

UCLA

UCLA Electronic Theses and Dissertations

Title

The contribution of prefrontal neuronal populations to the encoding of rule-related stimulus information during adaptive decision making

Permalink

<https://escholarship.org/uc/item/4503m8vj>

ISBN

9798190937442

Author

Vormstein-Schneider, Douglas

Publication Date

2026-09-11

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

The contribution of prefrontal neuronal populations to the encoding of rule-related
stimulus information during adaptive decision making

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy in Neuroscience

by

Douglas Carlton Vormstein-Schneider

2026

© Copyright by

Douglas Carlton Vormstein-Schneider

2026

ABSTRACT OF THE DISSERTATION

The contribution of prefrontal neuronal populations to the encoding of rule-related stimulus information during adaptive decision making

by

Douglas Carlton Vormstein-Schneider

Doctor of Philosophy in Neuroscience

University of California, Los Angeles, 2026

Professor Peyman Golshani, Chair

ABSTRACT

Cognitive flexibility, the capacity to update behavior when the relevance of environmental cues changes, depends on the medial prefrontal cortex (mPFC)¹²³⁴. However, how prefrontal populations represent stimulus information according to behavioral relevance, and how distinct neuronal classes contribute, are not yet fully understood. This dissertation examines how the mPFC represents rule-related information during an attentional set-shifting task, whether that representation reorganizes across rule switches, and whether somatostatin-expressing (SST) interneurons contribute differently from the non-SST population. Head-fixed, water-restricted mice performed a two-alternative forced-choice task with multiple un-cued rule shifts within single sessions,

while two-photon calcium imaging captured mPFC population activity with SST neuron identity recorded. Using dimensionality reduction and linear support vector machine decoding, we reveal three primary findings. First, the mPFC preferentially encodes the relevant over the irrelevant stimulus when the animal is expert at the current rule, an effect present in both SST and non-SST populations and stronger than raw stimulus encoding. Second, relevant-stimulus decoding is enhanced before a switch and weakens in the trials following the switch, while irrelevant-stimulus decoding shows no comparable pre-switch enhancement. Third, over the course of learning a new rule, information about the newly relevant stimulus gradually reoccupies a population subspace previously used for the formerly relevant stimulus. This occurs despite the two being of different sensory modalities. Meanwhile, irrelevant information does not undergo this transformation. Cell-type comparisons did not reveal strong differences in stimulus encoding. Together, these results support the idea that the mPFC organizes stimulus information by behavioral relevance rather than raw sensory modality. Findings are constrained by modest sample size and low statistical power, however our findings emerge despite these limitations suggesting a robust effect.

The dissertation of Douglas Vormstein-Schneider is approved.

Laura DeNardo

Dean Buonomano

Timothy Spellman

Peyman Golshani, Committee Chair

University of California, Los Angeles

2026

CONTENTS

ABSTRACT OF THE DISSERTATION	ii
ACKNOWLEDGEMENTS	vi
CV (Abbreviated)	vi
INTRODUCTION	1
Chapter 1 Calcium Imaging During Complex Mouse Behavior	12
Chapter 1.1 Mice learn to perform a two alternative forced choice attentional set- shifting task during neural recordings	12
Chapter 1.2 Mouse performance falls under discrete behavior states	22
Chapter 2 Extracting Complex Information From Prefrontal Neuronal Activity	26
Chapter 2.1 Choice and outcome decoding in discrete neuronal populations	26
Chapter 2.2 Event specificity in the firing of individual neurons	33
Chapter 2.3 Stimulus decoding in discrete neuronal populations	40
Chapter 3 Changes In Task Event Decoding Around Rule Switches	55
Chapter 3.1 Investigating informational changes immediately after a switch	55
Chapter 3.2 Investigating informational changes over many trials	61
DISCUSSION	67
METHODS	76
REFERENCES	86

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Professor Peyman Golshani for his support as the chair of my committee and a staunch supporter throughout the challenges of the project. His belief in my ability and level-headed approach to obstacles helped greatly in my progression through the program. Having an expert with an excellent tuition for drawing connections between data points was incredibly helpful in tying the results together. Thank you Peyman for all the support and guidance. I will always remember to “keep pushing”.

Second, I would like to thank my committee members: Tim Spellman, Laura DeNardo, and Dean Buonomano. Their guidance and orientation through the various trials that emerged over the course of the thesis was a valuable supplement to that of my advisor. I would like to extend additional thanks to Tim Spellman for making schematics, parts lists, and arduino code available for the behavioral rigs. The project would not have been possible without his contribution.

Next, I would like to thank the entire Golshani lab, past and present, for their support. Thanks to Blake Madruga for assisting in the engineering and customization of many of the physical components required for data collection, as well as for helping with the linear unmixing of two photon imaging signal. Blake Madruga and Conor Dorian were instrumental in training me in various aspects of calcium imaging, and Pingping Zhao contributed greatly to my learning of GRIN lens implants, all of whom I am very grateful to have worked with. I’d like to thank Arash Bellafard and Liron Sheintuch for their input into the best approaches in data analysis, and to Hesper Chen, Zoë Day, Samara Hecht, and Cai Ying for their help managing the many moving parts of the lab. Long Yang, Diego

Espino, Tamiris Prizon, Chrysa Iordanidou, Mina Shahi, Ronen Reshef, and everyone else in the lab for the interactions that fostered an environment for my professional growth.

A special thanks to everyone who worked regularly on the project. First and foremost to Zoë Day, whose brilliant mind and exceptional work ethic helped to push the project through its most trying times. In depth discussions with her on troubleshooting, optimization, and the philosophy behind the project characterize some of my favorite moments working on the project. Additionally, to Daniel Lee, who invested his whole undergraduate career to work on the project. I appreciate his persistence, hard work, and calm presence throughout the majority of the project. Finally, to the other undergraduate students that helped to drive things forward: Armen Babaian, Alex Gonzalez, Iman Zuberi, and Asher Adhoot.

I acknowledge the NSIDP graduate program and the UCLA Brain Research Institute, particularly Jenny Lee. Financial support was provided by the University of California, Los Angeles, the UCLA Neurotechnology Translation T32, and 5F31MH138093-02.

Research Experience:

PhD Candidate – University of California, Los Angeles

May 2021 –

Present

Principal Investigator: Peyman Golshani

Thesis Committee: Drs. Dean Buonomano, Laura DeNardo, and Timothy Spellman (UCONN)

Project Aim: Assessing the subtype-specific roles of cortical interneurons in regulating adaptive decision making

- Developed and optimized multiple 2 alternative forced choice head fixed behavior rigs.
- Designed and executed 2-photon calcium imaging experiments to record neuronal activity through a GRIN lens in head fixed mice.
- Performed optogenetic and chemogenetic experiments to investigate the sufficiency and necessity of functionally-distinct neuronal populations in a two alternative forced choice decision making task.
- Designed a custom Matlab pipeline for comparing metrics of behavioral performance across conditions, performing signal analysis of calcium traces, and decoding task information from neural activity.

Research Associate – Broad Institute of MIT and Harvard

August 2017 – June 2020

Principal Investigators: Jordane Dimidschstein, Gordon Fishell

Project Aim: Design and test subtype-specific enhancer AAVs for targeted neuronal labeling and manipulation.

- Selected and cloned hundreds of putative gene regulatory elements into AAV backbones.
- Systematically tested hundreds of AAVs through local cortical stereotaxic injections and systemic retro-orbital injections.
- Assessed viral specificity using immunofluorescence and in situ hybridization.
- Obtained and quantified histological colocalization images using confocal and widefield microscopes.
- Assisted in setting up and maintaining the infrastructure of a new lab, including managing a colony of over 200 transgenic mice.
- Profiled dozens of cell-type-specific small molecules for potential as drug-able targets.

Undergraduate Researcher – University of Vermont

January 2015 – May 2017

Principal Investigator: Sayamwong Hammack

Project Aim: Understanding the neuronal circuits that regulate stress, emotion, and resilience.

- Executed chronic variate stress behavioral assay on rats undergoing cocaine withdrawal
 - Performed deep brain cannula implantation surgeries in the Bed Nucleus of the Stria Terminalis for localized injection of Pituitary Adenylate Cyclase-Activating Polypeptide receptor antagonists
 - Managed a tissue processing pipeline for perfusing, cryo-sectioning, and nissl staining coronal brain sections
 - Performed deep brain stereotactic AAV injections into the Bed Nucleus of the Stria Terminalis and imaged axonal projections throughout the rat brain using widefield microscopy
-

Education:

University of California, Los Angeles; Los Angeles, CA
2020 – Present

September

Neuroscience Interdisciplinary Doctoral Program

Ruth L. Kirschstein National Research Service Award F31 (National Institute of Mental Health) Fellow, Neurotechnology Translation T32 Trainee, Summer Mentored Research Fellow

University of Vermont; Burlington, VT
– May 2017

August 2013

Major: Neuroscience (B.S.); Minor: Chemistry

Presidential Scholarship

INTRODUCTION

Cognitive flexibility and the demands of adaptive behavior

Survival in a changing world requires the ability to adapt in the face of changing relationships between cue, choice, and outcome. Animals must evaluate whether learned associations should be maintained, and when they should abandon a previously successful strategy in light of new information. This capacity to adjust behavior in response to shifting contingencies is termed cognitive flexibility, and is central to goal-directed behavior⁵⁶. Cognitive flexibility requires the extinction of learned behaviors as cue–reward contingencies change, and the reorganization of attention when the features that predict reward change⁷⁶⁸⁹.

The behavioral study of cognitive flexibility has a long history in the attentional set-shifting paradigm, a class of tasks that have been used to quantify executive function in humans for more than sixty years⁶¹⁰. In these tasks, a subject learns to attend to one of multiple stimuli while ignoring others and is eventually required to shift attention to a previously ignored stimulus. The development of set-shifting tasks for rats, and later for mice, extended this framework to species amenable to circuit-level interrogation, and much of that literature has focused on modeling the cognitive deficits associated with human pathologies¹⁰¹¹¹²¹³¹⁴. These rodent adaptations opened the door to asking whether the prefrontal cortex is necessary for flexible behavior and how populations of identified neurons contribute to the computations driving such behavior.

