Y, Windolf C, Lee EK, Yang F, Shelton AM, Winter O, Laboratory IB, Dyer EL, Chandrasekaran C, Steinmetz NA, Paninski L, Hurwitz CL (2024) In vivo cell-type and brain region classification via multimodal contrastive learning. Available at: <https://openreview.net/forum?id=10JOIFIPjt> [Accessed March 25, 2025].

- Yu J, Hu H, Agmon A, Svoboda K (2019) Recruitment of GABAergic Interneurons in the Barrel Cortex during Active Tactile Behavior. *Neuron* 104:412-427.e4.
- Zhang Y, He T, Boussard J, Windolf C, Winter O, Trautmann E, Roth N, Barrell H, Churchland M, Steinmetz NA (2024) Bypassing spike sorting: Density-based decoding using spike localization from dense multielectrode probes. *Adv Neural Inf Process Syst* 36 Available at: https://proceedings.neurips.cc/paper_files/paper/2023/hash/f499387f191d6be56e68966181095878-Abstract-Conference.html [Accessed April 9, 2024].

Acknowledgements

We thank NIH program officers Ned Talley, Sahana Kukke, and Michele Pucak for their support, and we thank Tim Gardner and Cindy Chestek for their advice and guidance. We thank Bill Karsh for SpikeGLX software modifications enabling use of NP Ultra probes.

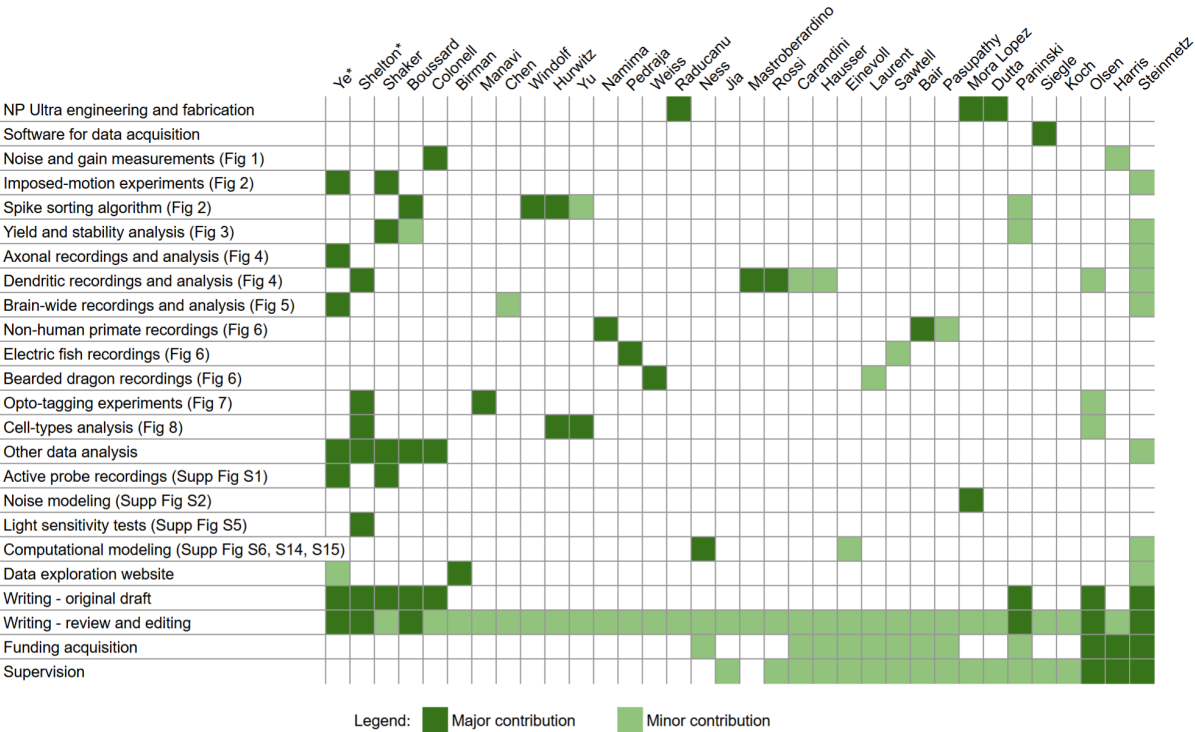
Funding

This research program was funded by the NIH BRAIN Initiative (U01NS113252 to NAS, SRO, and TDH). Additional support was provided by the Pew Biomedical Scholars Program (NAS), the Klingenstein-Simons Fellowship in Neuroscience (NAS), the Max Planck Society (GL), the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement No 834446 to GL and AdG 695709 to MH), the NIH (R01 NS118448 & R01 NS075023 to NBS), the Wellcome Trust (PRF 201225 and 224688 to MH, SHWF 221674 to LFR, collaborative award 204915 to MC, MH and TDH), the Giovanni Armenise Harvard Foundation (CDA to LFR), the Human Technopole (ECF to LFR), and the NSF (IOS 211500 to NBS). GM is supported by a Boehringer Ingelheim Fonds PhD Fellowship. The primate research procedures were supported by the NIH P51 (OD010425) to the WaNPRC, and animal breeding was supported by NIH U42 (OD011123). Computational modeling work was supported by the European Union Horizon 2020 Research and Innovation Programme under Grant Agreement No. 945539 Human Brain Project SGA3 and No. 101147319 EBRAINS 2.0 (GTE and TVN).

Author contributions

ZY collected brain-wide acute mouse recordings and pharmacology experiments. AMS and SM collected data with opto-tagging. ZY and JRS collected data with imposed probe motion and visual stimuli. JC performed noise and gain measurements. AMS performed light sensitivity measurements. SC collected acute mouse recordings. JHS wrote software for data acquisition. TN and WB collected non-human primate recordings. FP and NBS collected recordings from electric fish. SW collected recordings from bearded dragon lizards. LFR and GM collected and analyzed data on dendritic back-propagation. JB, CW, CH, and LP developed the spike sorting methods. HY and CH performed cell classification analyses with NEMO. ZY, AMS, JC, JB, JRS,

and NAS analyzed data. DB developed the data exploration website. CML, BR, and BD performed engineering and fabrication of NP Ultra probes. TVN wrote the computational model and NAS analyzed the output. ZY, AMS, JC, JB, JRS, LP, SRO, and NAS wrote the original draft manuscript. All authors reviewed and edited the manuscript. SRO, TDH, and NAS obtained primary funding. XJ, MC, MH, LFR, GTE, GL, NBS, WB, AP, CML, BD, LP, JHS, CK, SRO, TDH, and NAS supervised aspects of the work.



Declaration of interests

CK holds an executive position, and has a financial interest, in Intrinsic Powers, Inc., a company whose purpose is to develop a device that can be used in the clinic to assess the presence and absence of consciousness in patients. This does not pose any conflict of interest with regard to the work undertaken for this publication. BR, CML, and BD are employees of IMEC vzw, a nonprofit research institute that manufactures, sells, and distributes the Neuropixels probes, at cost, to the research community. IMEC vzw holds patents US10811542B2, US10044325B2, and US9384990B2 related to the Neuropixels 1.0 technology that is built upon in this work. MC, MH, and LP are members of Neuron's Advisory Board. All other authors have no competing interests.

Data and materials availability:

As of the time of writing, NP Ultra probes are available for purchase at <https://www.neuropixels.org/specialty-products> under the name “Neuropixels 1.0 HD” (for “high density”). Future availability is subject to supply constraints.

Spatiotemporal waveforms of 4,666 single units recorded across cortex, hippocampus, thalamus, striatum, and midbrain areas in awake mice, along with spike times are available at <https://doi.org/10.6084/m9.figshare.19493588>; and waveform movies are browse-able at <https://npultra.steinmetzlab.net/>.

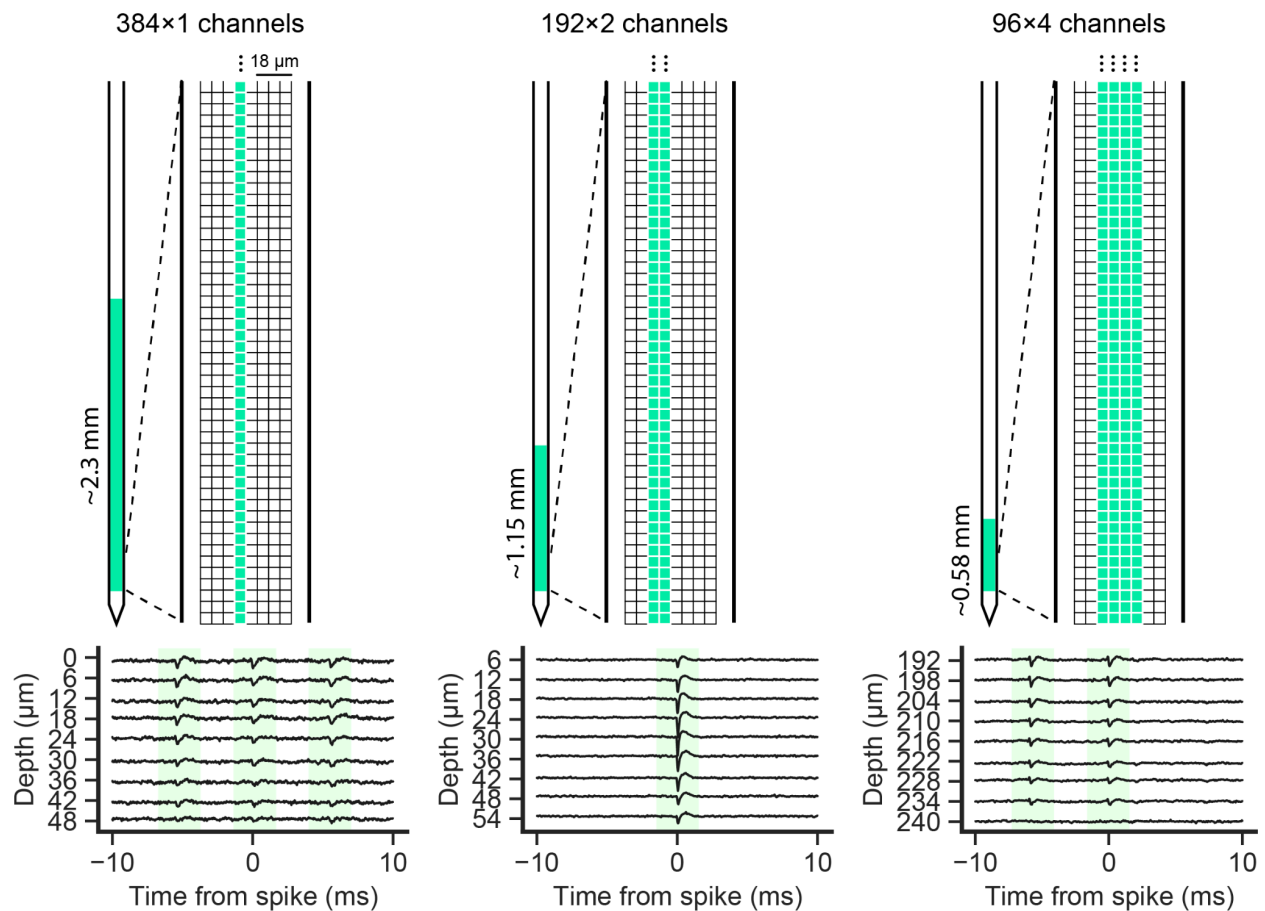
NP Ultra dataset for imposed motion with visual responses pre- and post-motion in the visual cortex is available at: <https://dandiarchive.org/dandiset/000957>.

Code for computing spatial footprints of recorded neurons, in python and in matlab, is provided at: <https://github.com/zhiwen10/Neuropixels-footprint>

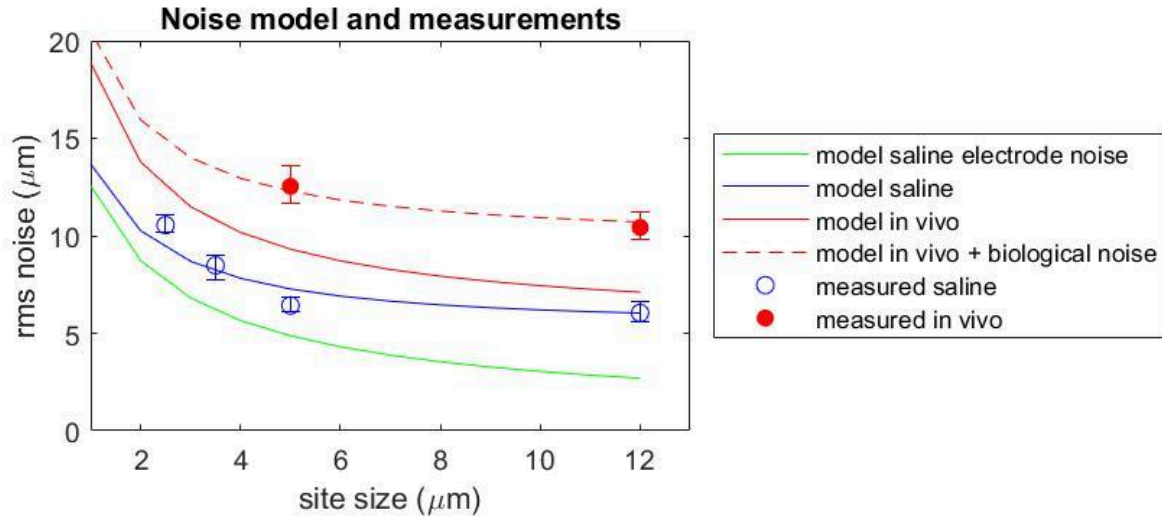
DARTsort code is available at: <https://github.com/cwindolf/dartsort>. The code used for sorting UHD recordings is in the branch `uhd\_np`. In this branch, the script `dartsort.scripts.run_pipeline_uhd_resampled.py` shows how Dartsort was run on UHD and resampled datasets.

Code used for the analysis presented in Figure 2 and the comparisons of different patterns is available at: <https://github.com/julienboussard/uhd-resampled-sorting-analysis>.

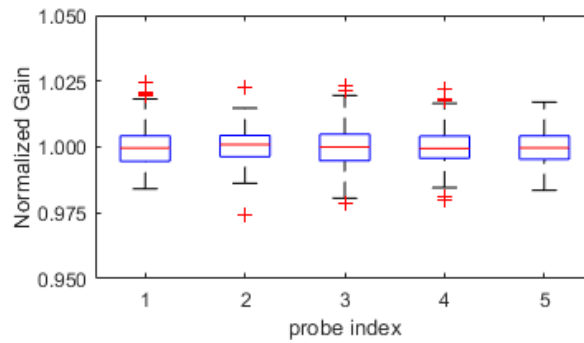
Supplemental Figures



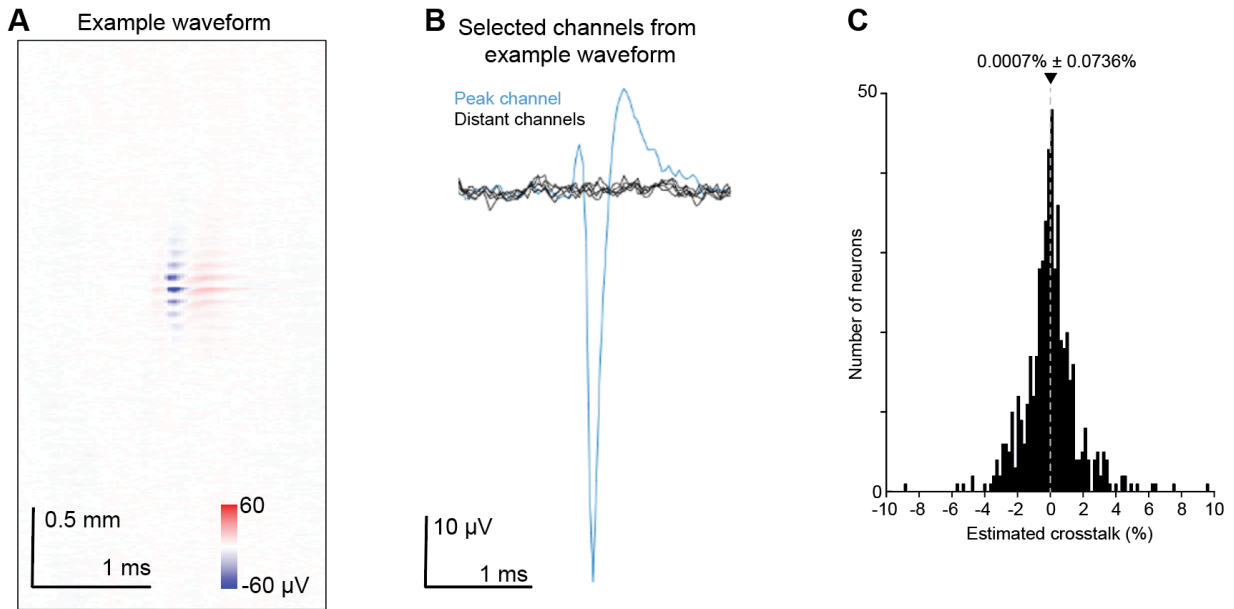
Supplemental Figure S1. Schematics of site layout and raw data traces from three different configurations of the NP Ultra probe. (Top, left to right) 384 x 1 configuration of sites, 192 x 2 sites, and 96 x 4 sites. For both the 384 x 1 and 192 x 2 configurations, not all recorded sites are in the same column, instead there are lateral shifts in the columns of the selected sites (not shown). The selected sites can be moved up and down the length of the probe, chosen from among the total 6144 sites covering ~4.6mm span. (Bottom) Raw data traces 20 ms about a detected spike excerpted from ~5 minute-long recordings in mouse cortex. Detected spikes highlighted in green.