The clinical significance of understanding cognitive flexibility is considerable. Impairments in behavioral flexibility and impulse control are characteristics of disorders

with a prominent prefrontal component, including obsessive-compulsive disorder, autism spectrum disorders, and schizophrenia^{6,10,13,14}. Set-shifting tasks have been used to quantify cognitive deficits associated with schizophrenia, attention-deficit/hyperactivity disorder, and dementias related to Parkinson's, Huntington's, and Alzheimer's diseases^{6,10}. In schizophrenia in particular, cognitive deficits have come to be recognized as a core feature of the illness rather than a consequence of positive symptoms, and modern theories consider these deficits in terms of dysfunction in the cognitive control mechanisms supported by the prefrontal cortex^{13,14}. Establishing how prefrontal circuits normally support flexible behavior is a prerequisite for understanding how those circuits fail in disease.

The medial prefrontal cortex as a substrate for flexible cognition

Across species, the medial prefrontal cortex (mPFC) has a well-established role in the cognitive processes that underlie flexible behavior, from the updating of cue–reward contingencies to organizing goal-directed action^{15–19}. Its reciprocal connections with sensory, motor, limbic, and neuromodulatory systems place it in a position to integrate information about stimuli, actions, outcomes, and internal state, and to apply top-down control to structures that shape behavior^{20–25}. The extensive maturation period of the mPFC is thought to contribute to its ability to drive complex behaviors, but this prolonged window also makes these behaviors vulnerable to disruption²⁴. This is a common theme across the study of prefrontal contributions to memory, decision making, and mood, and a critical disease-relevant context for research in this region^{13,24}.

A central question about the mPFC concerns the nature of the information it represents and how that information is formatted. Single neurons in the prefrontal cortex can be difficult to interpret: rather than encoding a single task variable, they often respond to mixtures of many variables, producing heterogeneous response profiles^{16,20,26–32}. This property, termed mixed selectivity, has been recognized as computationally advantageous^{29,33,34}. Neurons that respond nonlinearly to combinations of task variables generate high-dimensional population representations, and these representations enable flexible encoding of information^{32,33,35–37}. The dimensionality produced by mixed selectivity is predictive of behavior, collapsing on error trials, which places population-level geometry, rather than individual neuron activity, as the primary means of analysis^{32,38}.

This population-level view has influenced how prefrontal function is studied. Rather than asking what a neuron encodes, questions are framed in terms of how task variables are represented in the high-dimensional space of population activity, and how that supports the computations required to perform complex behavior^{33,37}. Work in the primate prefrontal cortex has shown that context-dependent decision making can be understood as a dynamical process occurring at the level of the population, in which contextual signals integrate the relevant sensory inputs while suppressing the irrelevant ones¹⁸. The geometry of these representations is not fixed. It evolves with learning, and prefrontal representations can change from being high-dimensional, nonlinear, and randomly mixed early in learning to low-dimensional and rule-selective once a task is mastered^{7,34}. These findings promote an approach to the mPFC that treats task-related information as a

property of neural populations and asks how that information is transformed in the face of changing reward contingencies.

A recurring theme in this literature is that the mPFC does not simply represent sensory features but represents them in the context of their behavioral relevance. Prefrontal responses can reflect both the identity of individual cues and their meaning as task rules^{7,39}. When the relevance of a stimulus modality changes, prefrontal populations preferentially encode the currently relevant information while the encoding of irrelevant information is actively suppressed, a selection process that distinguishes the mPFC from primary sensory areas, where relevant and irrelevant stimuli are often represented indiscriminately^{18,31}. Attentional set-shifting tasks are uniquely suited to exploring this capacity to gate information according to context, and here we use one to ask a critical question in the field of cognition: how does the mPFC reorganize its representation of relevant and irrelevant stimuli when a rule shift occurs?

Distinct neuronal types and the division of labor in cortical circuits

Cortical networks are primarily built from glutamatergic excitatory neurons and a diverse population of local GABAergic inhibitory interneurons that gate signal flow and sculpt network dynamics^{40–42}. Although interneurons constitute roughly 25% of cortical neurons, they are highly heterogeneous, forming classes distinguished by their morphology, electrophysiology, molecular identity, connectivity, and activity profiles^{41,42}. A common perspective in the field is that this diversity reflects a division of labor: distinct interneuron types, each with a specific set of cellular properties and connectivity motifs, expand the computational power of a relatively small number of canonical circuit motifs^{41,42}.

Understanding how the prefrontal cortex supports flexible behavior requires attention not only to population activity but to the cell types underlying it.

The major interneuron classes are defined largely by molecular markers—parvalbumin (PV), somatostatin (SST), and vasoactive intestinal peptide (VIP). PV neurons typically provide fast perisomatic inhibition, VIP neurons frequently act to disinhibit the network by inhibiting other interneurons, and SST neurons target the dendrites of pyramidal cells, where they are positioned to control the integration of synaptic input^{41–43}. Direct comparison of these subtypes during behavior has revealed that they preferentially signal different task-related events. In the dorsomedial prefrontal cortex, for instance, SST neurons have been found to primarily signal motor action, VIP neurons to respond strongly to action outcomes, and PV neurons to respond less selectively across sensory cues, action, and outcome, while pyramidal neurons show functional heterogeneity that varies across cortical layers⁴⁴. Studies of probabilistic decision making have dissociated the contributions of PV and SST populations, suggesting a role for SST neurons in predicting the value of upcoming events and a role for PV neurons in value updating^{45–47}. These findings suggest that task-related information is not distributed uniformly across cell types, and that the cell-type composition of prefrontal populations is an important feature in its functionality.

SST-expressing neurons are of particular interest for the study of flexible behavior. As a major class of dendrite-targeting inhibitory cells, they are well positioned to shape the integration of inputs onto pyramidal neurons, and they display response plasticity during learning^{48,49}. SST output onto pyramidal neurons can undergo durable, target-specific reductions during sensory learning, and this plasticity is sensitive to the

relationship between stimulus and reward rather than to stimulus exposure^{50–53}. SST activity is also modulated by behavioral state, including arousal and attention, and recent transcriptomic work indicates that the class comprises multiple subtypes with distinct properties, raising the possibility that subtype-specific analysis will be necessary to fully understand their computational roles^{53–58}. At the same time, the importance of SST neurons extends to disease. Alterations in SST neurons have been found in the prefrontal cortex of individuals with schizophrenia, and these alterations are associated with measures of premorbid cognitive function⁵⁹. The convergence of these observations makes the SST population a compelling focus for asking whether a genetically identified interneuron class makes a distinct contribution to the encoding of task-related information during performance of a set-shifting task.

A core aim of this dissertation is to compare how SST interneurons and the broader non-SST population encode task events during set-shifting, and in particular whether the two populations differ in how they represent stimulus, choice, and outcome information across a rule switch. Because the same physical stimuli change their meaning when the rule changes, this offers a direct test of whether SST neurons encode rule-related information in a distinct manner from the surrounding excitatory and inhibitory populations.

Choice, outcome, and stimulus as task events in cognitive behavior

Effective performance of a cognitive task requires the simultaneous representation of several variables that determine a stimulus-outcome relationship. In a simple discrimination task, these variables minimally include the identity of the presented

stimulus, the choice the animal makes, and the outcome that follows. Even primary sensory cortex represents task variables beyond the sensory stimulus itself, including expectation, choice, and context²⁷, and the challenge of simultaneously representing many variables is a general feature of cortical coding^{27,35,60,61}. In higher-order regions such as the mPFC, the representation of stimulus, choice, and outcome is core to how flexible behavior is constructed, and each of these events has been shown to be decodable from prefrontal population activity^{8,60,62}.

Outcome, or trial feedback, is particularly important in this regard. A foundational body of work has demonstrated that the mPFC enables set-shifting by encoding trial feedback information, with this information organized as a gradient in which cells located in deeper layers are more strongly selective for feedback¹⁹. In dynamic and foraging environments, prefrontal neurons maintain representations of action value that bias subsequent choices, and the discrepancy between expected and obtained reward, the reward prediction error, serves as the teaching signal that updates these representations^{61,63–65}. Choice signals, in turn, emerge in prefrontal populations in coordination with the animal's developing commitment to an action^{4,66}, and stimulus signals are represented in a manner that depends on whether the stimulus is currently relevant to behavior^{18,67}. The way these three classes of task event are encoded, and how their encoding changes when a rule switch renders a previously relevant stimulus irrelevant, is a central focus of this thesis.

Behavioral state and the internal context of decision making

An animal's behavior on a given trial reflects not only the external stimuli and task rules but also an internal state that fluctuates over the course of a session. Classical models of perceptual decision making assume that an animal applies a consistent strategy, but recent work has shown this view to be incorrect⁶⁸. Choice reflects a complicated interplay among multiple interleaved strategies. These can be characterized as states in a hidden Markov model, that persist for tens to hundreds of trials before switching, often several times within a single session⁶⁸. These states typically comprise an engaged state, in which decisions depend heavily on the sensory stimulus, alongside biased or disengaged states in which errors occur frequently⁶⁸.

Behavioral state is relevant to the interpretation of neural data for two reasons. First, an animal's engagement determines the extent to which its choices are governed by task variables, and trials drawn from disengaged, biased states can defy assumptions applicable to trials from an engaged state. This consideration impacts how population decoding results should be interpreted. Second, brain state influences cortical activity, in part through the modulation of local inhibitory circuits^{23,56,58,69,70}, including SST neurons that have their activity shaped by transitions in arousal and engagement^{55,56,71,72}. This dissertation adopts a latent-state model of behavior to segment trials into engaged and disengaged states, characterizing the choice biases, performance, and response latencies underlying them.

Population approaches to decoding information in cortical circuits

The idea that task-related information is a property of neural populations rather than isolated neurons has driven a shift in analytical methods. Rapid growth in the number of

neurons that can be recorded simultaneously has made it possible to ask what can be learned from a population beyond what is available from each neuron individually, and dimensionality reduction has been used as a tool for revealing structure in population activity^{34,38,73,74}. By projecting high-dimensional activity into a small number of dimensions, methods such as principal component analysis allow for the isolation of important features enabling greater organization and simplified visualization of data.

Complementing these descriptive methods, decoding approaches ask how much information about a task variable is present in population activity by testing whether that variable can be predicted from the activity on a trial-by-trial basis^{19,75}. A linear decoder, such as a support vector machine, can be used to ask whether a given variable is represented in neural activity, and its performance provides an interpretable measure of the information available within the circuit. Applied with temporal precision, decoding can reveal not only whether information is present but when it emerges. Comparing decoders trained and tested across different task epochs can reveal how a representation is transformed as conditions change. In the present work, dimensionality reduction and support vector machine decoding are combined to characterize how stimulus, choice, and outcome information is represented in the mPFC, how that representation differs between SST and non-SST populations, and how it evolves across rule switches.