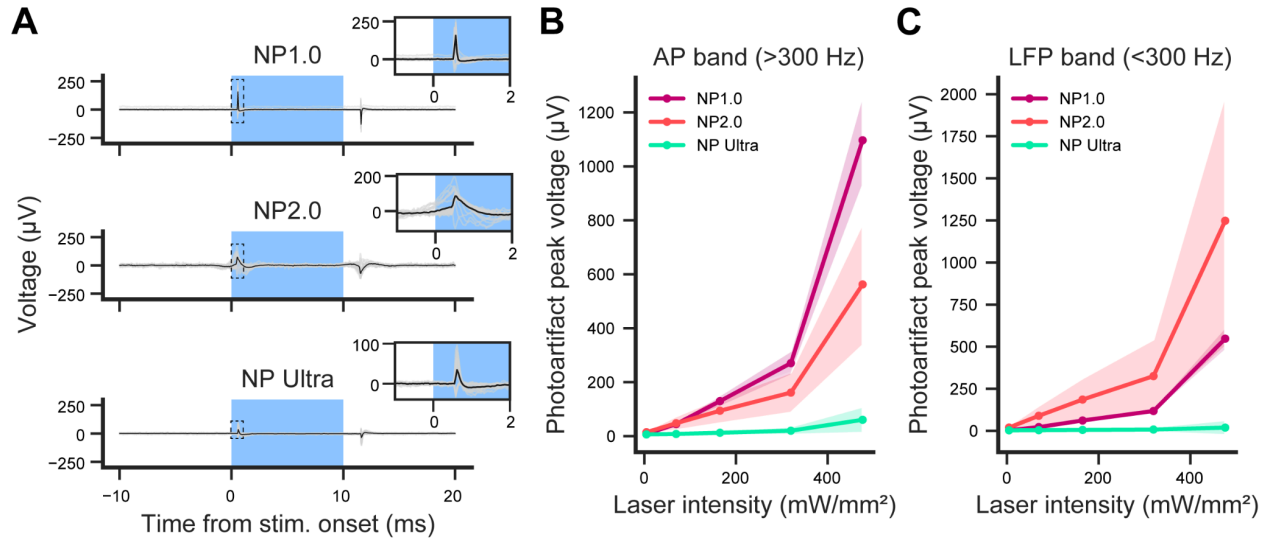
Supplemental Figure S2. Modeled and measured noise as a function of electrode site size. The solid curves show the predicted noise as a function of site size. The model uses estimated impedances given the characteristics of the TiN recording site material (see Methods). The green curve shows the predicted thermal noise at the TiN electrode in saline; the blue curve shows the predicted total noise in saline, which includes the thermal noise and the noise in the Neuropixels electronics. The solid red curve shows the predicted total noise in vivo, accounting for the higher impedance (taken to be 300 Ohm-cm), but not including biological noise; see Lopez, et al., 2012 for details). Data points show experimental measurements of noise in saline and in vivo; error bars show the 25th and 75th percentile of all sites measured. The measured in vivo noise is expected to be higher than the model because it includes biological noise, i.e. small signals from distant units. The dashed red curve shows the predicted in vivo total plus biological noise, fit to the two in vivo data points. The inferred biological noise is 8 μV .



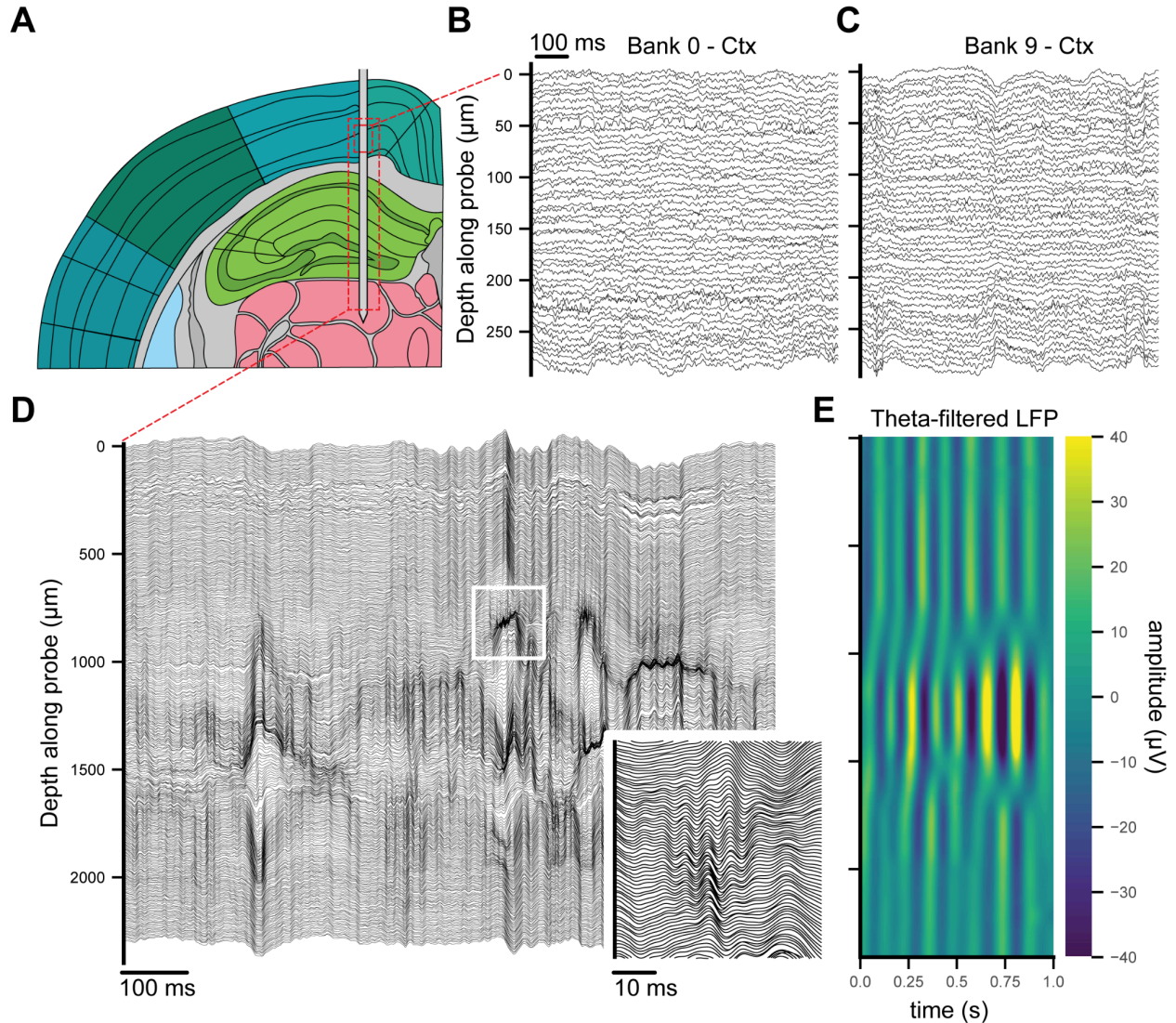
Supplemental Figure S3. Channel-to-channel uniformity in gain in NP Ultra. Peak to peak voltage was measured at each site when applying a 3 kHz, 2mV sine wave to a saline bath for 5 probes. Standard box and whisker plots show distribution of normalized gain measured across the 383 recording sites on the probe (reference site is excluded). Box limits = 25th-75th percentile; whiskers = max/min within 1.5X the 25-75th percentile; points = outliers outside the whisker range. Note that these data show that error intra-site voltage measurements is < 5% for all sites.



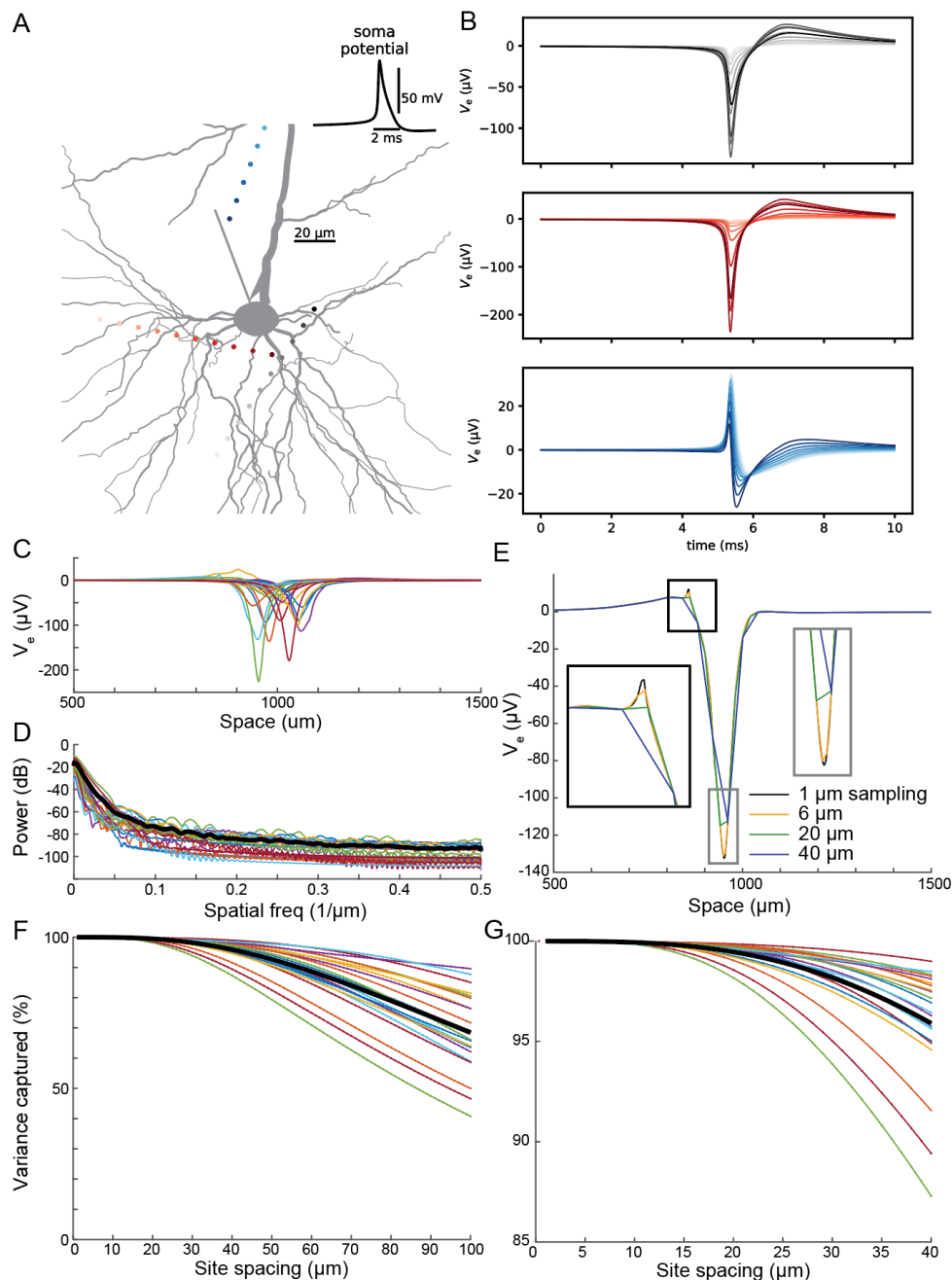
Supplemental Figure S4. Crosstalk analysis of NP Ultra recordings. **A**, An example neuron's average waveform across channels (y-axis) and time (x-axis). **B**, The waveform on the peak channel and on distant channels. For this example neuron, the amplitude at the distant channels was 1.25% of the peak amplitude. **C**, Summary histogram of estimated crosstalk across the $n=519$ neurons included in the analysis. The estimated crosstalk was on average $0.0007\% \pm 0.0736\%$ (mean $\pm$ s.e.m.).



Supplemental Figure S5. Light sensitivity tests. **A**, Example experiment showing the light-induced artifact measured on NP 1.0, NP 2.0, and NP Ultra probes. Gray traces are all channels, black line is the average across channels. Insets focus on the initial positive peak of the photoartifact occurring shortly after light onset. **B**, Peak photoartifacts for each probe type in **A** averaged across trials in the AP band (>300 Hz). **C**, Same as in **B** for LFP band data. The physical explanation for smaller artifacts in NP Ultra probes is unclear, but it is important to note that the NP Ultra probes were the passive, non-switchable version whereas 1.0 and 2.0 probes included CMOS switching circuitry.

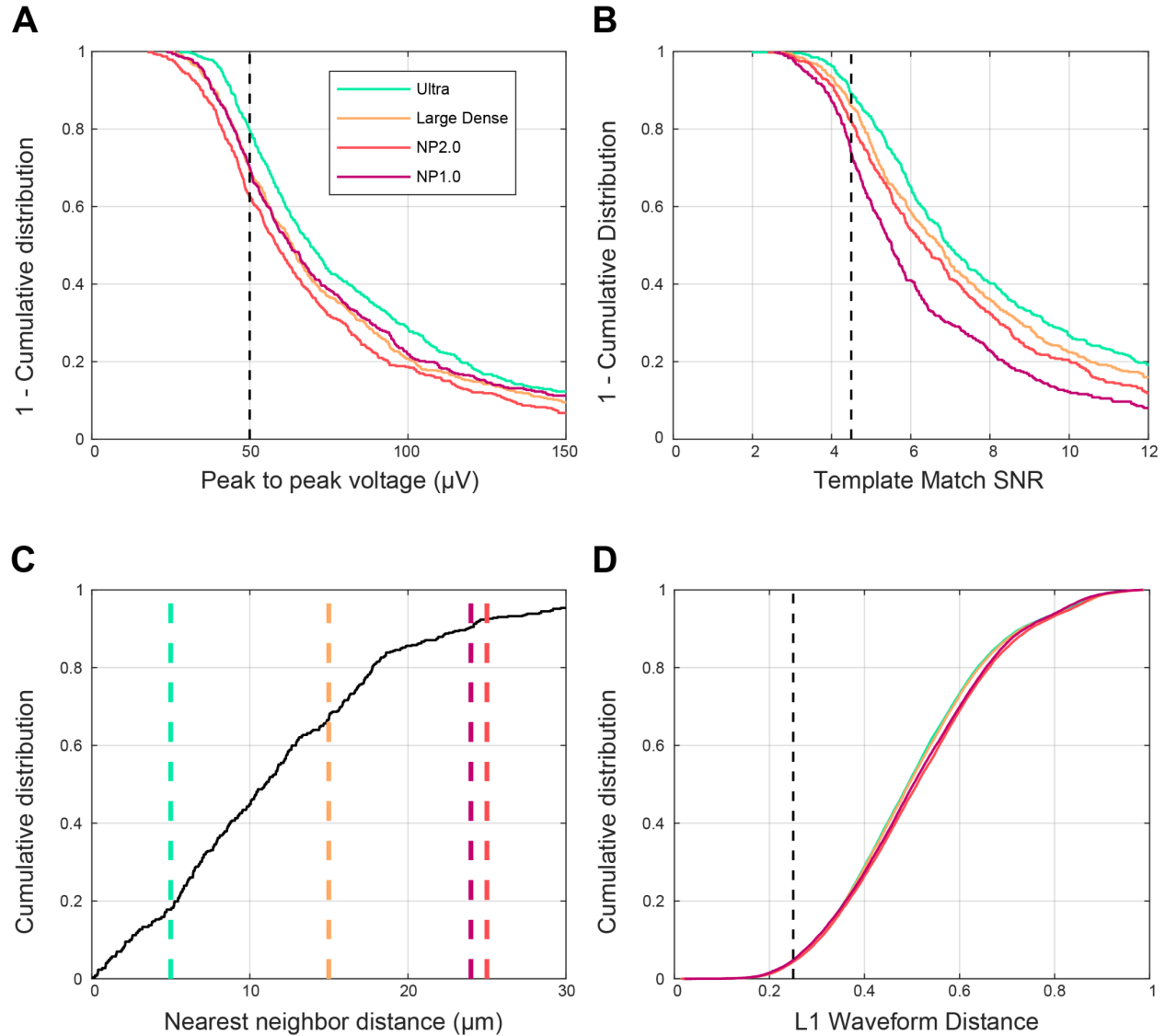


Supplemental Figure S6. LFP signal in cortex, hippocampus, and thalamus using standard and linear NP Ultra probe configurations. **A**, Schematic of insertion location. A switchable NP Ultra probe was inserted sequentially through cortex, hippocampus, and thalamus and recordings were made in the standard 48 x 8 and linear 384 x 1 probe configurations. **B**, Initial LFP recording from cortex made using the 48 x 8 probe configuration. LFP data was filtered using a 3rd order Butterworth filter between 0.5 Hz and 300 Hz. **C**, Final LFP recording from cortex using a bank of contacts (48 x 8 configuration) higher up on the probe after full insertion to thalamus. **D**, LFP recording from thalamus to cortex in the linear probe configuration. Inset shows sharp wave ripple localized to a subset of channels, filtered between 20 Hz and 250 Hz. **E**, Heatmap showing the theta-band LFP signal filtered between 4 Hz and 15 Hz across the span of the probe in the linear configuration as in **D**.



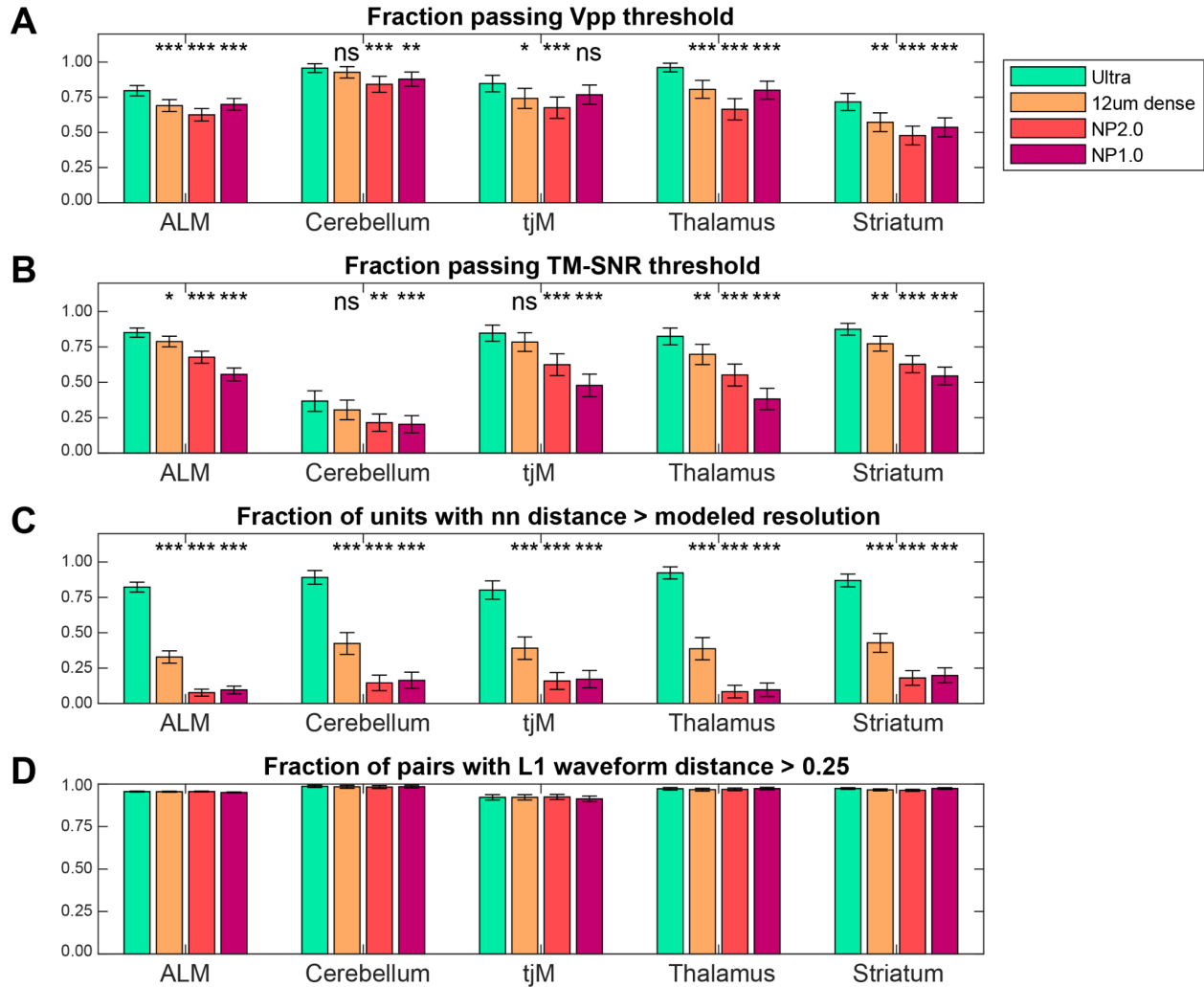
Supplemental Figure S7. Computational modeling shows that most spatial features of extracellular action potentials are captured by measurements with the site spacing of NP Ultra and NP 2.0. **A**, A detailed biophysical simulation of the extracellular action potential (EAP) of a multi-compartmental single neuron (gray) was sampled along random linear trajectories through the 3D space near the neuron. The somatic intracellular waveform is shown in the inset. Colored dots indicate example points where the EAP was computed. **B**, Waveforms (voltage, V_e , over time) from these example EAP sample points, with colors matching the dots in A. **C**, The spatial profile at the peak time point was extracted from each random trajectory, at 1 μm resolution. Each colored trace represents a different trajectory from near the same neuron. The exact coordinates of each trajectory were chosen randomly, so the peak of each profile (representing the point of closest approach to the neuron) has a different position on the x-axis. **D**, The spatial power spectrum of each trajectory (colors) and mean power spectrum (thick

black line) reveal a gradual fall-off in power to higher frequencies. **E**, To visualize and quantify the impact of sampling at coarser spatial resolution, the original spatial voltage profile at 1 μm resolution was downsampled and reconstructed by linear interpolation at 6, 20, and 40 μm sampling resolutions (roughly corresponding to the sampling resolution of NP Ultra, NP 2.0, and NP 1.0 along the axis of the probe), here shown for a single example spatial profile. The peak amplitude is underestimated and some small features are missed with coarser sampling, as highlighted in the insets. **F**, Success of reconstruction quantified as the percent variance of the original spatial profile that is maintained by the downsample, for each trajectory sampled in panel C. **G**, Zoom in over the range 1 to 40 μm sampling resolutions. The spatial profiles are reconstructed with high fidelity (~99% on average) with sampling of NP 2.0 but begin to lose significant information by the 40 μm spacing of NP 1.0, especially along certain trajectories. This failure to capture the spatial profile well presumably corresponds to the failure to track neurons sampled with 1.0-like patterns across probe motion as reliably as with NP Ultra and NP 2.0 (**Fig. 3**).

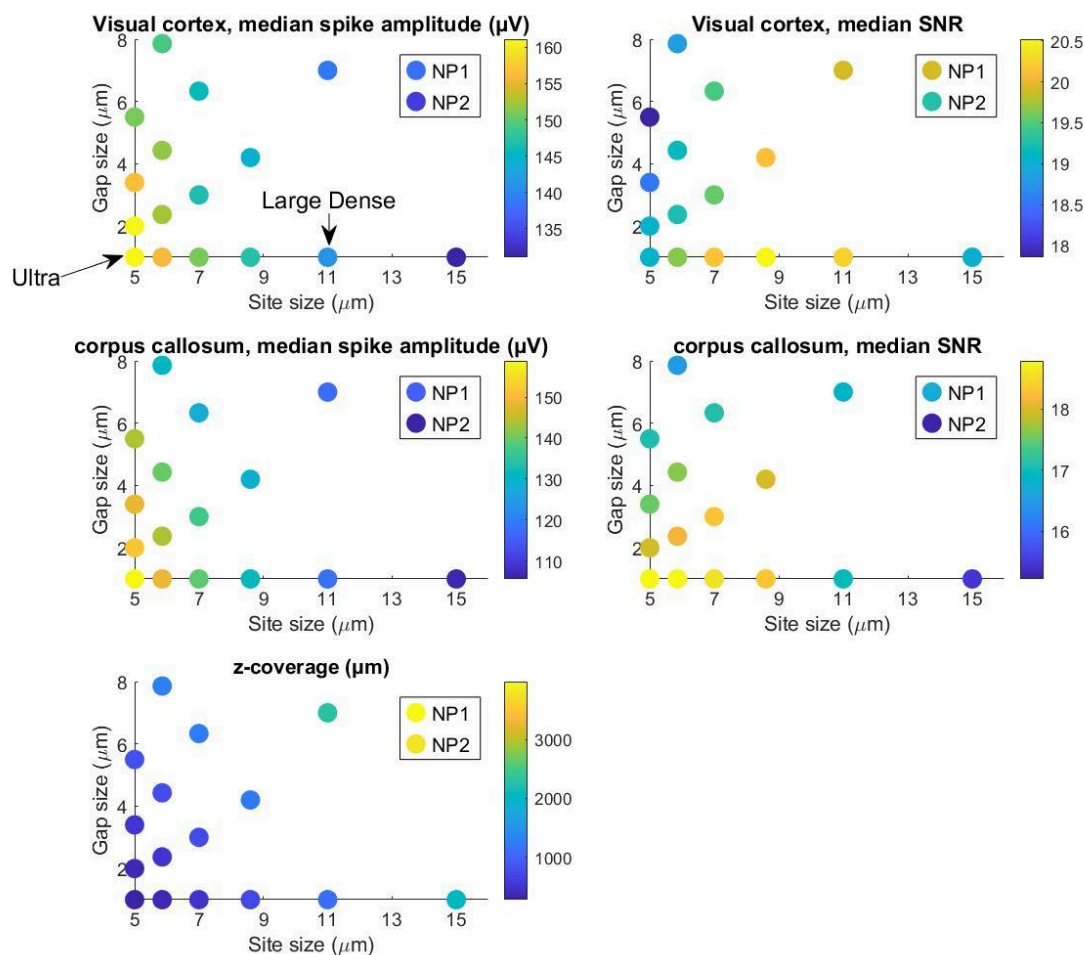


Supplemental Figure S8. Detailed spike sorting analysis. The goal of this ‘sortability’ analysis is to parse the effects of increased site density on the ability to detect and sort spikes. To optimize for the largest spatial range while detecting the largest number of sortable units, we want to understand what aspects of the data make spikes more detectable – that is, more readily found by thresholding or template matching – and then correctly assigned to a specific unit. The ideal sampling is a function of the signal: Very small footprint units require dense sampling to detect them, while units with larger footprints, spaced well apart from each other, will be adequately detected at lower site density. The distribution of unit characteristics is expected to vary across brain regions. With curated NP Ultra data, we can examine the most complete set of units available and estimate the impact of lower spatial sampling. Measurements of the features of extracellular electrophysiology signals that contribute to sorting are summarized in this and the following figure. See **Fig. 2E** for site patterns. Panels **A-D** show distributions of these features for units measured in 11 separate NP Ultra recordings in mouse ALM. **A**, Cumulative distributions of peak to peak voltage, which is the most important feature for detecting spikes via thresholding. The shift to lower V_{pp} for sparser patterns is due to sampling further from the peak signal, so that a larger fraction of units will not be detected at a given threshold. Although NP 1.0 is slightly sparser than NP2.0, it covers a larger area on the probe, so its V_{pp} distribution is slightly higher than NP 2.0. The dashed vertical line at 50 μV

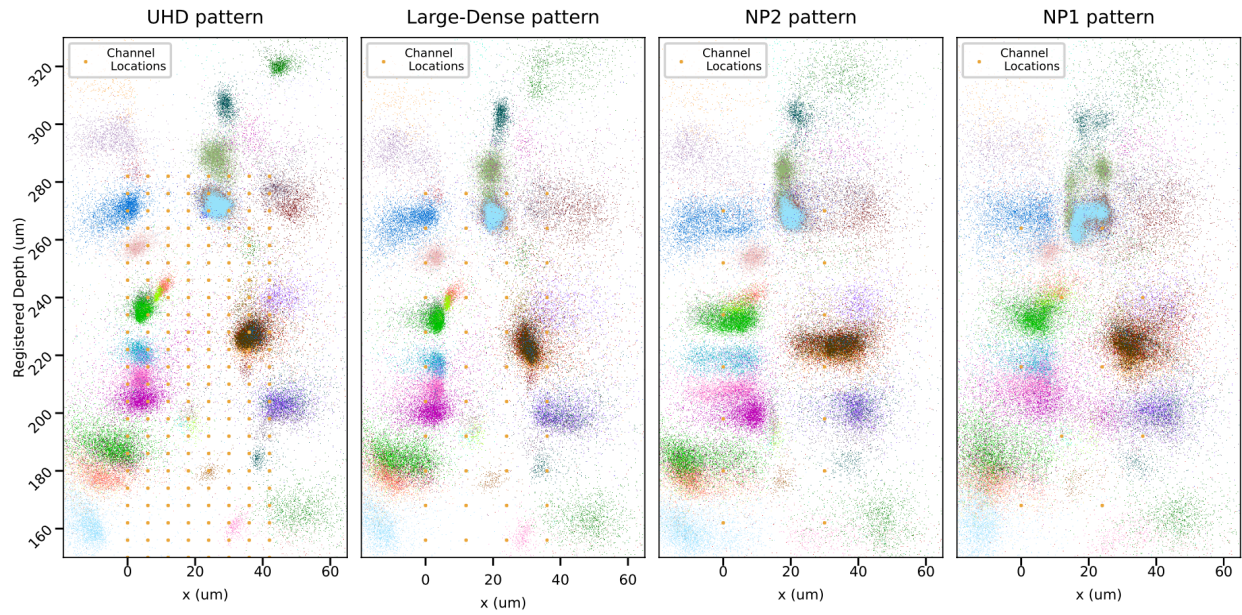
indicates the threshold used in **Supp. Fig. S9. B**, Cumulative distributions of Template Match SNR showing significant shifts to lower SNR at sparser patterns; as discussed in the Methods section, TM-SNR is an estimate of the SNR for spike detection via thresholding of data filtered by a template. The dashed vertical line at 4.5 indicates the threshold used in **Supp. Fig. S9. C**, Cumulative distribution of physical nearest neighbor distances, including ‘unimodal’ and ‘overlapped’ units (see Methods for Analysis of sorting quality), to build as complete a picture of the activity as possible. The effect of the lower spatial resolution of sparser patterns is that more units have overlapped distributions of spike locations. Pairs of units separated by less than spatial resolution of the probe may still be sorted by other features, but can not be fully separated by spike localization alone. **D**, Distribution of all pairwise waveform shape distances for all unimodal units in this set of recordings; see Methods for a description of the waveform distance metric. Similar distributions across site density patterns, indicates that the spatial variation of waveforms for most units is smooth enough to be adequately detected at lower site density. See **Supp. Fig. S9** for quantification of units passing a threshold of 0.25 (vertical line; corresponds to 25% of detected signal differs between two templates). Note that this assessment takes no account of noise or waveform variability, and thus underestimates the difficulty of real sorting.



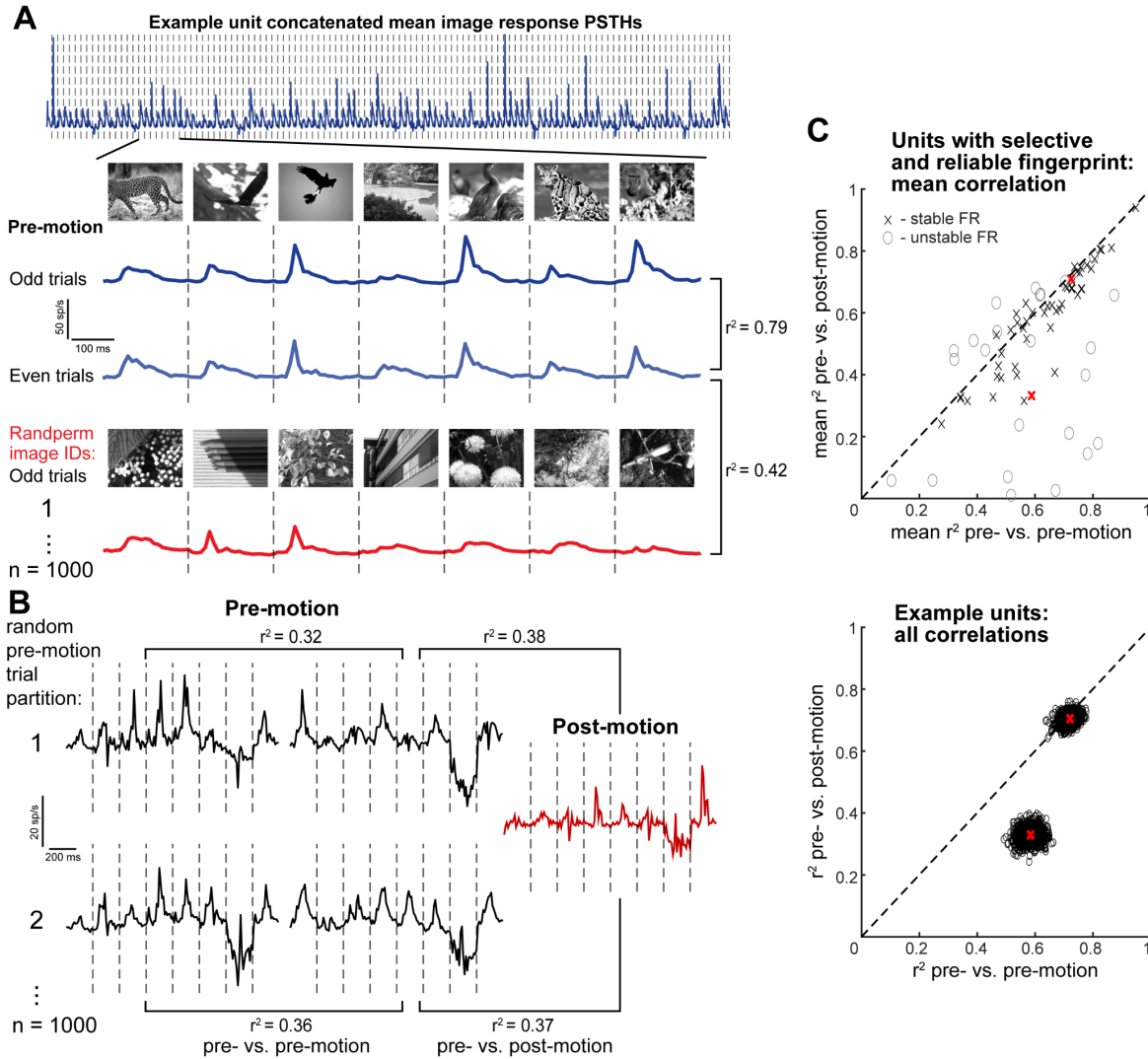
Supplemental Figure S9. The impact of varying site patterns on key features for sorting, across a variety of brain regions. Regions with a larger fraction of small footprint units will have more units that fall below reasonable detection thresholds at lower site density. Regions with closely spaced units will have more pairs that fall below the spatial resolution of low site density probes, which can lead to sorting errors. **A**, Fraction of units with peak-to-peak voltage (V_{pp}) > 50 μ V. **B**, Fraction of units passing TM-SNR > 4.5. **C**, Fraction of units with separation greater than the ideal spatial resolution of the site pattern (see thresholds in figure **Supp. Fig. S8**). **D**, Fraction of pairwise distances greater than 0.25. For all, errors bar shows 2*standard error of the proportion above threshold assuming an underlying binomial distribution. For each region, asterisks show significance of difference from Ultra pattern, calculated using 2-proportion z-test. Differences in fraction of pairs with L1 > 0.25 (panel D) are not significant according to this test. ALM: Anterolateral motor cortex; tjM: tongue-jaw motor cortex.