Encoding of latent variables: a feature of mPFC

Latent variables, such as value, context, and rule are not directly observable on a given trial, yet they shape how a stimulus should be interpreted and how a choice should be made. A major theme of prefrontal research has been to understand how these latent

variables are encoded in population activity, and how their encoding enables the animal to consider past experiences in the context of present decisions^{3,27,28,61,63,64,72}. A defining feature of the prefrontal cortex is its capacity to represent the current rule and to use that representation to determine how incoming stimuli are processed^{7,76}. In flexible decision making, an instruction cue can instruct neural activity into a rule-dependent configuration, so that the same stimulus evokes different patterns of neural activity^{18,77}. The relevant and irrelevant stimulus representations must therefore be disentangled from their raw sensory representations, and any reorganization of the population code across a switch can be attributed to the updating of rule rather than to a change in input. The analyses in this dissertation exploit this, using population decoding to ask how the mPFC represents the relevant rule, whether that representation differs between SST and non-SST populations, and how newly relevant information encoded as an animal learns a new rule, including whether it comes to reoccupy the subspace vacated by the stimulus that was relevant before the switch.

Overview of the dissertation

The chapters that follow apply this framework to a within-session attentional set-shifting task in head-fixed mice, using two-photon calcium imaging of the mPFC with somatostatin interneurons identified by a transgenically targeted fluorophore. Chapter 1 establishes the behavioral paradigm and imaging approach, demonstrates that mice perform multiple uncued rule shifts within single sessions, and characterizes the discrete behavioral states that structure their performance. Subsequent chapters extract stimulus, choice, and outcome information from prefrontal population activity using dimensionality reduction

and support vector machine decoding; compare the encoding of these events between SST and non-SST populations; examine the specificity of individual neurons for rule-related stimuli; and track how the decoding of task events changes across rule switches, both in the trials immediately following a switch and over the course of learning a new rule. The aim of this work is to understand how the mPFC represents the relevance of stimuli, how that representation is reorganized when relevance changes, and what distinct contributions discrete neuronal populations make to flexible behavior.

Chapter 1

Calcium Imaging During Complex Mouse Behavior

Chapter 1.1: Mice learn to perform a two-alternative forced choice set-shifting task during neural recordings

To investigate the circuit mechanisms underlying cognitive flexibility we leveraged a two-alternative forced choice attentional set-shifting task. This task requires a head fixed, water-restricted animal to choose one of two water delivering lick ports and report its choice with a lick during a 1.5 second response window. The correct spout is indicated by the presentation of a 2.5 second stimulus presentation (Figure 1.1.1). A correct choice results in a 3uL water reward being administered immediately upon choice execution. An incorrect choice results in the withholding of reward. The reward-indicating stimulus modality, either whisker stimulation frequency or odor presentation, consists of two possible stimuli: Stimulus A or stimulus B. If odor is the reward-indicating stimulus, also referred to as the relevant stimulus, the mouse will be presented with either odor A or odor B. Likewise, if whisker stimulation is the relevant stimulus, the mouse will be presented with either frequency A or frequency B. Trial initiation is signaled with a 500 millisecond white noise trial start cue. The stimulus is then presented for 2.5 seconds, after which the mouse reports its choice with its first lick during the response window. The response window ends immediately after the first lick, or after the full 1.5 seconds have passed. After the response window has closed, there is an inter-trial interval of a randomized time between 8 to 10 seconds before the next trial start cue is presented (Figure 1.1.2a-b). The task described here is known as Simple Discrimination (SD) and

is the first task among a series of increasingly difficult training tasks before a final Serial Extradimensional Shift (SEDS) task.

To learn SEDS, the mouse must first successfully master SD after which it will learn Compound Discrimination (CD), where the one of the two stimuli from the non-reward indicating stimulus modality is now presented alongside the relevant stimulus. This new stimulus is known as the irrelevant stimulus. The mouse must continue to follow the relevant stimulus despite the added presence of the irrelevant stimulus. Upon mastery of CD, the mouse is presented with various set shifts. First, the relevant stimuli A and B are replaced with new stimuli of the same modality: stimulus C and D. This is known as Intradimensional Shift 1 (IDS1). When the mouse learns the new stimulus-response relationships for water reward, it progresses to Reversal (Rev) sessions, where the side associations flip. Upon mastery of reversal, the mouse is exposed to its first extradimensional shift. Here the relevant stimulus modality is switched with the irrelevant stimulus modality, such that a previously reward-indicating odor modality is now non-reward indicating while the previously non-reward indicating whisker modality is the new relevant stimulus. After mastering this, a second Intradimensional Shift (IDS2) with the new relevant stimulus modality is learned before the animal progresses to SEDS (Figure 1.1.3a-g).

These set shifts across sessions train the mouse to a level of flexibility such that it can now recognize set shifts within single sessions. In SEDS, an extradimensional shift occurs in an uncued manner whenever a hidden performance criteria is reached. Upon reaching 80% performance in the last 40 trials, an extradimensional shift occurs and the relevant stimulus and irrelevant stimulus switch (Figure 1.1.2c). Upon this shift, the

mouse needs to use incongruent (ICG) trials to learn that the relevant stimulus modality has switched to become irrelevant, and that the irrelevant stimulus modality now indicates which spout to lick. ICG trials are trials in which the two stimuli presented indicate opposite spouts when they are the reward-indicating stimulus. For example, if odor A indicates the left spout when it is the relevant stimulus, and whisker frequency B indicates the right spout when it is the relevant stimulus, the two stimulus modalities are incongruent and the mouse must choose one stimulus modality to follow. Alternatively, both stimuli could indicate the same side, in which case the trial would be a congruent (CG) trial. Mice are able to use the ICG trials to learn the new relevant stimulus and can perform multiple un-cued rule shifts within a single session (Figure 1.1.2e). Mice perform the task with no preference for sensory modality, with average trials to reach the hidden performance criterion comparable across rules (Figure 1.1.3h). These results display a strong ability of the mice to learn and perform the complex cognitive flexibility task required to probe neuronal contributions to adaptive decision making.

To investigate the neural activity associated with this within-session learning, we performed two-photon calcium imaging of the mPFC during behavior once the mice had reached the ability to perform at least two shifts in a single session. To examine the distinct contributions of SST neurons, experiments were performed in transgenic mice. These mice express the dTomato fluorescent protein in SST-expressing neurons. This protein can be imaged simultaneously with the GCaMP8f calcium indicator by isolating signal from different ends of their emission spectra. Using GCaMP8f signal as a proxy for neural activity, we were able to record the activity of all neurons while retaining the identities of the SST interneurons (Figure 1.1.2d) for the duration of behavioral sessions.

In line with other findings in the mPFC literature (reference), we found individual neurons that had preferential firing to task events such as stimulus, choice, and outcome. Figures 1.1.2f displays neural activity from a single neuron during discrete trial types. Throughout the session, the neuron fires infrequently after the choice is reported on unrewarded response trials. However, the neuron displays increases in z-scored fluorescence shortly after response on rewarded trials. Similarly, a separate example neuron displays choice preference, with greater rises in fluorescence during anticipatory licking and after outcome presentation on right choice trials relative to left choice trials (Figure 1.1.2g).

Taken together, the success of our behavioral training and our ability to capture neuronal activity at the level of individual neurons set a strong foundation for exploring how genetically-distinct neuron classes contribute to the encoding of rule-related information.

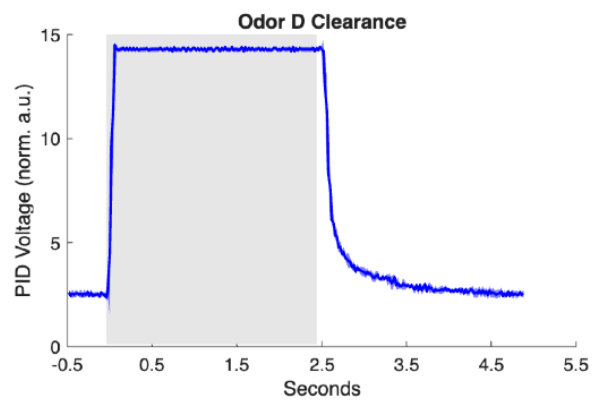
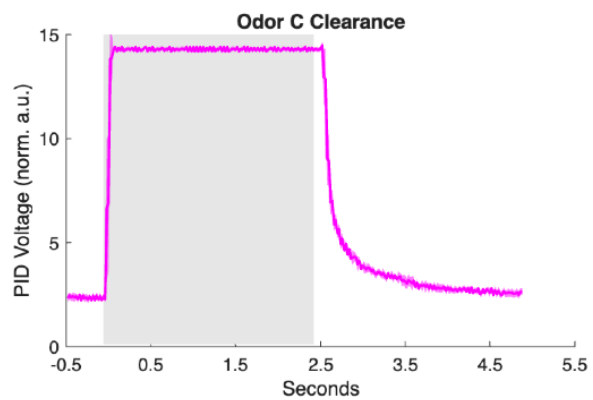
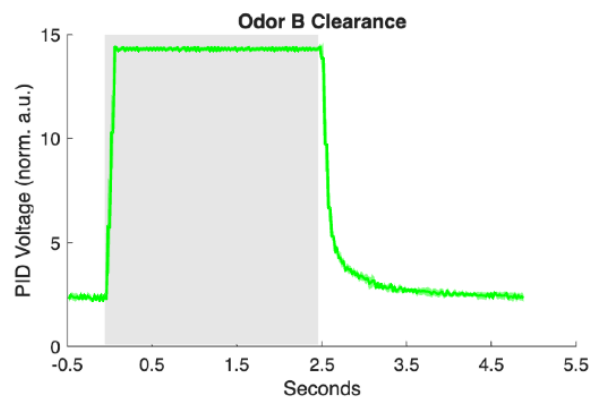
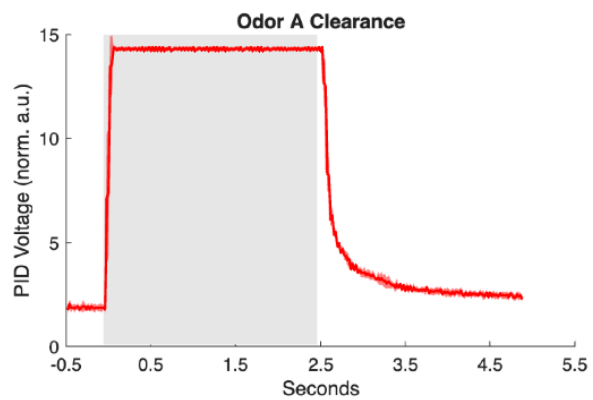


Figure 1.1.1: Odor release and clearance for discrete stimulus presentation

Measurements of odorant particles using PID measuring system throughout the first five seconds of trial structure. Traces for each odorant are an average of 5 'trials' of release. Grey shaded area indicates stimulus presentation window.