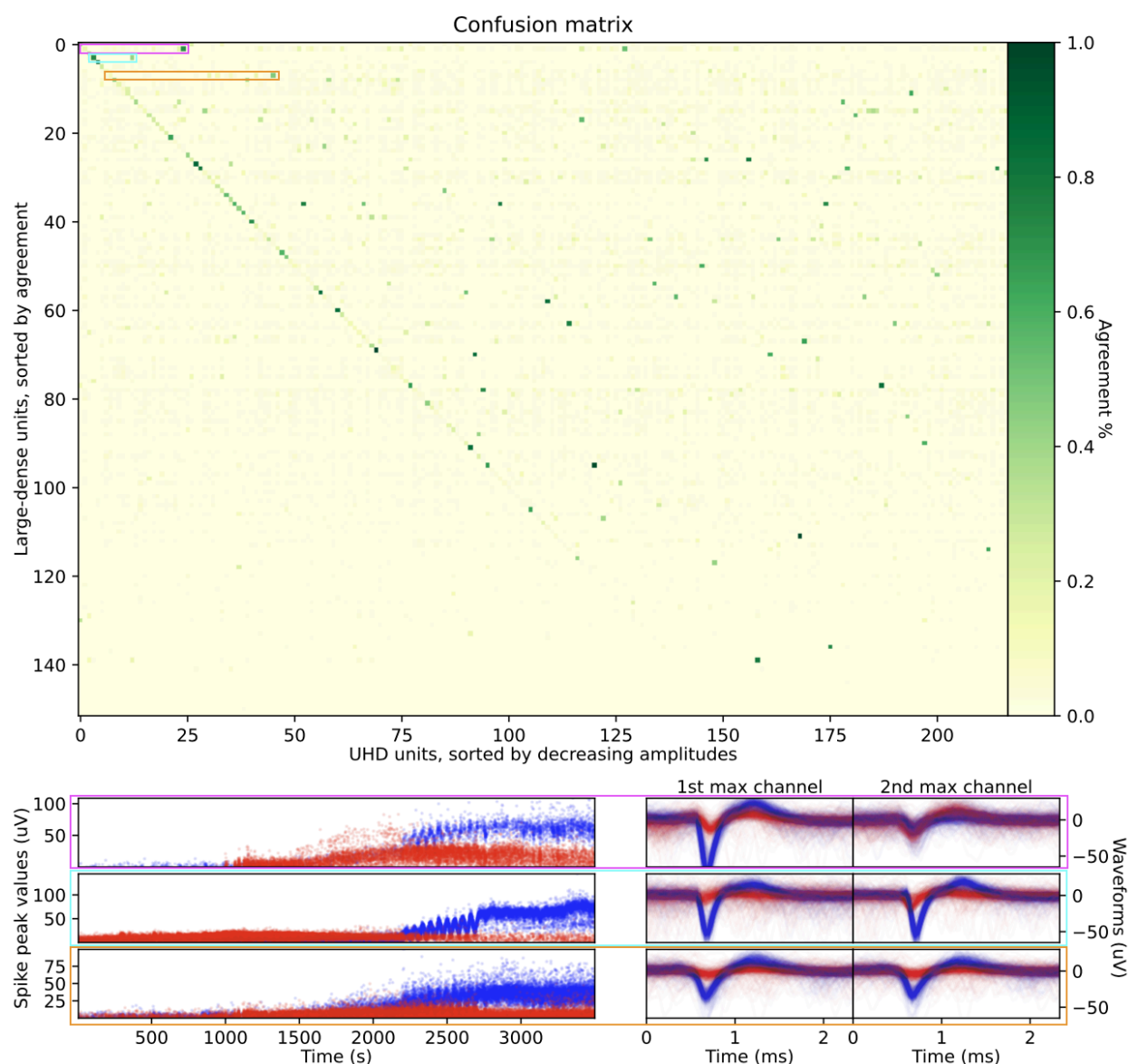
Supplemental Figure S10. Predicting optimal site layout. Using interpolations of the measured average spike waveforms from the recordings discussed in Figure 5, we can predict the observed signal amplitude (peak to peak voltage) and signal to noise ratio (calculated here as amplitude/noise, with noise taken to be the predicted value from the model shown in S2) for a range of site patterns. In these plots, each point is colored by the median amplitude (right column) or median SNR (left column) for all units in a region, detected with a square site pattern with a given site size (x-axis) and edge-to-edge gap size (y-axis). All patterns cover the same $48 \times 288 \mu\text{m}$ area, matching the area of the Ultra probe. With small site sizes, the amplitude is higher, because the electrode averages over a small area and can sample the peak of the electric field. However, thermal noise increases with reduced site size, so SNR is reduced. The optimal site layout is a function of the average spatial footprint of units. In regions with many large footprint units, e.g. the visual cortex (top row) the peak SNR occurs with larger sites with small gaps (see points along the site size axis). In contrast, in corpus callosum (second row), units with small footprints dominate, and the best SNR is achieved with smaller site sizes. The predicted amplitude and SNR for NP 1.0 and NP 2.0 patterns are plotted as separate points. See the colorbar on each plot for the magnitude of these differences. The best detection is always achieved with the smallest gap size, but at the cost of reduced tissue coverage at a given site count. The map in the third row shows the z-coverage for the patterns tested, illustrating the trade off between density and coverage with a fixed channel count of 384.



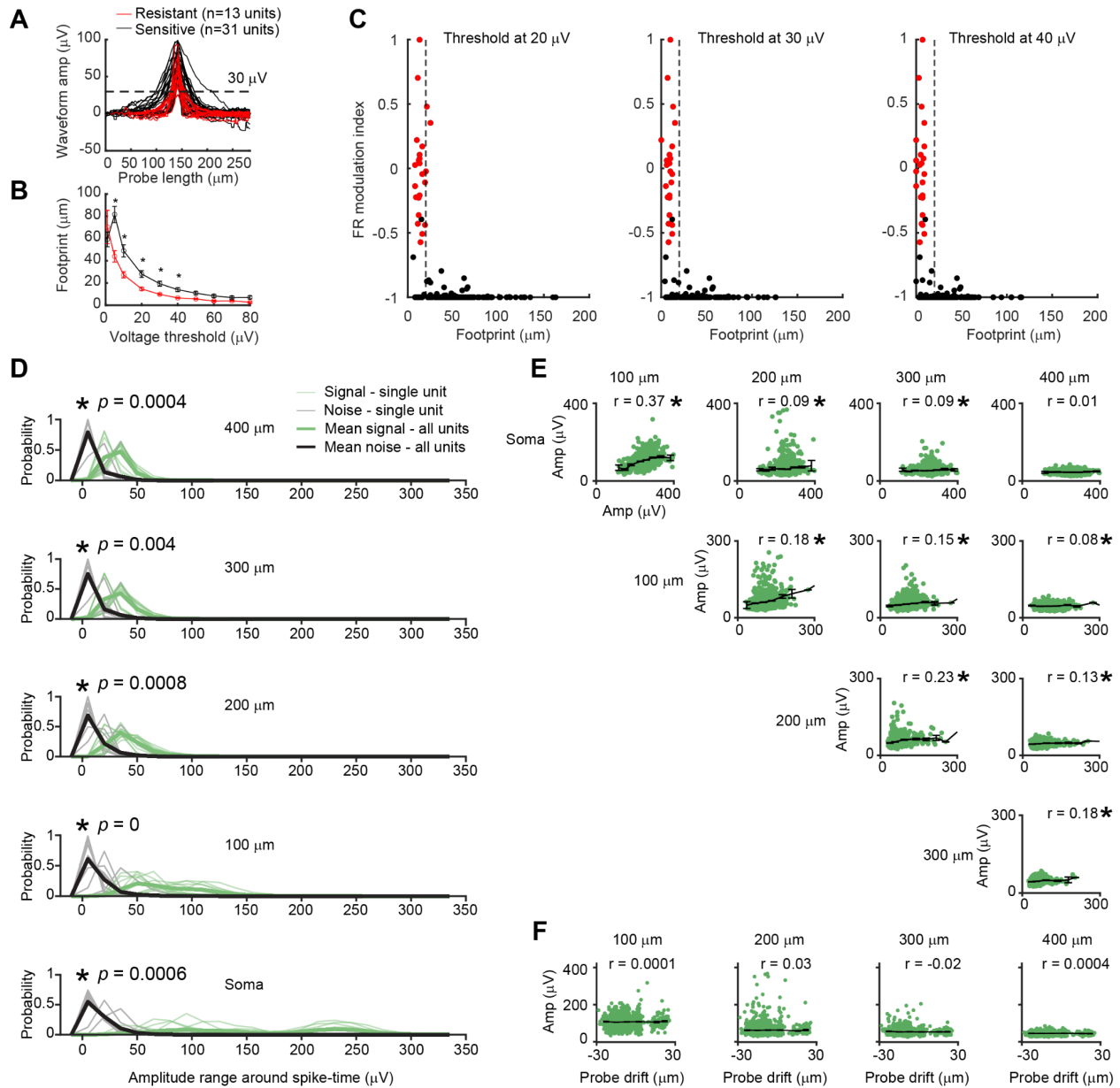
Supplemental Figure S11. The impact of varying site patterns on spike localization. Spikes were detected and clustered according to their location and amplitude obtained on the original UHD recordings. They were then “relocalized” for each re-sampling pattern. The four panels show the spike locations above 150 μm , on the plane of the probe. We see that clusters are more spread out as channel density decreases. This effect is amplified when the source neuron is outside of the probe boundaries.



Supplemental Figure S12. Methods for stability and yield analyses. **A**, Fingerprint calculation and inclusion criteria. An example neuron's visual response "fingerprint" is shown in the top panel; it is calculated by concatenating the PSTHs (0 to 200 ms) aligned to the onset of 118 different natural images. Inset: Fingerprint of the same neuron for 6 images calculated from the odd-numbered trials prior to imposed motion and the even-numbered pre-motion trials. A null distribution is created by randomly permuting image identities and re-concatenating into the fingerprint vector 1000 times. This preserves the noise and nonspecific visual responses of a neuron while destroying the structure of any selective visual response pattern it may have. A neuron passes the inclusion criteria if the pre-odd vs. pre-even correlation is greater than all 1000 pre-even vs. shuffled correlations. **B**, Assessment of fingerprint tracking stability. Of the neurons passing the inclusion criteria for fingerprint reliability and selectivity, 1000 unique random partitions of the 30 pre-motion trials are created (2 sets of non-overlapping 15 trials for each partition) and fingerprints are constructed from each set. Fingerprints from each subset of pre-motion trials are correlated with each other, and one pre-motion fingerprint subset is correlated with the post-motion fingerprint. **C**, Top plot: mean pre- vs. pre- and pre- vs. post-motion correlations for all neurons that passed the inclusion criteria for fingerprint reliability and selectivity in an example NP Ultra session. Gray circles represent neurons with an unstable firing rate ($>4\times$ difference in mean PSTH firing rate pre- vs. post-motion). Bottom plot, all 2000 correlations from each random partition are shown for 2 example neurons. Each dot represents the 2 correlation scores from the 1000 random partitions. Mean correlation scores for these two example neurons are marked by red X's in bottom and top plots.



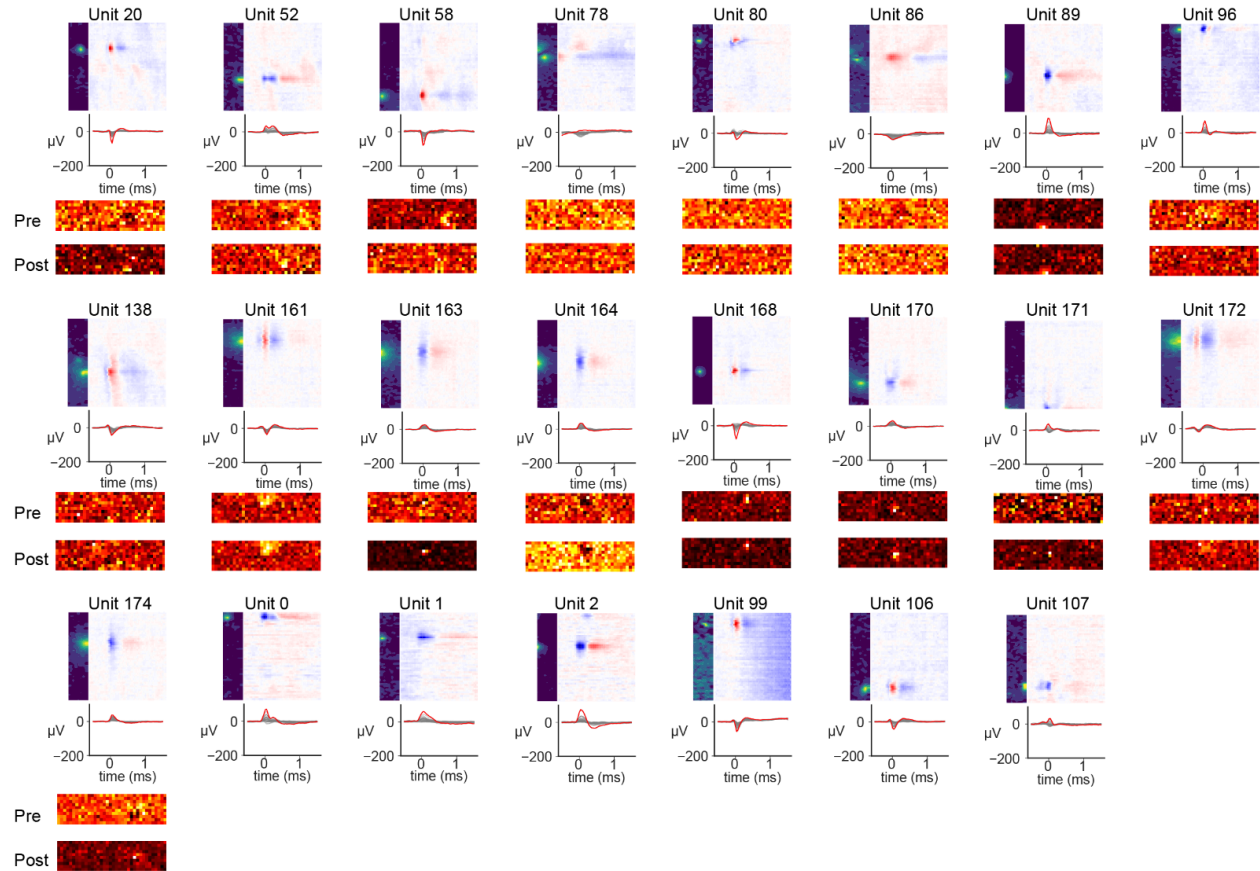
Supplemental Figure S13. Improved yield for NP Ultra is not due to over-splitting units. To check that larger yield from NP Ultra probes is not due to oversplit units, thus artificially leading to a larger number of good units compared to resampled data sorts, we examined the confusion matrix between the Ultra and Large-dense sort (Top). The Ultra units (x-axis) are sorted by decreasing amplitudes, and Large-dense units (y-axis) by agreement percentage. Although most units have low agreement, partly because many more spikes are discovered in the Ultra data, we found some high amplitude units being potentially split i.e. multiple Ultra units having high agreement with a single Large-dense unit. We isolated, for the three highest-amplitude potential split units, the combination of units being merged in the Large-dense sort (pink, cyan and orange rectangles). For these three pairs, we plot their peak absolute value (Bottom left), and 500 randomly sampled waveforms on each unit's max channel (Bottom right) in red and blue respectively. Waveforms are sampled after the imposed motion period ending around 3000s to avoid motion-induced variability. The lack of overlap between peak amplitudes and waveforms supports the notion that each of these pairs of units are indeed distinct.



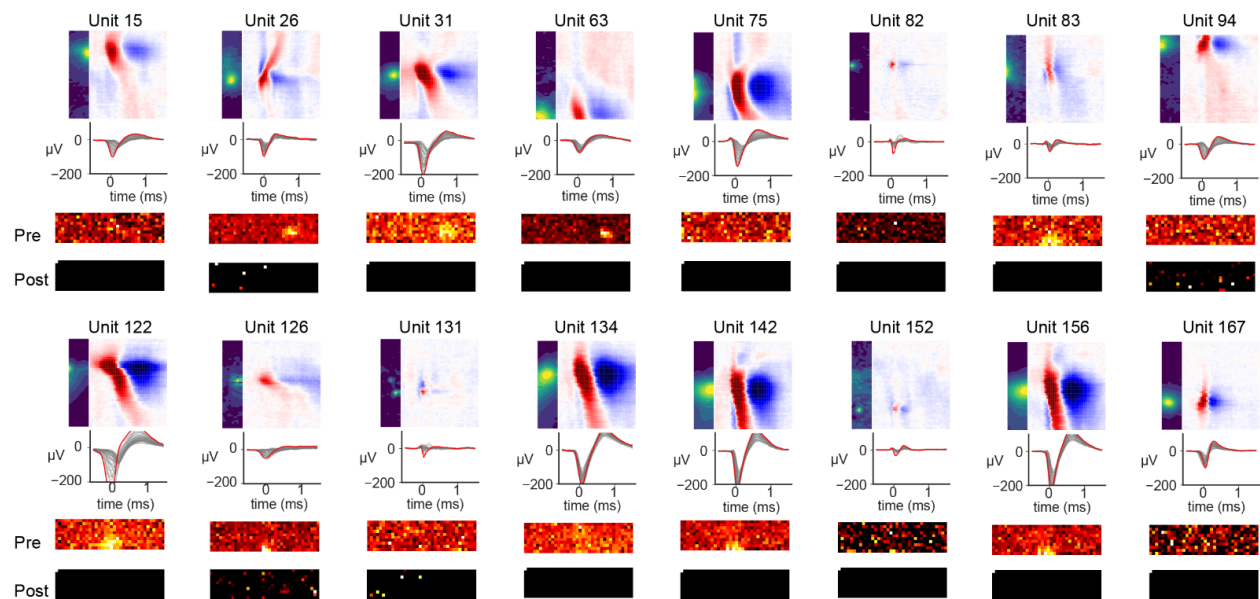
Supplemental Figure S14. Voltage threshold selection for footprint metrics and dendritic back-propagation signal variance. **A-C**, To determine a suitable voltage threshold above the noise floor for footprint calculation, we examined the spatial distribution of waveform peak amplitudes across sites. We used data collected in an 8×48 probe site configuration in the muscimol dataset, which includes ground truth axonal waveforms that were muscimol resistant. **A**, Spatial spread of waveform peak amplitudes along the probe length (along the 48 sites/ $288 \mu\text{m}$ length direction) across different single units. To intuitively visualize the spatial fall-off of peak amplitudes across probe sites around the max peak site in different units (i.e., the definition of footprint metric), we extracted peak amplitudes along a column of sites that contained the max peak sites, and aligned max peak sites to the center position for all units. To ensure a fair comparison, we selected units with comparable max peak amplitudes for both muscimol-resistant ($n = 13$, red) and muscimol-sensitive units ($n = 31$, black). The muscimol-resistant units exhibited a narrow amplitude spread from the peak site, while the muscimol-sensitive units showed a range of spatial spread, both narrow and wide. **B**, Relationship between voltage thresholds and calculated footprints. We then calculated the footprints for the selected units in panel A using varying

detection thresholds at [1,5,10,20,30,40,50,60,70,80] mV. Significant differences in footprints between muscimol-resistant and muscimol-sensitive units were observed across all thresholds ranging from 5 to 40 μ V (two-sample t-test, $p < 0.05$). **C**, The scatter plots for firing rate (FR) modulation index against footprint at different voltage thresholds [20, 30, 40 μ V] for all units in muscimol dataset. Black, muscimol sensitive units; Red, muscimol resistant units. Note there is an overall decrease of footprint with increasing voltage threshold, implicit to the definition of footprint. However, the footprints of muscimol sensitive units are restricted within 20 μ m (dashed line) across all three thresholds. Based on these observations, we selected 30 μ V as the detection threshold for footprint calculation for all units. **D-F**, The occurrence of dendritic back-propagation is likely condition-dependent, including sensory and state-dependent factors (Stuart and Spruston, 2015). To investigate whether we are able to capture some of these variations with NP Ultra probes, in addition to the increased template SNR (**Fig. 4**), we analyzed the spike-to-spike variations of dendritic back-propagation and signal correlation across dendritic distances within single units. **D**, Distributions of spike-to-spike peak amplitudes (green color) and noise (black color) at different dendritic locations within a single unit (thin lines) and averaged across all units (thick lines). Noise distributions were randomly sampled within the recording session with matched spike counts as spike signals. Note a significantly wider, though mostly unimodal distribution of signal variance compared to noise at all dendritic positions (indicated by asterisks, $p < 0.05$, Student's t-test), suggesting the presence of additional signal variance unexplained by noise distribution. **E**, Spike peak amplitude correlations between nearby sites along the dendrite, in an example unit. The top row had somatic peak amplitudes on the X axis and back-propagation peak amplitudes at increasingly distal locations on the Y axis, at [100, 200, 300, 400 μ m] distance to the soma. Each dot represents a unique spike within the example unit. Note a significant correlation between most nearby sites (indicated by asterisks) and a decrease of correlation coefficients as the dendritic distance increased. **F**, No significant correlation between probe drift (extracted from kilosort) and peak amplitudes across dendritic distances, ruling out a contribution of probe drift to signal variance in **D**.

A all muscimol-resistant units (23/175 units)

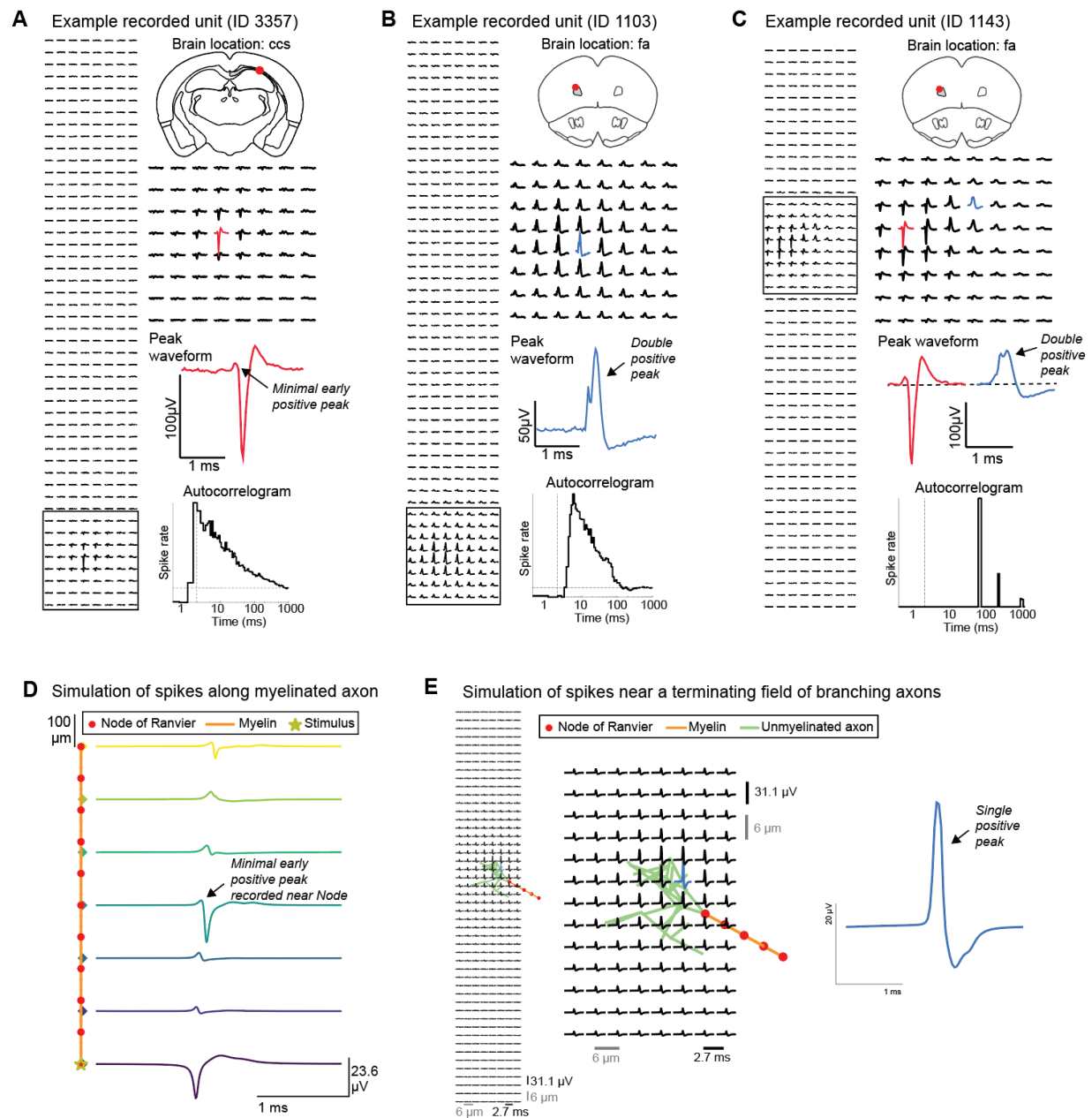


B example muscimol-sensitive units (152/175 units)



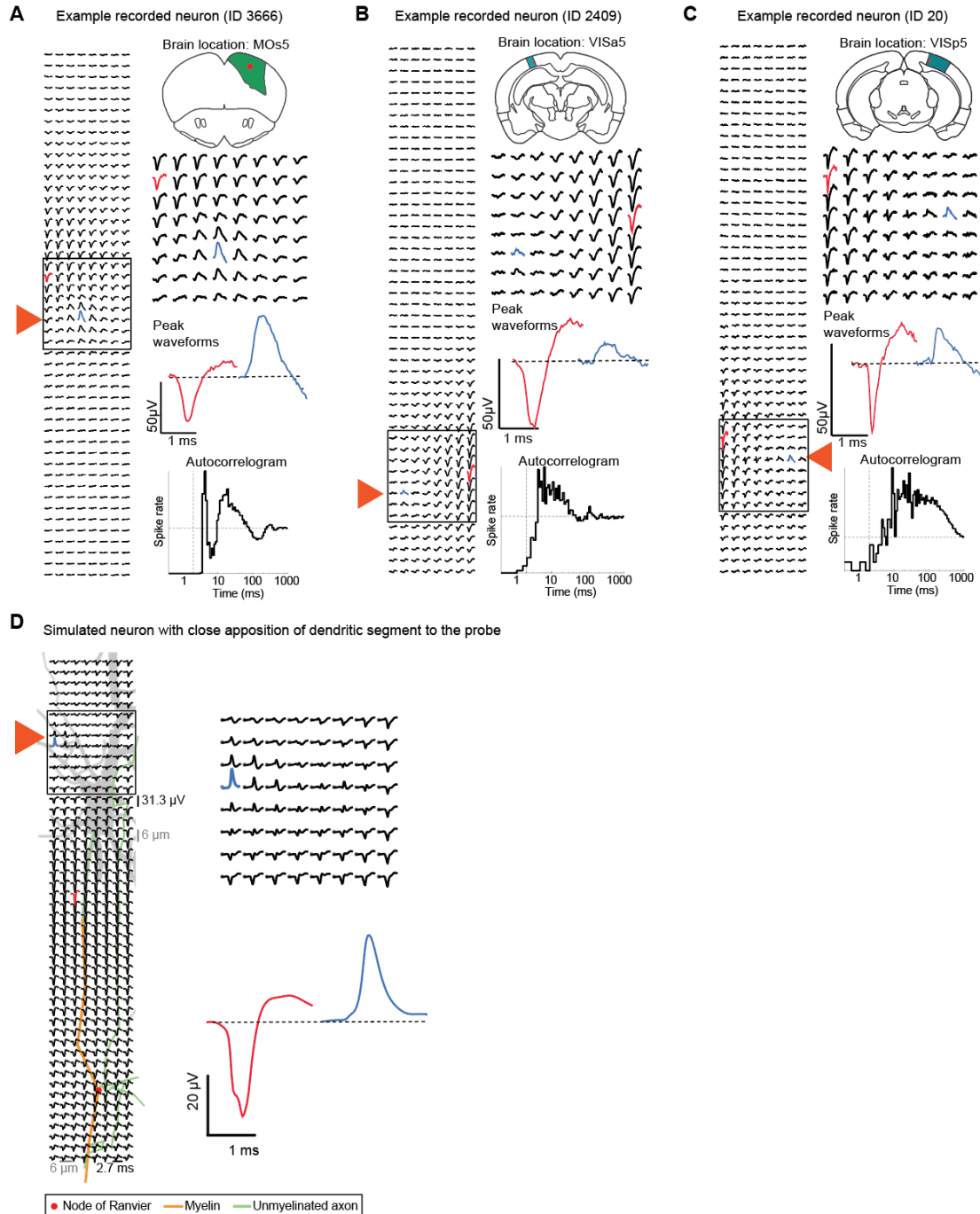
Supplemental Figure S15. Muscimol-resistant and muscimol-sensitive waveforms and receptive field maps before and after muscimol application. A, waveform shapes and receptive field maps of all muscimol-resistant units (23/175 units). 17/23 units in VISp had receptive field maps, while 5/23 units

were recorded in MOp without receptive field maps. Note that receptive field maps changed for some units but not others, and all muscimol-resistant units had small spatial footprint (spatial extent of channels with detectable signal $>30\ \mu\text{V}$). Top: waveform shape plots with three views: probe width x height, height x time, and peak channel over time. Middle: receptive field maps before muscimol. Bottom: receptive field maps after muscimol. **B**, Example muscimol-sensitive units with receptive field maps in VISp. Note that all receptive field maps disappeared due to the silence of neural activity in muscimol, and muscimol-sensitive units had a mix of large and small spatial footprints.