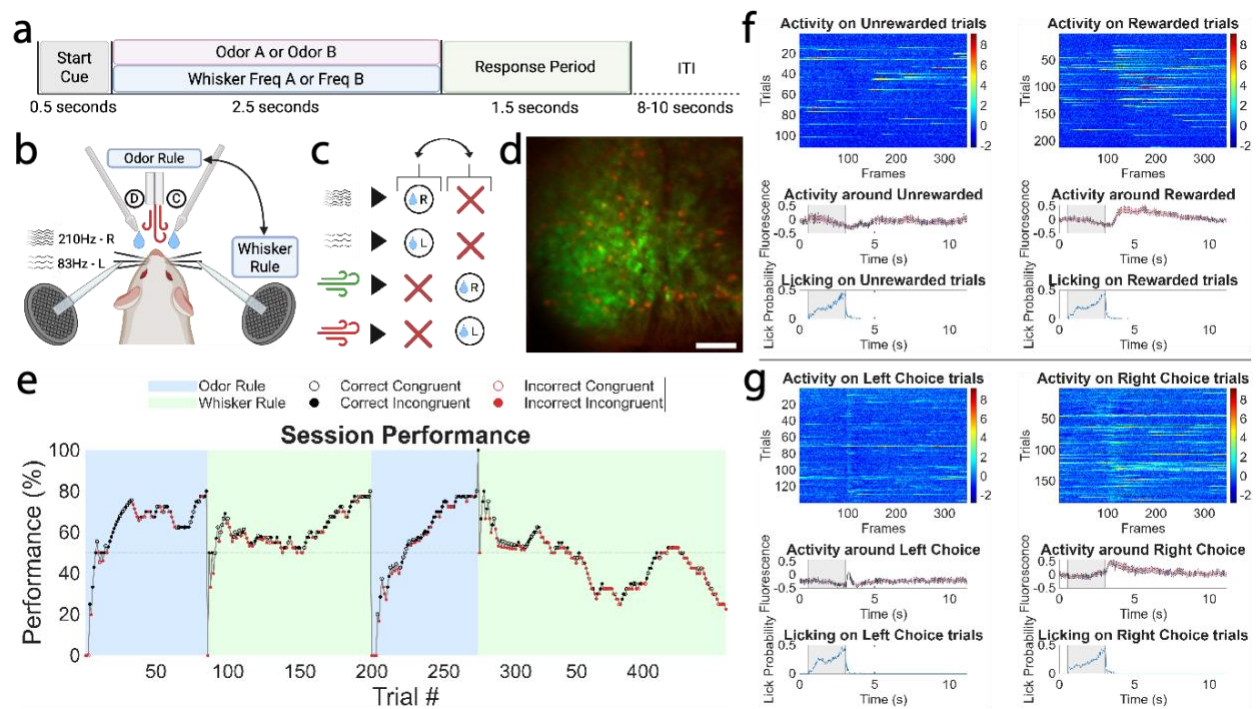


Figure 1.1.2: Imaging of mPFC population activity at the single neuron level during performance of an attentional set-shifting task

a) Schematic of trial structure. The response period is 1.5 seconds maximum, intertrial interval (ITI) is initiated upon completion of 1.5 seconds or after first lick in the response window. **b)** Schematic of behavioral setup. One of two odors are released from a single odorant spout. A left and right lick spout are both present. Speakers with funnels placed on mouse whiskers are used for whisker stimulus frequency presentation. Bilateral stimulation of whiskers with the same frequency indicates one spout, while bilateral stimulation of whiskers with the alternate frequency indicates the other spout. Rule-indicating stimuli switch between whisker and odor rule. **c)** Switches between whisker and odor rule occur such that one stimulus is the distractor while the other is rule-indicating. Upon a switch, the previously irrelevant / distractor stimulus becomes the rule-indicating / relevant stimulus and vice versa. **d)** Example mPFC imaging window using two photon excitation through a GRIN lens. **e)** Example performance throughout a single Serial Extradimensional Shift session. **f-g)** Heatmaps show trial-by-trial z-scored fluorescence on trials of the same event type. Middle plots show an average trace of neural activity across all trials of that type. Bottom plots show the probability of a lick occurring during that portion of the trial for that trial type.

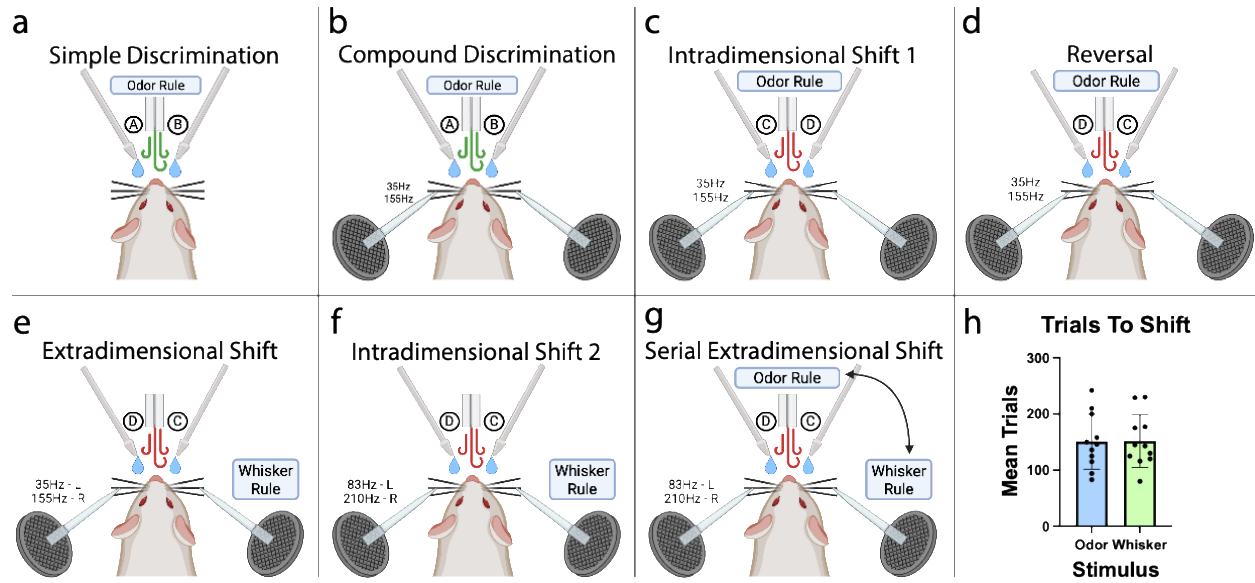


Figure 1.1.3: Progression of training to expert level on a serial extradimensional shift task.

a) Simple Discrimination: The mouse reports its choice with a right or left lick in response to either odor A or odor B. Odor A indicates the left spout is armed with a water reward while odor B indicates the right spout is armed with a water reward. **b)** Complex Discrimination: Irrelevant stimulus presentation is added in the form of a whisker stimulation distractor. One of two frequencies of stimulation is presented simultaneously throughout odor presentation. **c)** Intradimensional Shift 1: The relevant stimuli, odors A and B, undergo changes to two new odors, odor C and D. Odor C now indicates to lick left while odor D indicates to lick right. **d)** Reversal: Relevant stimulus side switches. Odor D now indicates to lick left while odor C now indicates to lick right. **e)** Extradimensional Shift: Relevant stimulus odor and irrelevant stimulus whisker now switch their rule-indicating and distractor natures. Whisker stimulus frequency becomes rule indicating while odor becomes irrelevant. **f)** Intradimensional Shift 2: The new relevant stimulus undergoes an intradimensional shift in which frequencies 1 and 2 are replaced by two new frequencies that are the new left and right indicating stimuli. **g)** Serial Extradimensional Shift: The testing stage of the behavior in which extradimensional switches between the two stimulus modalities occur in an un-cued manner multiple times within session. Shifts occur upon the mouse meeting a hidden performance criteria of 80% of the last 40 trials correct with a response rate of 80% or higher on the last 10 trials. **h)** Number of trials to reach hidden performance criteria does not differ when the mouse is performing under the odor rule versus when the mouse is performing under the whisker rule

Chapter 1.2: Mouse performance falls under discrete behavioral states

While mice were able to perform multiple switches within single sessions, not all trials reflect full engagement in the behavior. Mice fail to perform the task withing the paradigm of its rule structure. They may be disengaged, by means of being largely satiated or potentially for other reasons. To investigate the approach that the mice take to solve the task, we leveraged the work of Ashwood et al., 2022⁶⁸ to develop a GLM-HMM (Generalized Linear Model – Hidden Markov Model) to identify the behavioral states that the mice occupy while performing the task. The model was trained to identify 3 states using behavioral data from both training and testing behavioral sessions. Predictor weights for each of the states indicated distinctly different relationships to predicting a right choice on a present trial. One state, which we labeled the “engaged” state, placed roughly equal weights on each predictor, with the greatest weight coming from the stimulus for the present trial itself. The other two states, which we labeled the “right-biased” and “left-biased” states show strong but opposite weight on the lick bias constant predictor, suggesting that bias plays a greater role in determining choice than the other task-relevant predictors (Figure 1.2.1a). The proportion of right licks in each state provides further evidence for the disengaged states being defined by choice side bias, as the right-biased state showed a great proportion of right licks while the left-biased state showed an opposite directionality of equal magnitude (Figure 1.2.1b). Furthermore, a Shannon entropy analysis revealed that the engaged state showed 1 bit of choice entropy while both disengaged states showed 0.41 bits (Figure 1.2.1c). This further emphasizes the rigidity of choice in the left- and right-biased states as compared to the equal likelihood of left and right choices in the engaged state.

One feature of disengagement during behavior is poor task performance. While mice in the engaged state averaged 62.4% correct response trials, performance in both the left- and right-biased states was around 44% correct (Figure 1.2.1d). Typically on a two-alternative forced choice task the baseline performance would be roughly 50% correct, as there are only two possible choices. However, our task employed a lick bias correction feature such that bias towards one side over time initiates the increased presentation of opposite side trials. The existence of this mechanism to promote exploration highlights the prevalence of side bias in disengaged states, with mouse performance falling below 50%.

Disengagement and re-engagement occur on slower timescales than individual trials. Mice were prone to remaining in a single state for blocks of trials, with state switching occurring slowly. The likelihood of remaining in the current state on the following trial was well above 90% for all possible switch combinations (Figure 1.2.1e).

Independent of side choice reporting, other features of task engagement indicate discrepancies between states. An engaged mouse is likely to lick throughout the stimulus presentation period in anticipation of the response window. Disengaged mice may participate in anticipatory licking in a side-biased manner, but are also prone to lick less frequently. We analyzed the mean latency to choice reporting within the response window for each behavioral state across sessions. Mice reported choice with a shorter response time when engaged than when in disengaged states (Fig 1.2.1f). Together, these results highlight the complexity of mouse performance on the task, as well as confirm that strategic approaches must be taken in order to perform well.

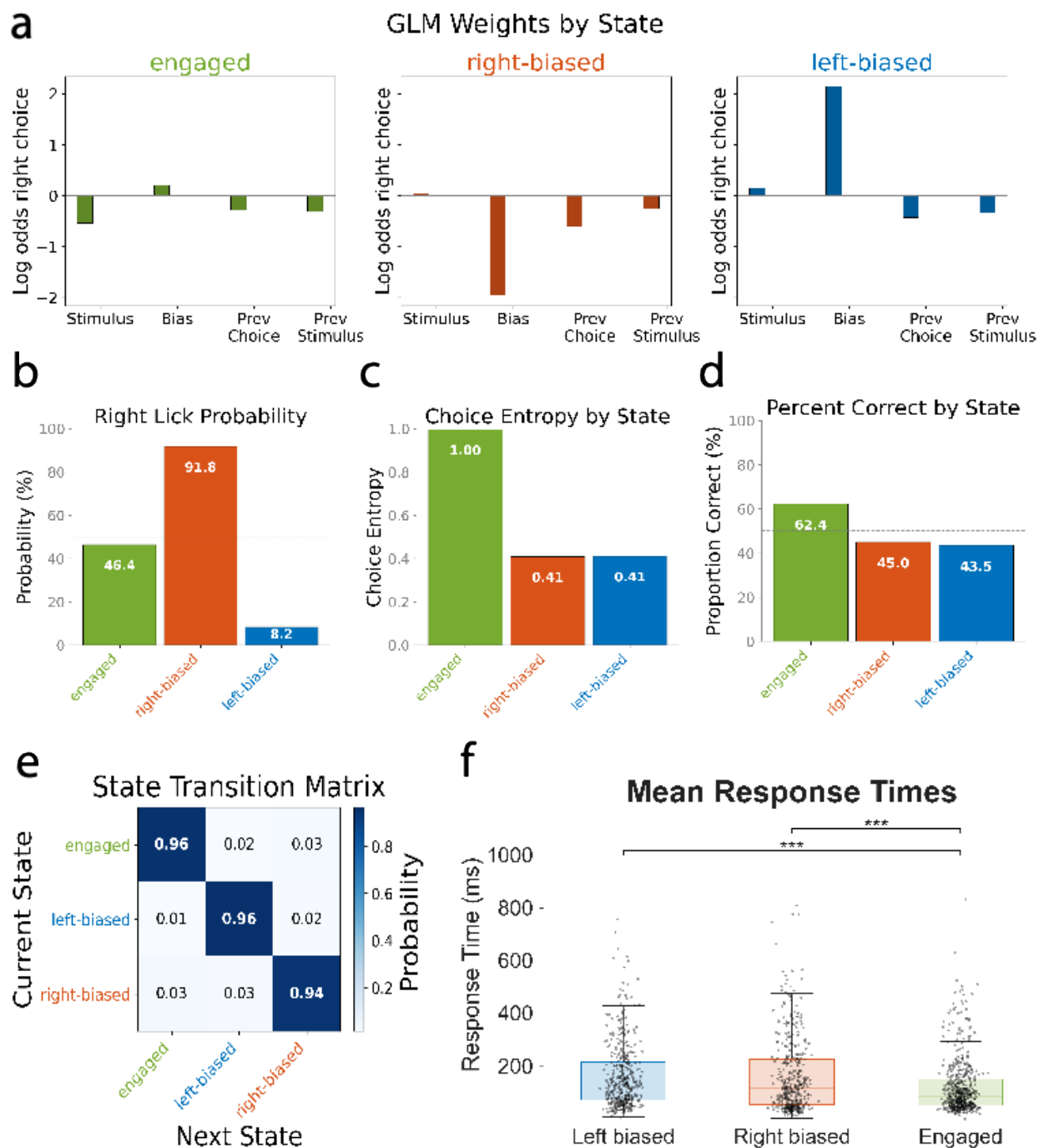


Figure 1.2.1: Discrete behavioral states during the performance of a two-alternative forced choice attentional set shifting task

a) GLM predictor weights relative to the prediction of a right lick on the present trial. **b)** Probability of right lick choice reporting on a given trial labeled as being of the state shown on the x axis. **c)** Shannon entropy in bits calculated for left versus right choice probability in a given state. **d)** Proportion correct trials in a given state. left- and right-biased states fall below chance due to lick bias correction mechanism presenting additional trials indicating the avoided spout. **e)** Transition matrix indicating the probability of switching to, or staying in, a behavioral state given the state on the present trial. **f)** Mean response time in milliseconds from initiation of response window to first lick for a given state. Dots indicate individual sessions. Animals are fastest to respond when engaged with the task.

Chapter 2

Extracting Complex Information From Prefrontal Neuronal Activity

Chapter 2.1: Choice and outcome decoding in discrete neuronal populations

There is abundant evidence that higher order cortical regions such as mPFC encode task-relevant information, such as choice and outcome, through population dynamics^{31,37,39}. We used principal component analysis (PCA) to reduce the dimensionality of our recordings, finding that trials could be segregated by task event when classified by neuronal activity in a low dimensional PCA subspace (Figure 2.1.1a). By leveraging the classification abilities of linear support vector machine (SVM) decoding, we were able to train models to classify trials by event type using population dynamics from neural activity (Figure 2.1.1b-c). For each group, we trained multiple SVM decoders to use calcium data from 500 millisecond windows within trials to discriminate between rewarded and unrewarded trials, or left choice and right choice trials. By tiling our window across the trial structure by leaps of 100 milliseconds, we were able to identify where task-relevant information peaks within the neural activity in mPFC. Decoding accuracy peaked within the first second after outcome presentation (Figure 2.1.1b-c) and we therefore focused our downstream analyses on this window of the trial. In order to look at equal numbers of trials before and after rule switches, we included disengaged trials in our analyses. Notably, this resulted in slight decoding confounds for non-stimulus task events, as same-side choice trials cluster in behavioral states with side bias. Similarly, disengaged trials tend to be unrewarded and these clusters of trials may allow for decoders to identify these events above chance early in the trial (Figure 2.1.1b-c).

However, decoding rises substantially as task events are reported within trial. Choice information emerges alongside anticipatory licking, and outcome decoding accuracy does not rise above baseline before outcome presentation (Figure 2.1.1b-c). With these caveats, we next investigated neural activity from trials before and after rule switches.

By isolating trials before a switch, when the mouse is an expert on the given rule, from trials after a switch, when the mouse is naïve to the new rule, we were able to look at how task-relevant information evolves with the presentation of new information. Comparing decoding from trials before a switch to decoding using trials after a switch, we found that encoding of these two event types in population activity across all neurons does not change significantly across a rule switch (Choice: Before 76.31 vs After 74.79% mean window accuracy, $p = 0.470$; Outcome: Before 66.73 vs After 62.25% mean window accuracy, $p = 0.258$; paired Wilcoxon, $n = 12$ sessions) (Figure 2.1.1d-e). Next, we investigated if decoders trained using balanced neuron numbers from SST and Non-SST populations performed differently for these event types. For choice, SST and Non-SST populations encode information comparably both before and after a switch (Before: SST 67.03 vs nonSST 68.13% mean window accuracy, $p = 0.519$; After: SST 62.65 vs nonSST 64.57% mean window accuracy, $p = 0.206$; paired Wilcoxon, $n = 12$) (Figure 2.1.1f). When comparing the magnitude of change in decoder performance from before a switch to after a switch, no difference was found between the two populations (Choice Before minus After: median nonSST 2.73 vs SST 3.74% mean window accuracy, $p = 0.910$; paired Wilcoxon, $n = 12$) (Figure 2.1.1g).

When decoding outcome information before a switch, SST and Non-SST populations performed similarly (Before: SST 56.68 vs nonSST 57.72% mean window

accuracy, $p = 0.622$; paired Wilcoxon, $n = 11$) . However, performance on trials after a switch revealed that decoding using activity from SST neurons was better able to discriminate between rewarded and unrewarded trials (After: SST 57.80 vs nonSST 54.55% mean window accuracy, median SST minus nonSST 3.13, $p = 0.032$; paired Wilcoxon, $n = 11$) (Figure 2.1.1h). Despite this, the magnitude of change across switches was not statistically significantly different between these two neuronal types (Outcome Before minus After: median nonSST 3.41 vs SST 0.21% mean window accuracy, $p = 0.123$; paired Wilcoxon, $n = 11$) (Figure 2.1.1i).

To look broadly at differences between the two neuronal groups, we took the trial average activity trace for each neuron and found the mean activity across all trials for the two groups. In none of the trial structure's 346 imaging frames did the mean activity traces significantly differ, including those during outcome presentation (Figure 2.1.2a-b). Together, these results provide some evidence for cell-type-specific encoding of outcome information upon altered reward contingencies.