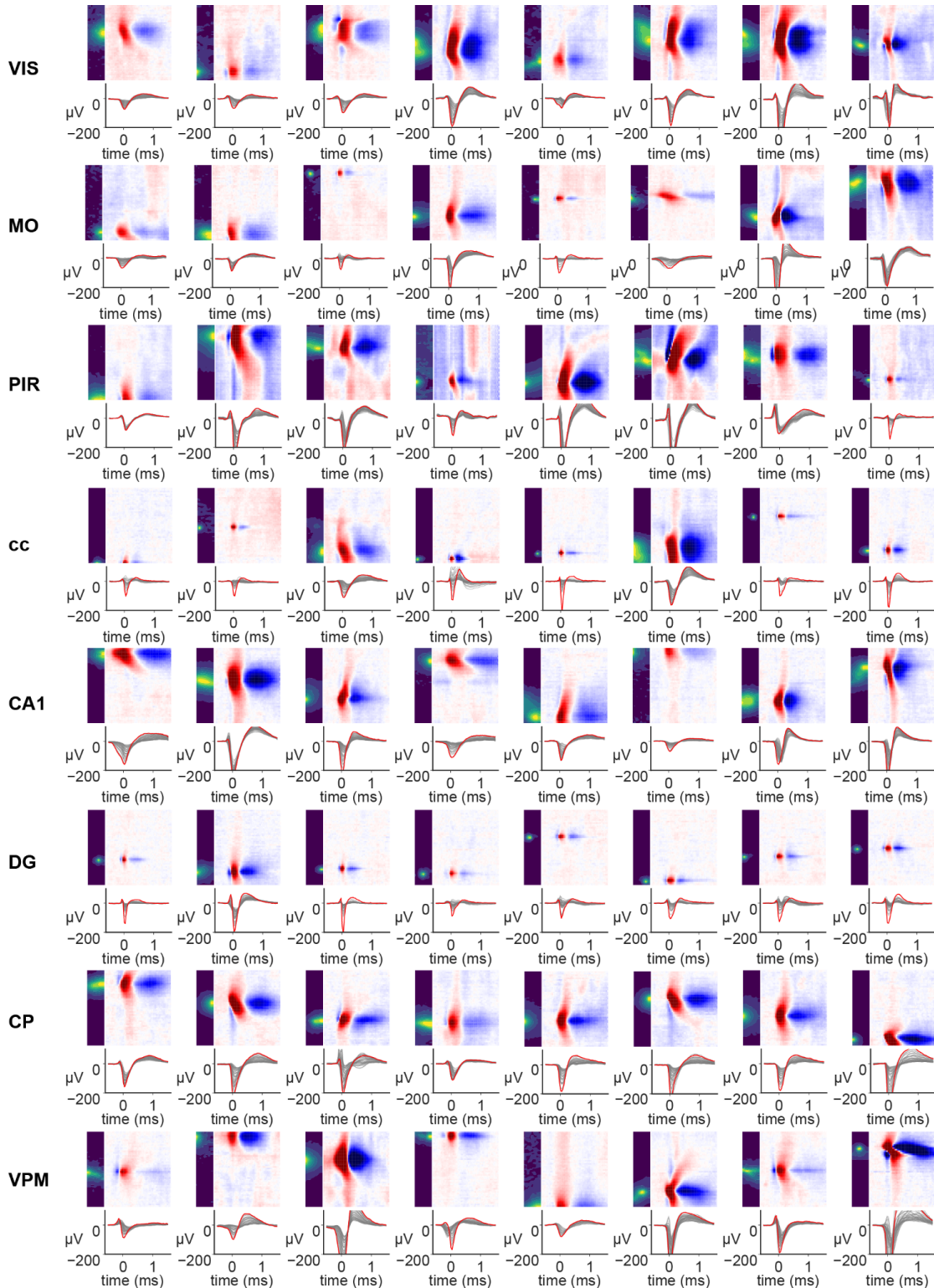


Supplemental Figure S16. Examples of white matter putative axonal recordings, and biophysical simulations of axonal extracellular action potentials. **A**, An example unit recorded in the white matter which has a small-footprint, negative-going action potential. Left, Waveform on all 384 channels, showing a small footprint. Right, unit location (red dot on coronal slice; ccs: corpus callosum splenium), peak channel waveform exhibiting minimal early positive peak, and autocorrelogram (ACG) on log time axis (vertical dashed line at 2 ms). ([link](#) to browse this neuron and view waveform movie) **B**, An example unit recorded in the white matter (fa: corpus callosum anterior forceps) which has only positive-going action potentials on every channel. Waveforms exhibit a double positive peak. Format as in **A**. ([link](#)) **C**, An example unit recorded in the white matter which has a region with negative-going action potential (similar to **A**) and a spatially-adjacent region with double-peaked positive-going action potential (similar to **B**). Format as in **A**. ([link](#)) **D**, Simulation of extracellular action potential recorded at various positions along a

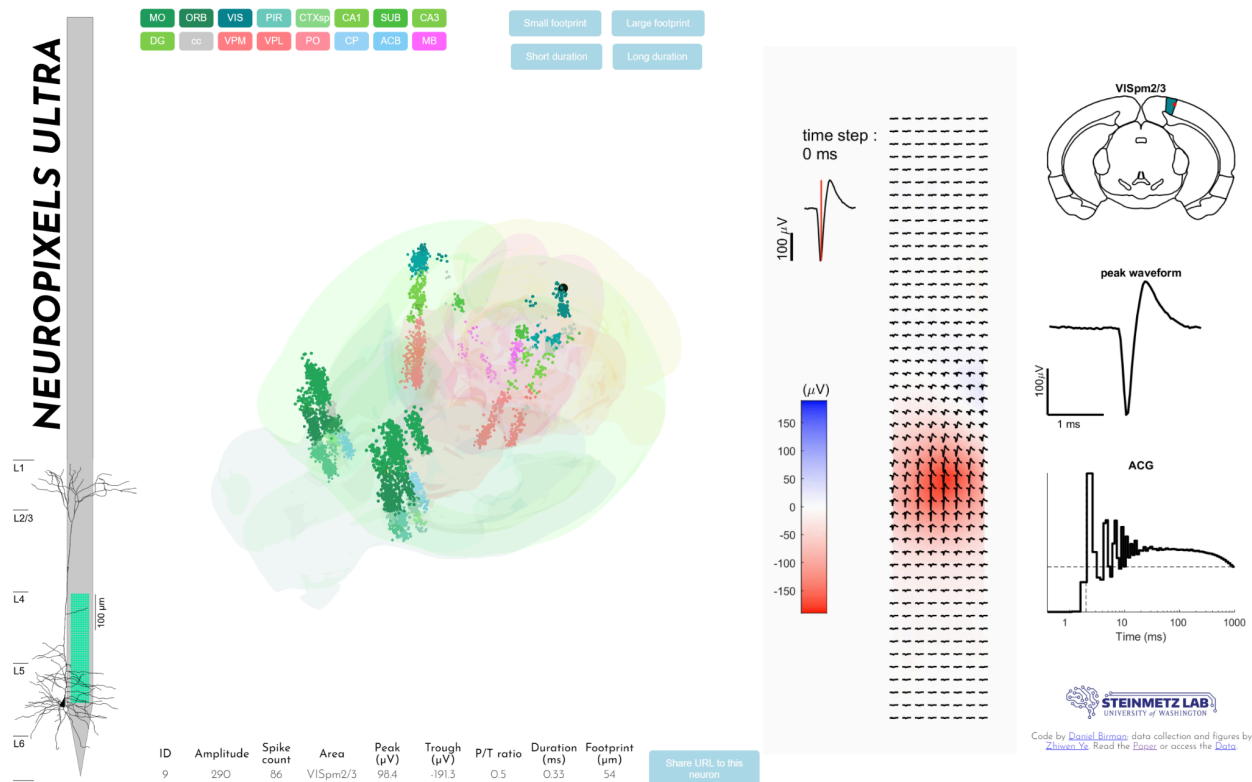
myelinated axon. The physical arrangement of the simulation is shown at left, with simulated traces recorded at each of the colored diamonds shown to the right. Near a Node of Ranvier (red dots), the waveform is negative-going and narrow with minimal positive-going component (similar to the unit from data shown in panel A). See also simulations in Chapter 7 of (Hanes et al., 2024). The early positive peak, when it occurs, reflects return currents of the distant spike as it approaches the recording site along the axon, but can have minimal amplitude depending on the parameters of the simulation. This simulation uses 1 μm diameter axons, 100 μm inter-Nodal spacing, and 1000 μm total axonal length. **E**, A branching axon termination field also produces positive-going spikes, as demonstrated in (McColgan et al., 2017), but typically without a double-peaked shape. The double-positive peak therefore remains mysterious. Here, recordings are simulated from a grid of sites matching the NP Ultra site layout, with the physical layout of the simulated axon shown behind the grid of waveforms.



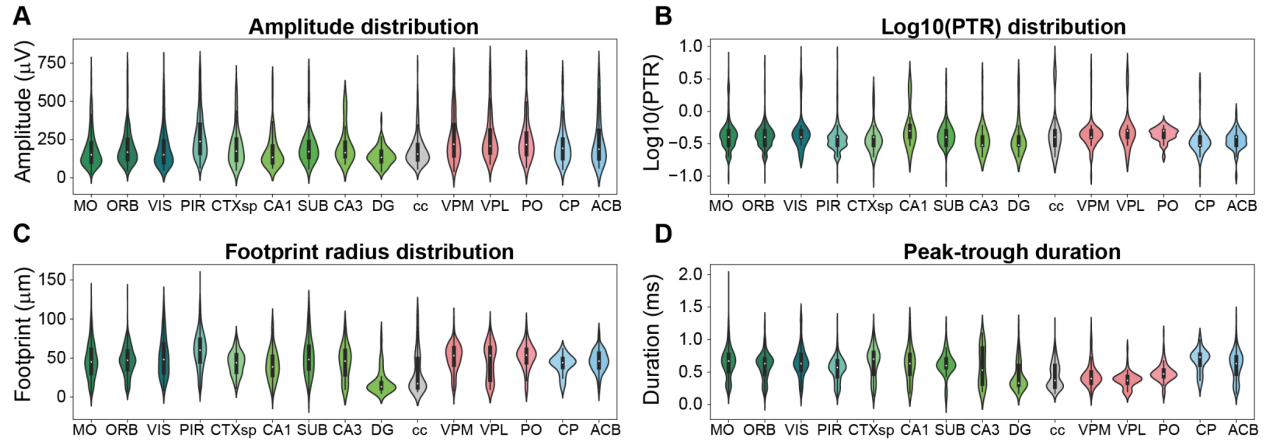
Supplemental Figure S17. Putative observation of electrical fields from dendritic branches. **A**, Example neuron for which the common extracellular action potential pattern of cortical somatic recordings (negative-going spike, large spatial footprint) is accompanied by a very small region with positive-going potentials, indicated by an orange triangle. Figure format as in Supp. Fig. S15A. ([link](#) to browse this neuron). MOs5: secondary motor cortex, layer 5. **B**, Another example neuron as in A ([link](#)). VISA5: anterior visual cortical area, layer 5. **C**, A third example neuron as in A ([link](#)). VISp5: primary visual cortex, layer 5. **D**, A simulated neuron with similar features to the example real neurons. The grey silhouette shows the spatial position of the simulated neuron relative to the probe, with a dendritic segment in close apposition to the probe near the orange triangle.



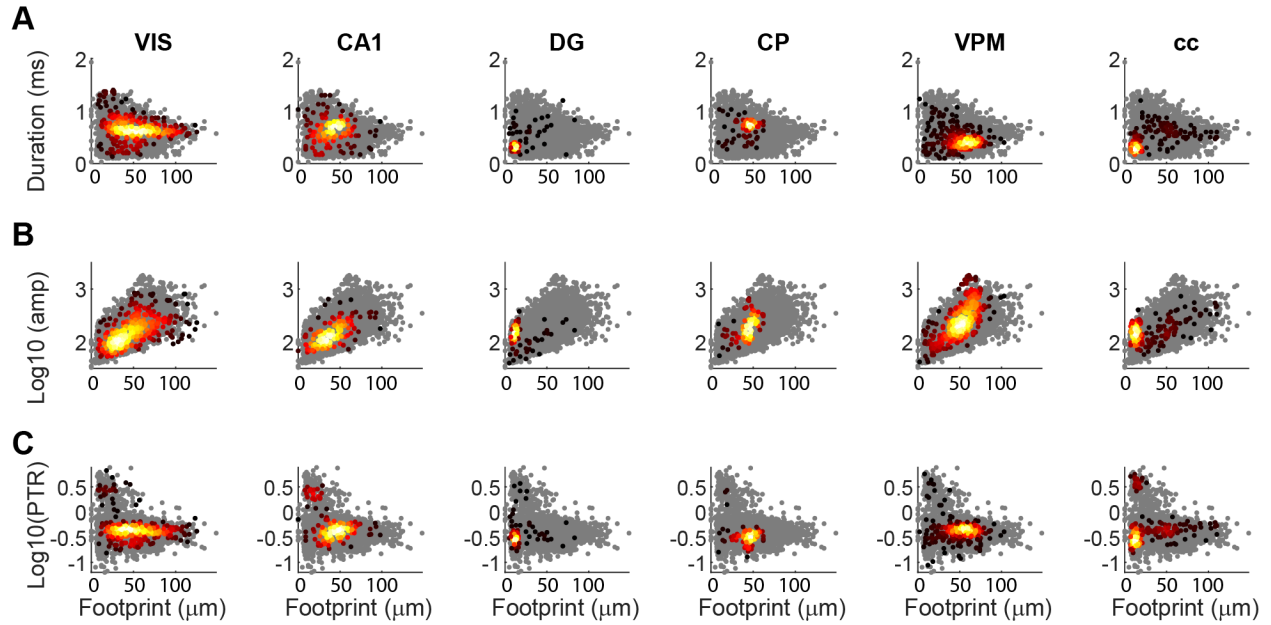
Supplemental Figure S18. Eight example neurons recorded in each of eight example brain regions. Each panel depicts the spatial, spatiotemporal, and temporal waveform of a single neuron. VIS: visual areas; MO: Somatomotor areas; PIR: Piriform area; cc: corpus callosum; CA1: hippocampal subfield CA1; DG: Dentate gyrus; CP: Caudoputamen; VPM: ventral posteromedial nucleus of the thalamus.



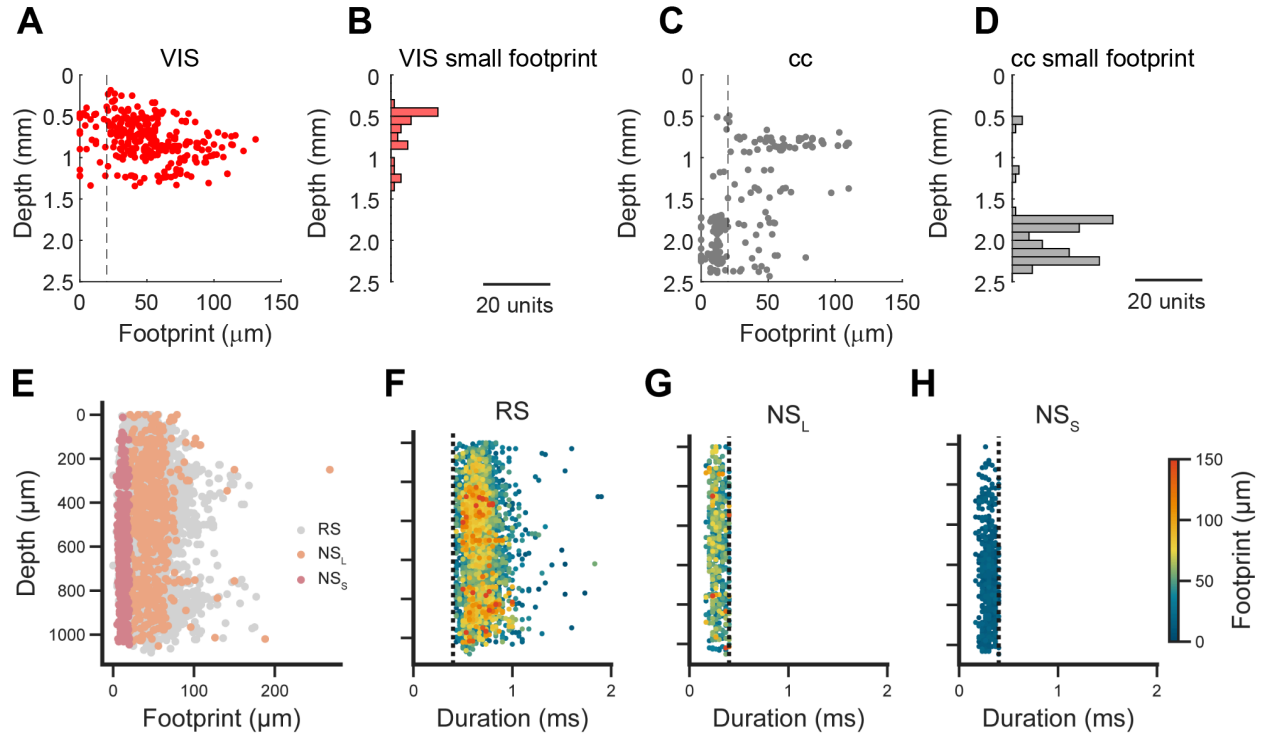
Supplemental Figure S19. Browsable web interface for dataset of brain-wide recordings (<https://npultra.steinmetzlab.net/>). Main panel to the left shows a 3D semi-transparent mouse brain with dots representing the location of each recorded neuron in CCF coordinates. The brain can be rotated by clicking and dragging, or zoomed with the scroll wheel. Neurons can be selected by clicking them. The neurons can be filtered with the buttons at the top of the panel, by brain region or features of footprint size and waveform duration. At the bottom of the panel, several descriptive statistics of the selected neuron are reported. The panels at right show details of the selected neuron: the waveform on all channels of the probe, with an animated movie of the waveform over time and space playing on loop in the background. Upper right shows the location of the neuron on a coronal slice and within its brain region of origin. Middle right, a zoom of the waveform on the peak channel. Bottom right, the autocorrelogram (ACG) of the neuron (note logarithmic x-axis). Finally, the “Share URL to this neuron” will copy a link to the clipboard which encodes the exact state of the website, so that another person clicking that link will allow them to see exactly the same view as you.



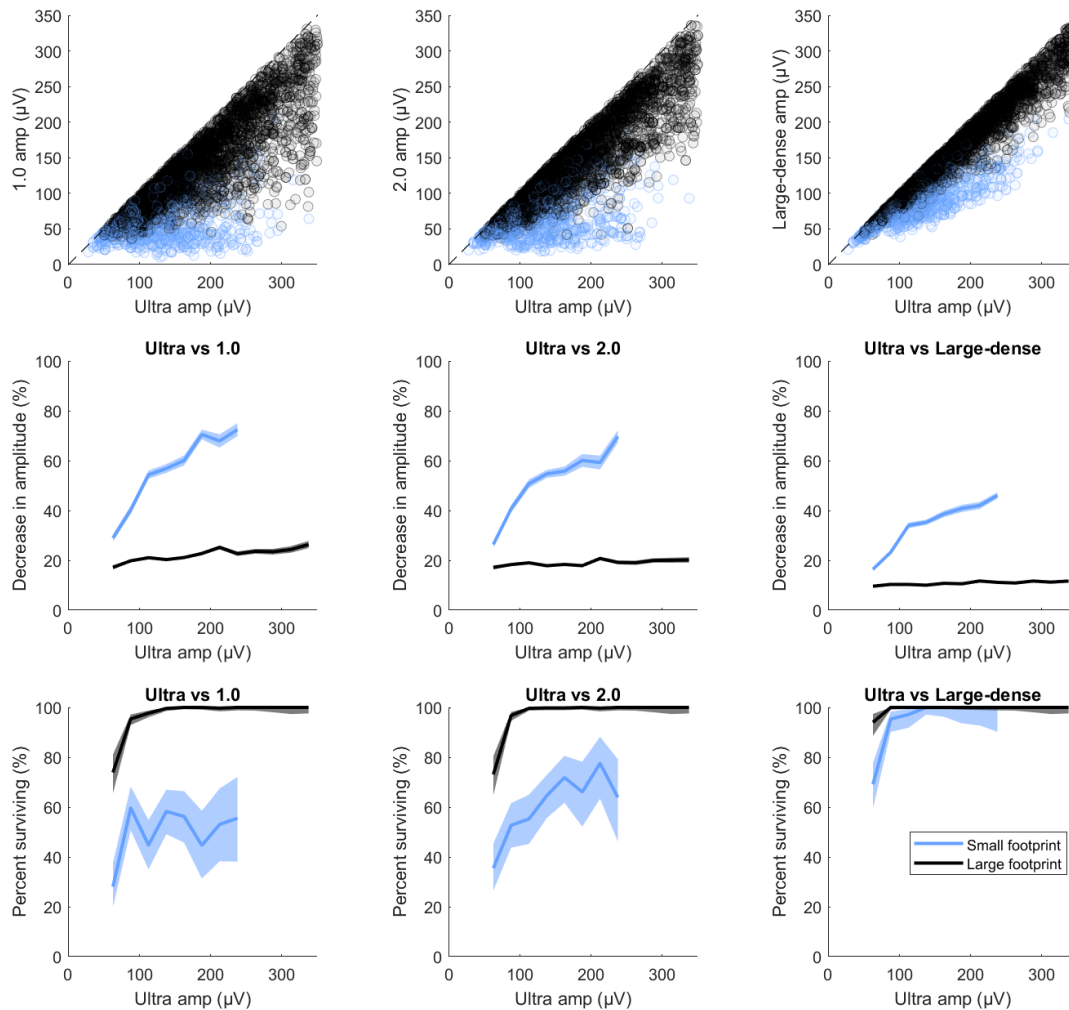
Supplemental Figure S20. Features extracted from some top recorded brain regions. A, Absolute peak amplitude distribution. **B,** Spatial footprint radius distribution. Note that footprint radius distribution of DG and cc waveforms are smaller than the other brain regions. **C,** Log10 of peak-to-trough ratio (PTR) distribution. Note that CP neurons tend to have smaller PTR index. **D,** Peak-to-trough duration (ms) distribution. Note that thalamic regions VPM, VPL, and PO have smaller peak-to-trough duration than other brain regions. MO: Somatomotor areas; ORB: orbital area; VIS: visual areas; PIR: Piriform area; CTXsp: cortical subplate; SUB: Subiculum; DG: Dentate gyrus; cc: corpus callosum; VPM: ventral posteromedial nucleus of the thalamus; VPL: Ventral posterolateral nucleus of the thalamus; PO: Posterior complex of the thalamus; CP: Caudoputamen; ACB: Nucleus accumbens.



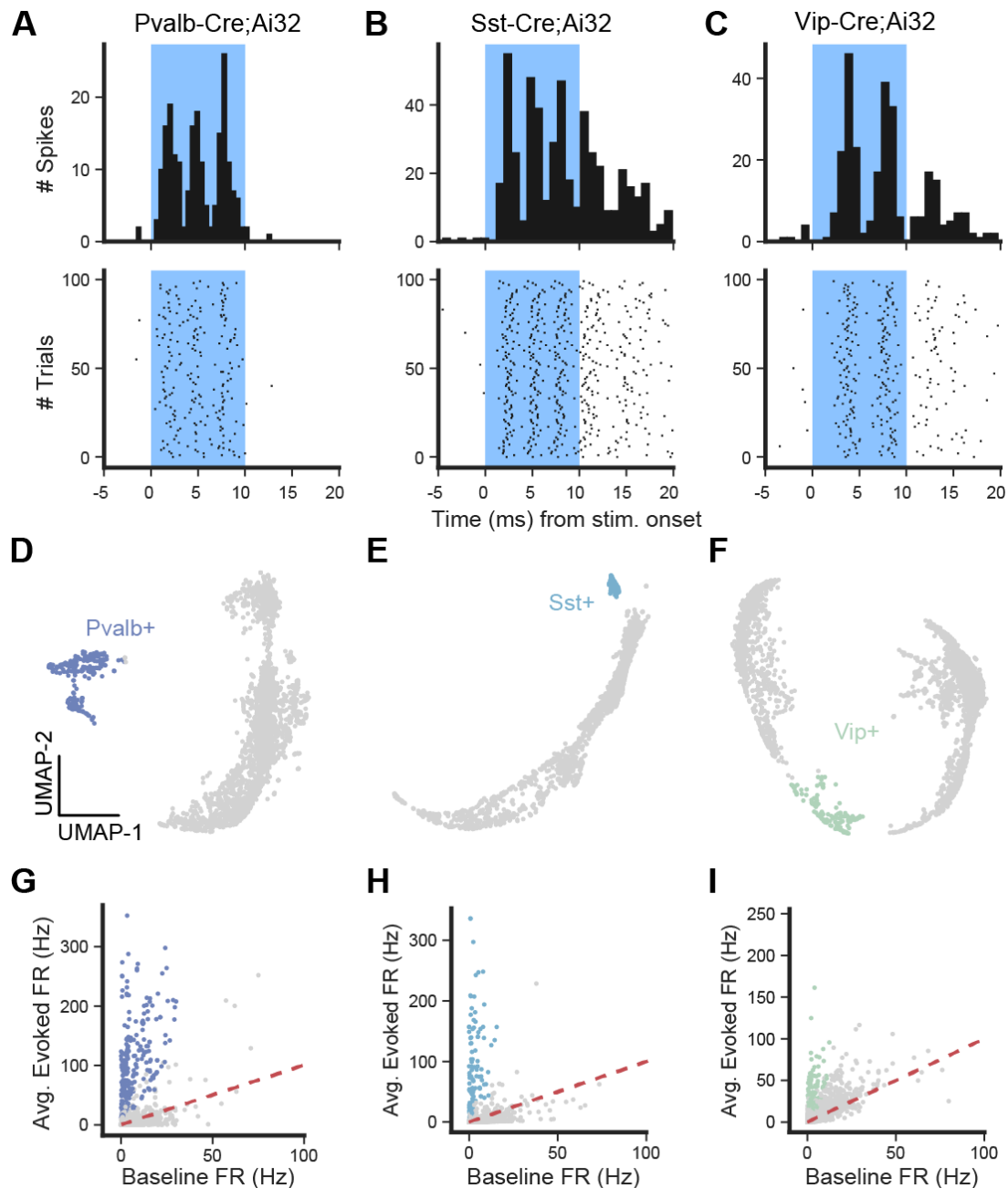
Supplemental Figure S21. Relationships between features extracted from some top recorded brain regions. **A**, Waveform duration vs spatial footprint. Waveform duration is defined as duration between peak and trough. Dots colored black to yellow are units from this region; gray background dots are the distribution of units from all areas for reference. **B**, Log10(amplitude) vs spatial footprint. **C**, Waveform spatial footprint radius vs log10(peak-to-trough ratio). Values of log10(PTR) above 0 indicate absolute peak amplitude are higher than absolute trough amplitude, Values of log10(PTR) below 0 indicate absolute peak amplitude are lower than absolute trough amplitude.



Supplemental Figure S22. Depth separation of units with small footprint in the corpus callosum and visual cortex. **A**, Footprint and depth relationship of units in the visual cortex, from the brain-wide dataset (from Fig. 4). Depth 0 indicates the surface of the cortex. Note that a number of small footprint units are intermingled with larger footprint units across cortical depths. **B**, Depth histogram of units in the visual cortex with footprint smaller than 20 μm . **C**, Footprint and depth relationship of units in the corpus callosum. **D**, Depth histogram of units in the corpus callosum with footprint smaller than 20 μm . The cc can be reached at different depths depending on the location and angle of insertion, which varied across recordings. Note that the depth estimations are calculated from incremental manipulator advancement, which could result in a small overestimation of the true depth due to tissue compression. The border between VIS and cc was estimated with electrophysiological features (excluding footprint) in combination with histology, and was independent of the depth estimation taken from the micromanipulator readings. **E**, Distribution of RS, NS_L , and NS_S units of the visual cortex by spatial footprint and depth from the cortical surface, from the visual cortex dataset (from Figs. 7,8). **F-H**, Depth by duration footprint distributions for each of the cell types in **E**, colored by footprint. Vertical dashed line indicates the threshold between RS and NS units.

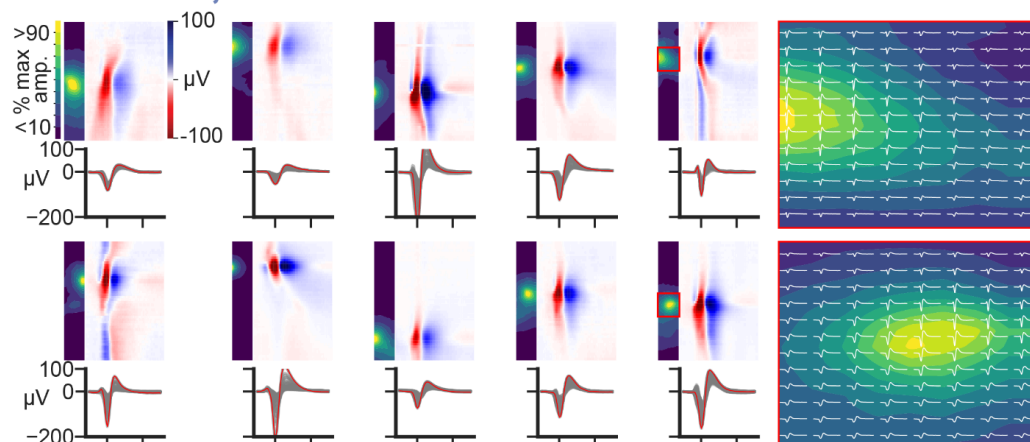


Supplemental Figure S23. Units with small footprint are preferentially more difficult to detect on probes with lower recording site density. Top row, the amplitude of each neuron in the brain-wide acute mouse dataset (Fig. 5; $n=4666$) as measured in the original recording (x-axis) and on each spatially resampled dataset (y-axis; for 1.0-like resampling, 2.0-like resampling, and 'large-dense' resampling (Fig. 2E) from left to right). Small-footprint units are plotted in blue. Middle row, the percentage decrease in amplitude between NP Ultra and each spatially resampled dataset as a function of the original amplitude of the unit in the NP Ultra recording. Small-footprint units (blue) decrease more than large-footprint units (black), especially for resampling at lower density (1.0-like, left panel) and for higher amplitude units. For example, small-footprint units decrease their amplitude by $56.4 \pm 0.8\%$ (mean $\pm$ s.e.m.) when spatially resampled in the NP 1.0-like pattern, but large-footprint units only decrease by $23.6 \pm 0.3\%$. Bottom row, Applying a cutoff of $40 \mu\text{V}$ minimum amplitude to approximately estimate the percentage of units that would 'survive' the spatial resampling (i.e. still be spike sortable), we find that nearly all large-footprint units of sufficiently large amplitude would remain above the amplitude threshold and thus survive to be recorded with the lower density probe, but approximately half of small-footprint units would instead fail to survive in a lower density recording.

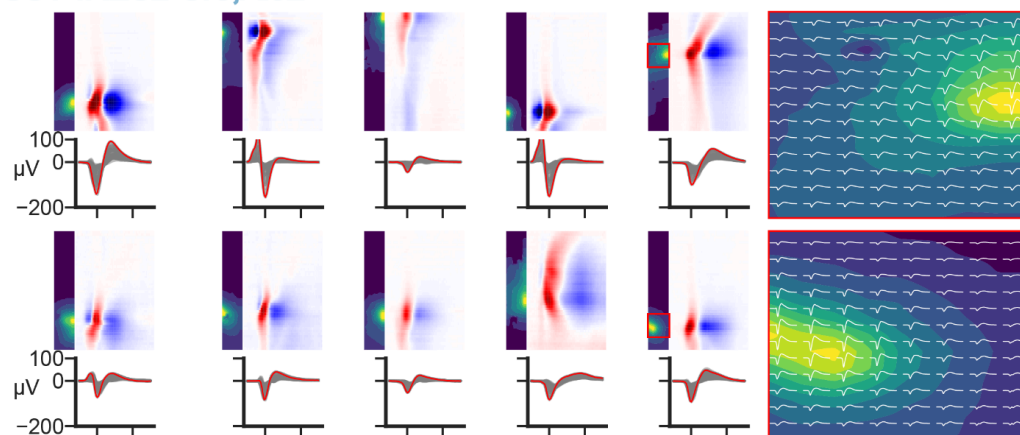


Supplemental Figure S24. Identification of optotagged units. Optotagged units were identified by photo-evoked changes in firing rate in average PSTHs from both 10ms (square-wave) and 1s (cosine ramp) pulses. **A-C**, PSTHs (top) and spike rasters per photostimulation trial (bottom) in three example optotagged units from each transgenic mouse line. **D-F**, 2-dimensional UMAP representations of concatenated firing rate responses (see Methods) for all trials and stimulation-protocols for each transgenic mouse line. Optotagged units were found via unsupervised clustering (DBSCAN). **G-I**, Scatter plots showing baseline firing rate against the average evoked firing for each transgenic mouse line. Optotagged units found via unsupervised clustering (**D-F**) are shown in color, confirming that the clustering successfully identified units that were strongly modulated by light delivery.