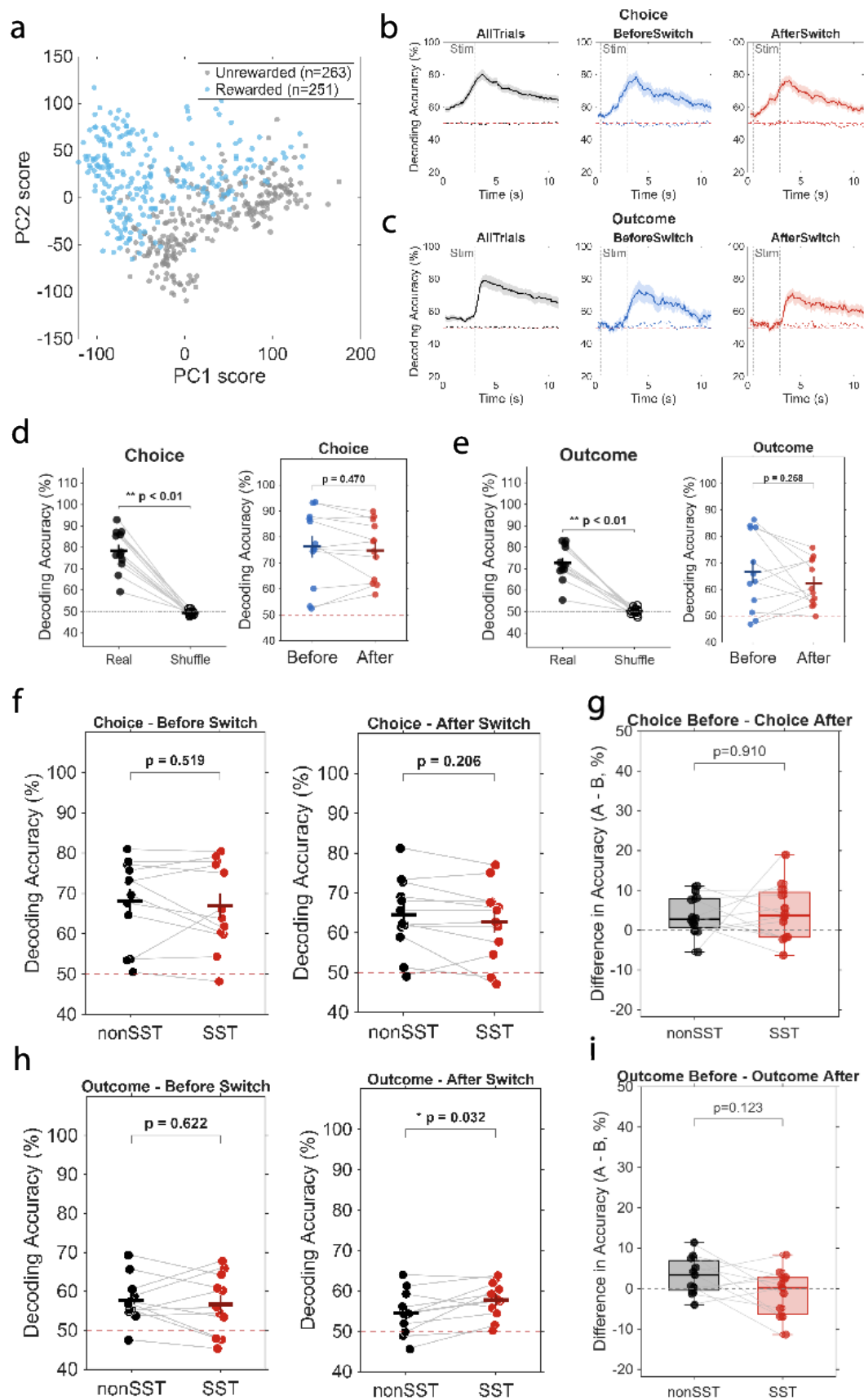


Figure 2.1.1: Decoding choice and outcome information from different populations of neurons during discrete trial groups

a) Principal component analysis plot showing each trial as a dot labeled by its trial type: Rewarded or unrewarded as indicated by dot color. Plot axes show value of that trial along the first and second principal components. **b)** Choice decoding performance from decoders trained and tested along 500 millisecond sliding windows across trial epochs. Training and testing performance is shown when training/testing on all trials, the 32 trials before a switch, and the 32 trials after a switch. Dotted lines indicate the stimulus presentation window of the trial. Decoding performance peaks after outcome presentation. **c)** Decoding performance for outcome as in panel c. **d)** Left: Decoding performance of choice during the 1 second window after stimulus presentation relative to decoding when trial identity (left or right choice) was shuffled relative to neural activity. Dots indicate decoding from individual sessions. Right: Decoding performance when training and testing using trials only within the 32 trials before or after each switch. Dots indicated decoding from individual sessions. **e)** As in d but for outcome decoding. **f)** Left: Choice decoding when training and testing using only trials within 32 trials before a switch and only neurons with a given identity indicated on the x axis. Neuron counts were balanced across decoders. Right: Same as left but with trials among the 32 after each switch. **g)** Comparison of the gaps between decoding of choice before and after a switch between SST and Non-SST neuron groups. Values reflect the difference in decoding before a switch versus after a switch. **h)** As in f but for outcome decoding. **i)** As in g but for outcome decoding.

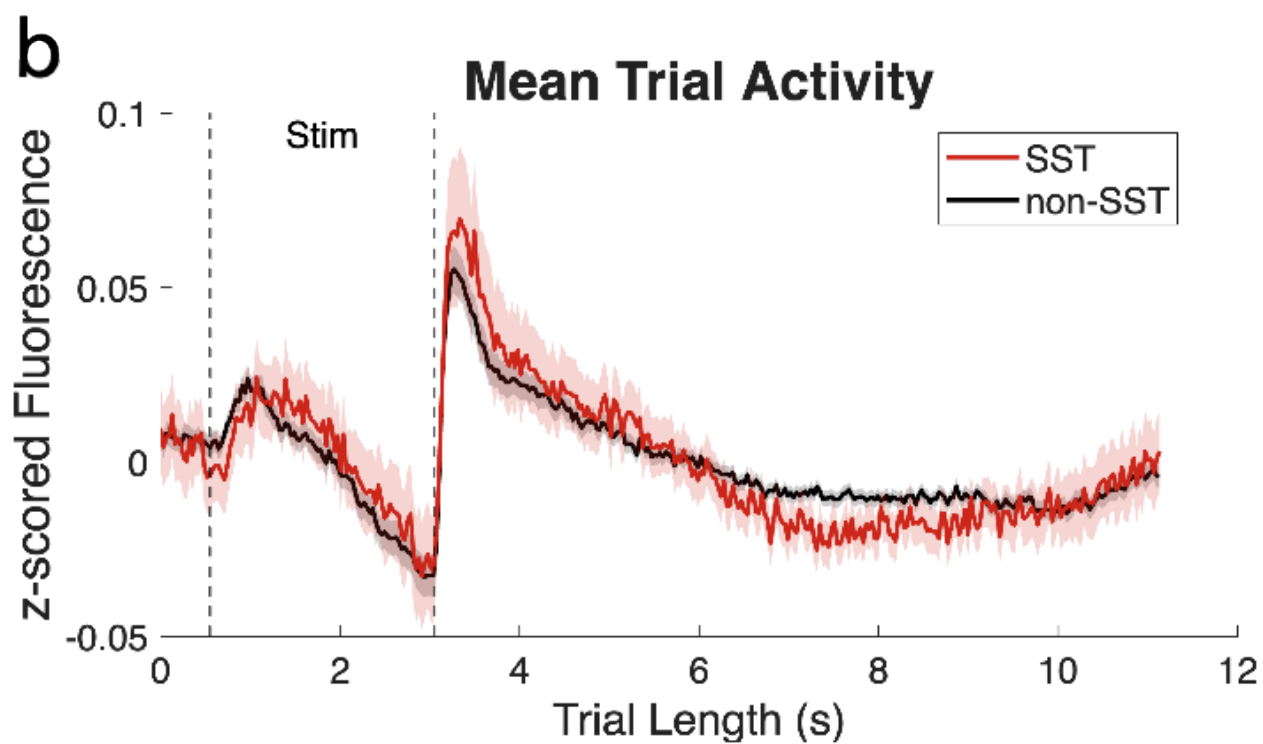
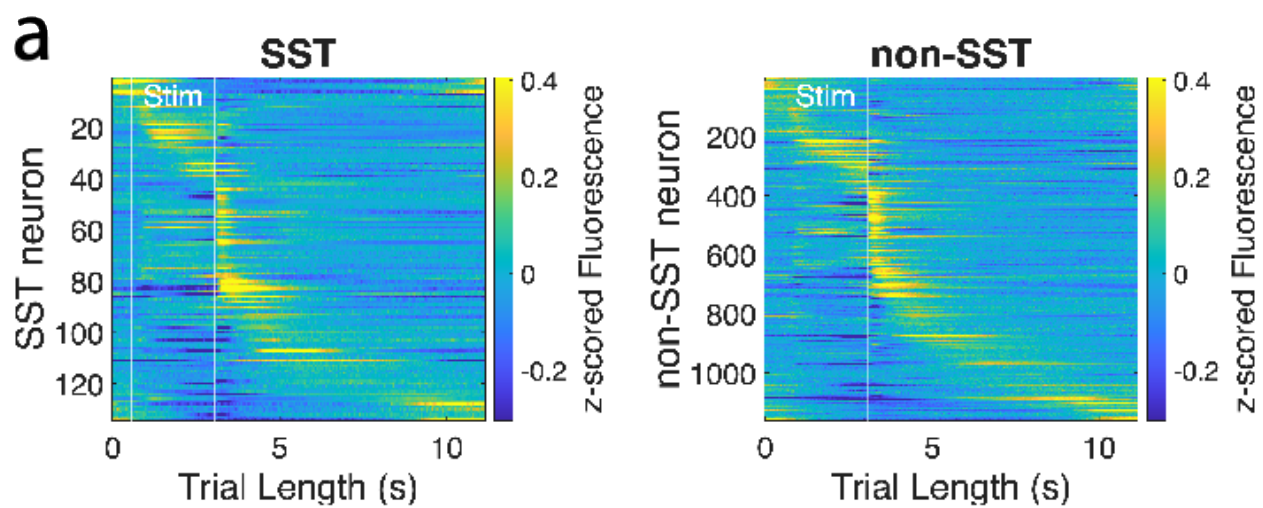


Figure 2.1.2: Trial-averaged comparison of neuronal activity in SST and Non-SST populations

a) Heat maps displaying the mean activity (z-scored fluorescence) of individual neurons (left: SST, right: Non-SST). **B)** Mean activity (z-scored fluorescence) across all neurons within a given neuron group overlayed.

Chapter 2.2: Event specificity in the firing of individual neurons

To investigate the contribution of individual neurons to the encoding of task-relevant information, we trained a Generalized Linear Model (GLM). Using task variables as predictors, and taking mean z-scored fluorescence during the peak decoding window, we were able to perform ordinary least squares analysis to identify whether a neuron had preferential firing for a given event type. For example, if a neuron had increased firing on left choice trials in comparison to right choice trials, the neuron would be considered a choice-specific neuron. We found that there was a large session-by-session variability in the proportion of event-specific neurons across event types. Further investigation revealed that the mean proportion of event-specific neurons across event types was positively correlated to the mean number of trials required to complete a set shift in a given session (Spearman $\rho = 0.671$, $p = 0.017$, $n = 12$ sessions; mean block length 139.0 ± 34.4 trials vs mean event-specific proportion $33.03 \pm 14.22\%$) (Figure 2.2.1a).

Following up on our decoding results, we looked into whether or not there were differences in the proportions of SST and Non-SST neurons that were choice or outcome specific (Figure 2.2.1b). While no significant differences were found for either event type (proportion event-specific, SST vs nonSST across sessions: Choice 49.61 ± 26.63 vs $49.09 \pm 22.38\%$, $p = 0.970$; Outcome 40.49 ± 27.57 vs $38.85 \pm 17.94\%$, $p = 0.677$; paired Wilcoxon, mean $\pm$ SD, $n = 12$), SST neurons have been implicated in error signaling⁷⁸. We therefore broke down the pools of outcome-specific neurons based on whether they had preferential firing for rewarded or unrewarded response trials. We found that out of the whole SST neuron population, a significantly higher proportion of these neurons were error-specific in comparison to those that were reward-specific (SST:

error 24.48 $\pm$ 16.57 vs reward 16.00 $\pm$ 13.71%, mean paired difference 8.48, $t(11) = 2.290$, $p = 0.043$; paired t-test, mean $\pm$ SD, $n = 12$ sessions) (Figure 2.2.1c). However, when a nested ANOVA was performed to remove mouse confounds the difference was not found to be significant (nested t-test, error 26.83 vs reward 16.46%, difference 10.37 $\pm$ 7.89 SEM, $p = 0.237$; 4 mice, 12 values) (Figure 2.2.1d).

Our GLM also allowed us to look at the proportions of event-specific neurons for other event types. Given the existing evidence for the mPFC containing rule-related information^{7,76}, we chose to look at the extent to which individual neurons had relevant or irrelevant stimulus preferences. We first analyzed all neurons together as a broader regional population. There were significantly more neurons that had preferential firing within the relevant stimulus modality than the irrelevant stimulus modality (All neurons: Relevant 32.97 $\pm$ 23.46 vs Irrelevant 11.33 $\pm$ 7.73%, median difference 19.96, $p = 0.0068$; paired Wilcoxon, mean $\pm$ SD, $n = 12$). This effect held up when performing a nested ANOVA to control for mouse confounds (nested t-test: Relevant 32.15 vs Irrelevant 11.62%, difference -20.52 $\pm$ 7.07 SEM, $p = 0.0082$; 4 mice, 12 values) (Figure 2.2.2a). When we divided the pan-neuronal population into SST and Non-SST groups, we found that the effect remained true across cell-types (nonSST: Relevant 32.65 $\pm$ 23.79 vs Irrelevant 10.92 $\pm$ 7.48%, $p = 0.0068$; SST: Relevant 35.70 $\pm$ 22.66 vs Irrelevant 13.47 $\pm$ 12.33%, $p = 0.020$; paired Wilcoxon, $n = 12$), indicating that both populations may be attuned to rule-related stimuli above distractor stimuli (Figure 2.2.2b-c). Nested ANOVA analyses within neuron group retained the significant difference (nested t-test: nonSST Relevant 31.76 vs Irrelevant 11.13%, $t(22) = 2.907$,

$F(1,22) = 8.452$, $p = 0.0082$; SST Relevant 35.30 vs Irrelevant 14.24%, $p = 0.0123$; 4 mice, 12 values each) .

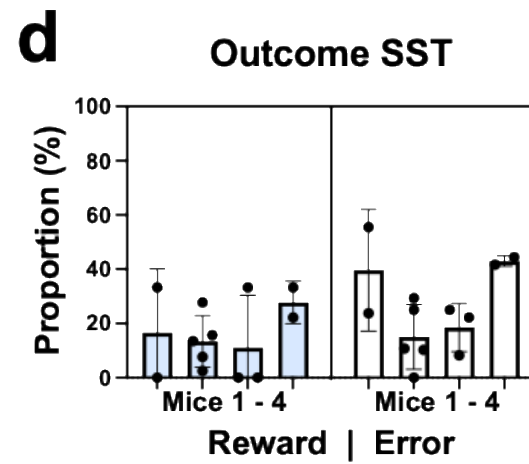
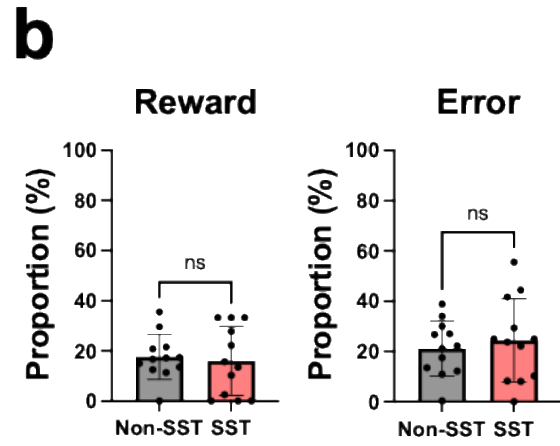
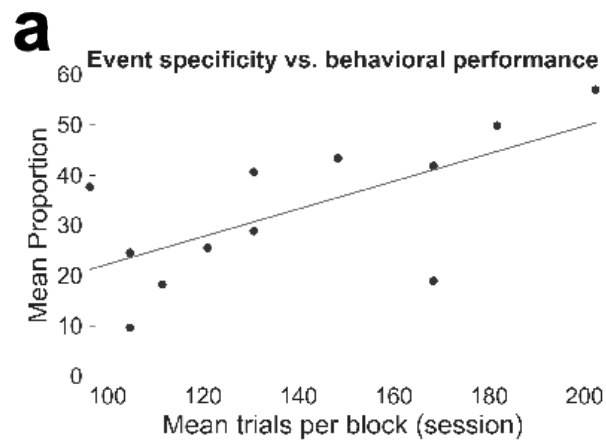


Figure 2.2.1: Single neuron event specificity proportions for outcome-related information

a) Plot indicating the relationship between the mean proportion of event specific neurons across all event types for a given session, and the mean trials within each rule block for a given session. **b)** Left: The proportion of SST and Non-SST neurons that have preferential firing to rewarded trials. Right: The proportion of SST and Non-SST neurons that have preferential firing to unrewarded trials. **c)** Left: The breakdown of Non-SST neurons that have preferential firing to rewarded versus unrewarded (error) trials. Right: The breakdown of SST neurons that have preferential firing to rewarded versus unrewarded trials. **d)** Nested ANOVA for preferential firing proportions of SST neurons for rewarded versus unrewarded trials. Columns indicate individual mice, with each dot representing one session. Results indicate no significant difference between rewarded and unrewarded proportions when accounting for variations across mice.

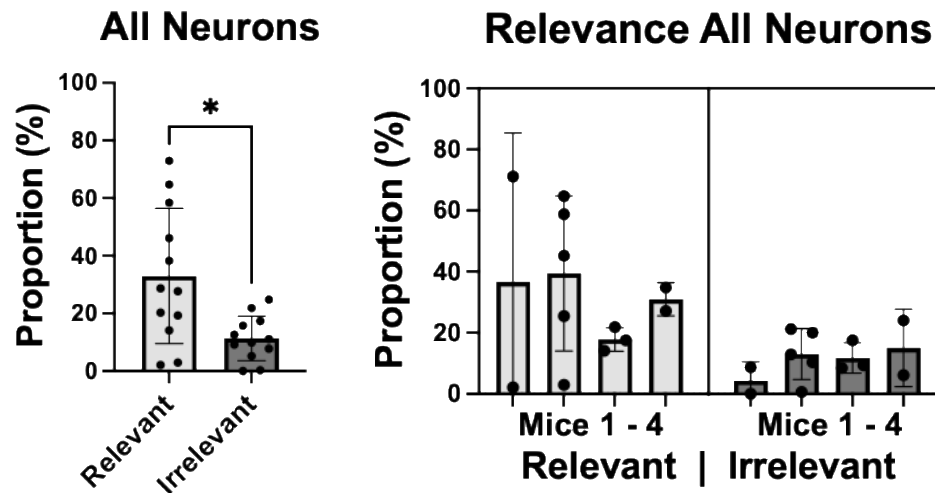
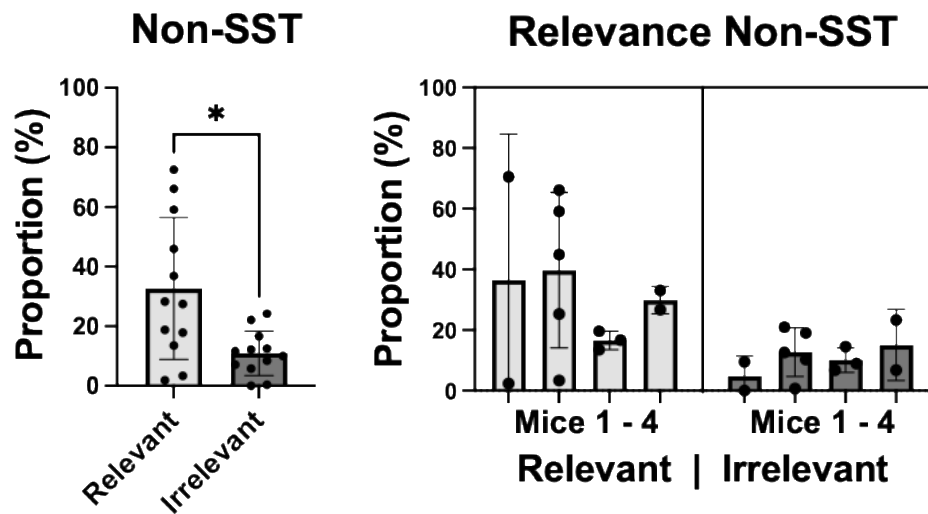
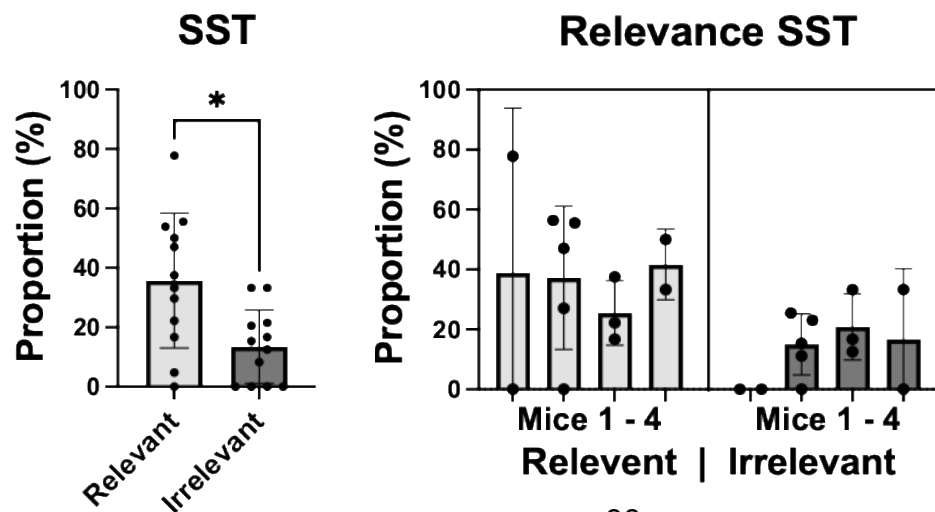
a**b****c**