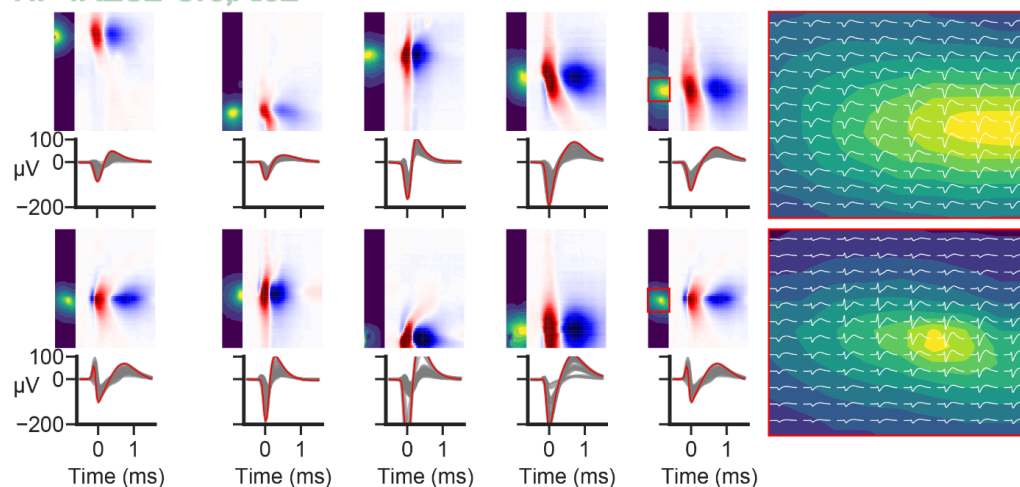
PV-IRES2-Cre;Ai32



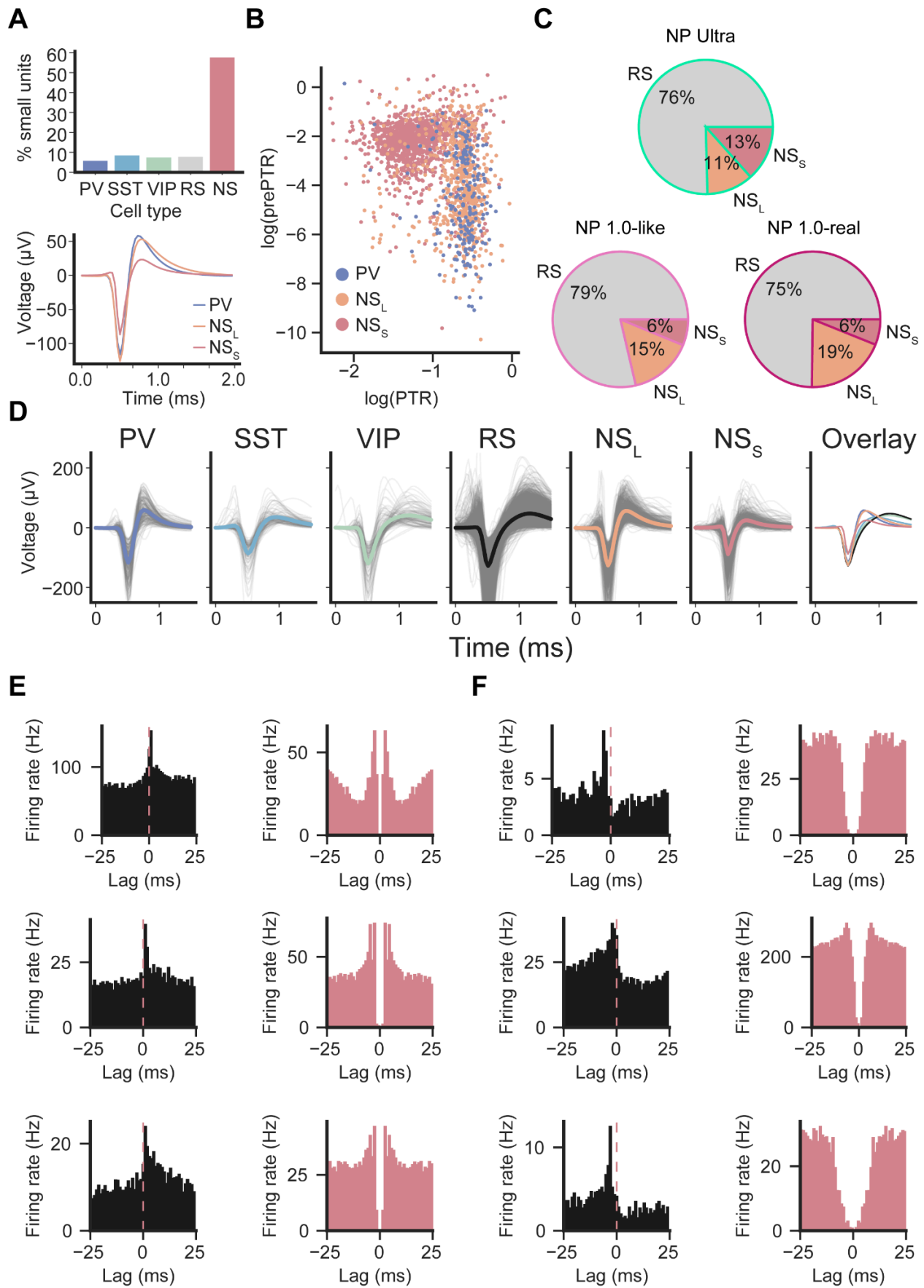
SST-IRES2-Cre;Ai32



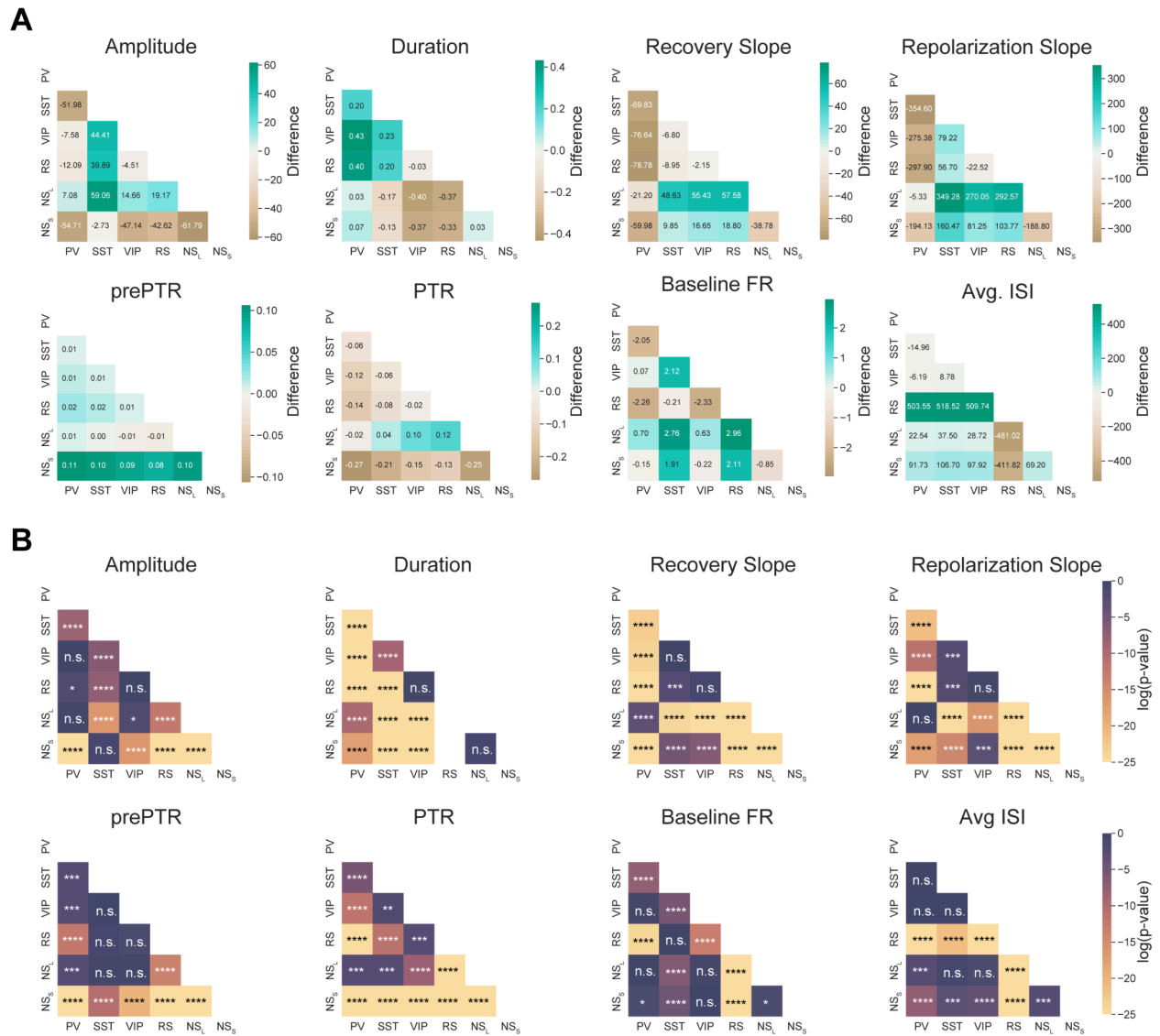
VIP-IRES2-Cre;Ai32



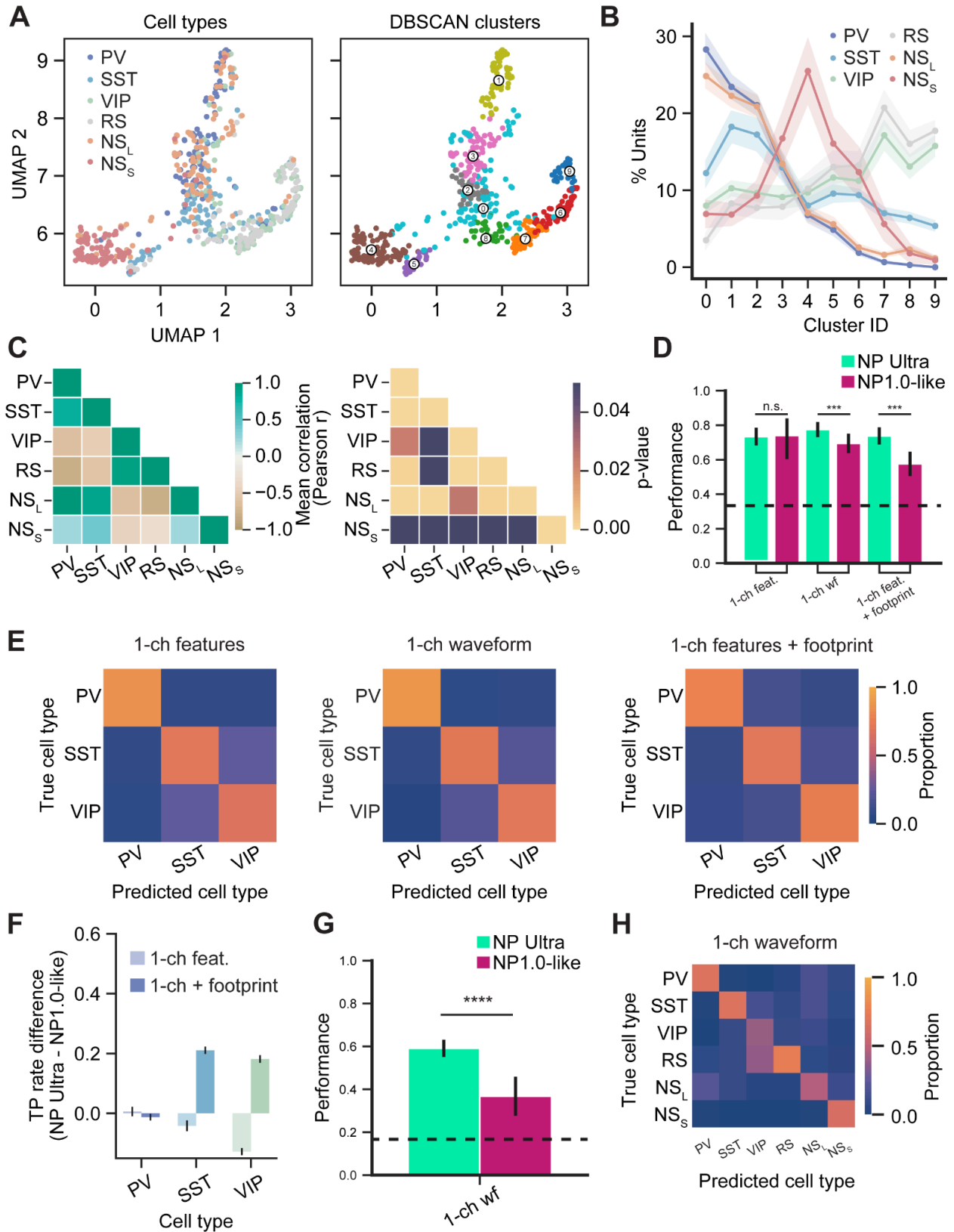
Supplemental Figure S25. Diversity of extracellular waveforms across transgenically-defined inhibitory types. Ten example units demonstrating the variability in extracellular waveforms within the three examined inhibitory neuron subclasses. Plotting conventions as in Fig. 1G. Heatmaps at right are insets (red box) of the right-most unit displayed in each row, overlaid with waveforms recorded on the 12 x 8 channels surrounding the peak amplitude channel.



Supplemental Figure S26. Waveform characteristics of each unit class, and inferred synaptic connections of NS_S units. **A**, Top: Percent small footprint units within each unit class category. Bottom: Comparison of grand-average waveforms from NS_L , NS_S , and PV units. **B**, Scatter plot showing the log-scaled prePTR and PTR values for units in each of PV, NS_L , and NS_S unit class, showing that NS_S units differ not only in spatial footprint but also in PTR. **C**, Proportion of RS, NS_L and NS_S units from NP ultra, spatially downsampled NP1.0-like pattern and a separate dataset with NP1.0 recordings (NP 1.0-real). Note a proportional decrease of NS_S units in NP1.0-like (6%) and NP1.0-real data (6%), compared to NP Ultra (13%). This results in a reduction of NS_S /NS unit ratio from 54% (NP Ultra) to 28% (NP 1.0-like) and 24% (NP 1.0-real). **D**, Grand-average mean waveforms for each identified unit class, overlaid with each other in the right-most panel. Gray traces are mean waveforms from individual units. **E**, Putative excitatory CCGs (left column) and ACGs (right column) from identified NS_S units. **F**, Putative inhibitory CCGs (left column) and ACGs (right column) from identified NS_S units.

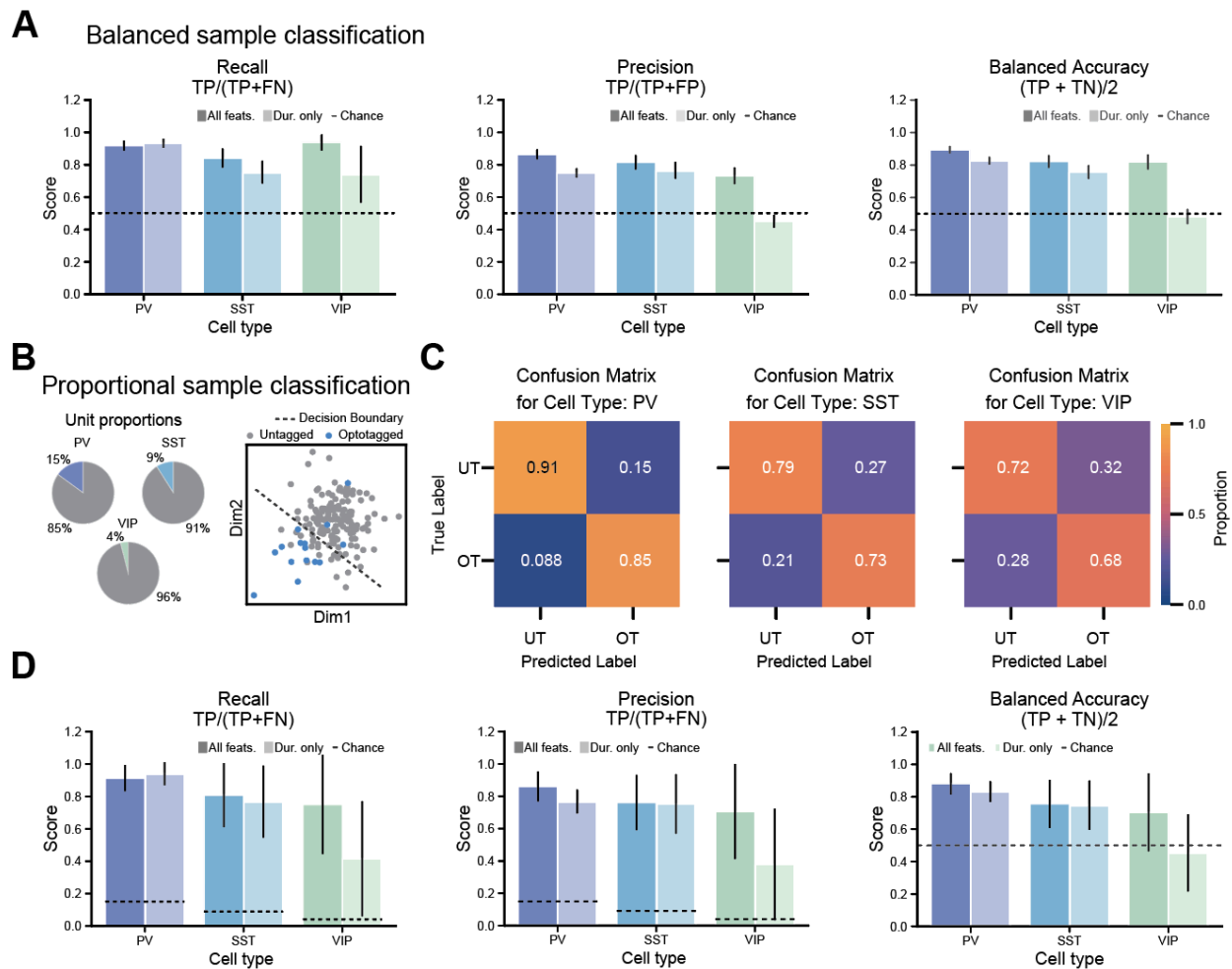


Supplemental Figure S27. Statistics of 1D feature comparisons between unit classes. A, Pairwise median differences of extracted 1D single-channel waveform features. Values are represented as the difference in distribution medians (row-column) between unit classes for each feature. As an example, NS_S had higher prePTR, but lower PTR compared to all other unit classes. B, Pairwise statistics (Mann-Whitney U test, Benjamini-Hochberg multiple comparisons corrected) between each unit class for each 1D single-channel feature, log scaled. Blank spaces in Duration heatmap result from no overlap between RS and NS distributions.



Supplemental Figure S28. Further analyses on unit class discrimination. **A**, Example UMAP projections of an equally subsampled set of units from each unit class, overlaid with unit class (left) and

cluster ID (right). **B**, Percentage of units found within each cluster through iterative application of UMAP on a subsampled dataset. Clustering was performed via DBSCAN and constrained to a maximum of 10 clusters. Error represents the 95% confidence interval about the mean. **C**, Average correlation of units found across all clusters in **B** (left) and the p-value of those correlation (right). **D**, Accuracy scores for optotagged-only classification. Bars represent mean scores $\pm$ SEM. **E**, Confusion matrices for optotagged only classification for each feature set. **F**, True Positive (TP) differences between NP Ultra and NP 1.0-like classification for 1-ch and 1-ch+footprint feature sets. **G**, Accuracy of 1-ch waveform classification when all units are used as samples. **H**, Confusion matrix of the classification performed in **G**.



Supplemental Figure S29. Detailed results for balanced and proportional classification of individual interneuron cell types. **A**, Classifier recall (left), precision (middle), and balanced accuracy (right) for each balanced interneuron classification. The classifier was logistic regression using balanced samples and all features. Darker bars indicate results from classification using all features, lighter bars indicate results using the spike duration metric alone. Recall measures the model's ability to find all positive instances, precision measures the accuracy of positive predictions, and balanced accuracy accounts for the imbalance in positive and negative classes. TP: true positive; TN: true negative; FP: false positive; FN: false negative. **B**, Schematic of the classification scheme using sample unit proportions approximately equal to that found in the cortex (Tremblay et al., 2016). **C**, Confusion matrices for each interneuron cell type classified via logistic regression using proportional samples and all features. UT = Untagged, OT = Optotagged. **D**, Classifier recall (left), precision (middle), and balanced accuracy (right) for each proportional interneuron classification.