Figure 2.2.2: Preferential firing of individual neuron populations for rule-related stimuli

a) Left: Proportions of neurons firing with preference to either right or left indicating relevant or irrelevant stimulus. Shows a higher proportion of neurons that have preference to left or right relevant stimulus relative to irrelevant stimulus. Right: Nested ANOVA analysis for four mice. Each column indicates data from one mouse, with each dot representing data from a single session. **b)** As in a, but using only Non-SST neurons. **c)** As in a and b, but using only SST neurons.

Chapter 2.3: Stimulus decoding in discrete neuronal populations

Following our findings of different single neuron preferences for rule-related stimulus information, we explored how population dynamics encoded whisker and odor stimuli. For both modalities, we were able to decode which of the two stimuli were presented on a given trial with significantly better performance than with shuffled data (all trials, All neurons: Whisker 66.59 vs shuffle 49.93% accuracy, $p = 0.00098$; Odor 65.52 vs shuffle 49.83%, $p = 0.00049$; paired Wilcoxon, $n = 12$) (Figure 2.3.1a&c). Furthermore, Odor decoding, but not whisker decoding, was significantly better on trials before a switch than on trials after a switch (Odor: Before 64.55 vs After 58.97% mean window accuracy, $p = 0.0068$; Whisker: Before 62.28 vs After 60.64%, $p = 0.424$; paired Wilcoxon, $n = 12$) (Figure 2.3.1a&c). This indicates preferential encoding of stimulus information when the mouse is attuned to the rule relative to when the mouse is naïve to the new stimulus-outcome relationship. Unlike with outcome decoding, no significant cell-type-specific differences were found for either stimulus modality before or after a switch (Figure 2.3.1b&d).

We next investigated how well SVM decoders would perform when blind to the stimulus modality. We trained decoders to discriminate whether the relevant stimulus, whether it be whisker or odor, indicated a left or right trial and found that we were able to decode well above chance (all trials, All neurons: Relevant 68.16 vs shuffle 50.08% accuracy, $p = 0.00049$; paired Wilcoxon, $n = 12$) (Figure 2.3.1e). Interestingly, decoding of relevant stimulus was significantly better when the mouse had learned of the rule-indicating nature of the stimulus than in the first trials after a switch (Relevant: Before 71.07 vs After 58.65% mean window accuracy, median After minus Before -12.19, $p =$

0.0049; paired Wilcoxon, $n = 12$) (Figure 2.3.1e). SST and Non-SST population decoding showed no differences in performance both before and after a switch (Relevant Before: SST 62.23 vs nonSST 63.53% mean window accuracy, $p = 0.424$; Relevant After: SST 54.78 vs nonSST 53.75%, $p = 0.569$; paired Wilcoxon, $n = 12$) (Figure 2.3.1f). While irrelevant stimulus decoding performed better than chance (all trials, All neurons: Irrelevant 57.67 vs shuffle 50.36% accuracy, $p = 0.00049$; paired Wilcoxon, $n = 12$), it did not display the same raised level of decoding before a switch relative to after a switch that was seen when decoding the relevant stimulus (Irrelevant: Before 57.91 vs After 58.56% mean window accuracy, $p = 0.910$; paired Wilcoxon, $n = 12$) (Figure 2.3.1g). Like the relevant stimulus, there were no observed differences in decoder performance when using SST or Non-SST populations. This was true both before and after a switch (Irrelevant Before: SST 52.38 vs nonSST 54.06% mean window accuracy, $p = 0.176$; Irrelevant After: SST 55.67 vs nonSST 54.62%, $p = 0.733$; paired Wilcoxon, $n = 12$) (Figure 2.3.1h).

With our ability to perform binary classification of stimuli using neural activity established, we moved on to examine if we could decode stimulus combination from the neural activity. With four possible combinations of stimuli presentations, we trained a one-versus-rest multiclass SVM decoder and found that we could decode trials better than with shuffled data (all trials: StimCombo 46.68 $\pm$ 9.60 vs shuffle 24.52 $\pm$ 0.82 mean window accuracy, median difference 22.03, $p = 0.00049$; mean 4-class accuracy 45.74%, chance 25%; paired Wilcoxon, mean $\pm$ SD, $n = 12$) (Figure 2.3.2a-b). When we compared the decoder's ability to perform in the above chance range relative to that of our relevant stimulus decoding, we found that relevant stimulus decoding performance

was significantly better than that of the stimulus combination decoder (all trials, proportion-of-distance-to-ceiling: Relevant 36.32 $\pm$ 12.93 vs StimCombo 28.91 $\pm$ 12.81, median difference 6.46, $p = 0.016$; paired Wilcoxon, $n = 12$) (Figure 2.3.2c-d). This provides support for the idea that stimulus information in the mPFC is predominantly represented in the context of rule.

To investigate this further, we looked at performance on whisker and odor decoding when they were the relevant versus irrelevant stimulus. For both stimulus modalities, decoder performance trended higher when they were the relevant stimulus on trials before a switch without reaching statistical significance (Before switch: Whisker Rel vs Irrel median 65.77 vs 55.82% mean window accuracy, $p_{\text{raw}} = 0.070$, $p_{\text{FDR}} = 0.141$; Odor Rel vs Irrel 67.90 vs 58.10%, $p_{\text{raw}} = 0.015$, $p_{\text{FDR}} = 0.059$; paired Wilcoxon, $n = 11$) (Figure 2.3.3a-b). This was not seen after a switch, when decoder performance was comparable across rule relevance for both modalities (After switch: Whisker Rel vs Irrel median 54.19 vs 64.56% mean window accuracy, $p_{\text{raw}} = 0.084$, $p_{\text{FDR}} = 0.336$; Odor Rel vs Irrel 58.66 vs 59.52%, $p_{\text{raw}} = 0.922$, $p_{\text{FDR}} = 0.922$; paired Wilcoxon, $n = 10$) (Figure 2.3.3a-b). Next we compared relevant stimulus decoder performance on trials where odor was the relevant stimulus versus whisker being the relevant stimulus. Decoding performance was similar between the two modalities both before and after a switch (Figure 2.3.3c). This was found to be true for both modalities in both trial epochs when decoders were trained to identify the irrelevant stimulus (Figure 2.3.3d).

We then turned our attention to how the mPFC encodes relevant and irrelevant information around a switch. Comparing decoding of relevant stimulus before a switch to irrelevant stimulus before a switch, we confirmed that mPFC population activity encodes

information on the relevant stimulus more thoroughly than information on the irrelevant stimulus (All neurons: Relevant-Before 71.07 vs Irrelevant-Before 57.91% mean window accuracy, $p = 0.00049$ vs shuffle for each; direct Rel-vs-Irrel median difference nonSST 9.00, SST 7.54 mean window accuracy) (Figure 2.3.4a). When comparing the differences between the decoding for the two stimuli using SST or Non-SST neurons, we found no significant difference between the two populations (Rel-Before minus Irrel-Before: median nonSST 9.00 vs SST 7.54% mean window accuracy, $p = 0.970$; paired Wilcoxon, $n = 12$) (Figure 2.3.4b). Decoding of the relevant stimulus before a switch versus after a switch showed better encoding of the relevant stimulus before a switch as opposed to the previously irrelevant stimulus, now the relevant stimulus, after a switch (All neurons: Relevant Before 71.07 vs After 58.65% mean window accuracy, $p = 0.0049$; paired Wilcoxon, $n = 12$) (Figure 2.3.4c). Here too we saw no significant difference between the gap between these two decoders when using SST or Non-SST neuron groups (Rel-Before minus Rel-After: median nonSST 9.40 vs SST 3.98% mean window accuracy, $p = 0.301$; paired Wilcoxon, $n = 12$) (Figure 2.3.4d).

Challenges in interpreting decoding results can arise when multiple event-types co-vary. In trials before a switch, when the mouse is regularly performing at or above 80% correct choices, left and right choice labels for these trials will co-vary greatly with those for the relevant stimulus. To search for confounds in before switch decoding, we first checked baseline decoder performance between those trained and tested on the relevant stimulus versus those trained and tested on choice (Figure 2.3.5a). We found that choice decoding accuracy was significantly higher (Rel Mean 70.88%, choice mean 76.73%, $p = 0.034$; paired Wilcoxon, $n = 12$). To assess covariance, we trained decoders

to differentiate left and right relevant stimulus and tested them on relevant stimulus as well. This performance was compared to decoder performance when training on the relevant stimulus and testing for ability to discriminate left and right choice (Figure 2.3.5b). Testing for choice showed significantly higher performance than when testing for relevant stimulus (Test Rel Mean 70.88%, Test choice mean 79.28%, $p = 0.005$; paired Wilcoxon, $n = 12$). This demonstrates strong evidence for confounds in before-switch decoding. We also checked how training with relevant trials and testing on choice trials compared to decoding performance when both training and testing on choice (Figure 2.3.5c). Interestingly, while there was no significant difference, accuracy when training with relevant stimulus labels showed a slight trend for higher than performance when trained on choice trials (Train Rel Mean 79.28%, Train Choice mean 76.73%, $p = 0.38$; paired Wilcoxon, $n = 12$).

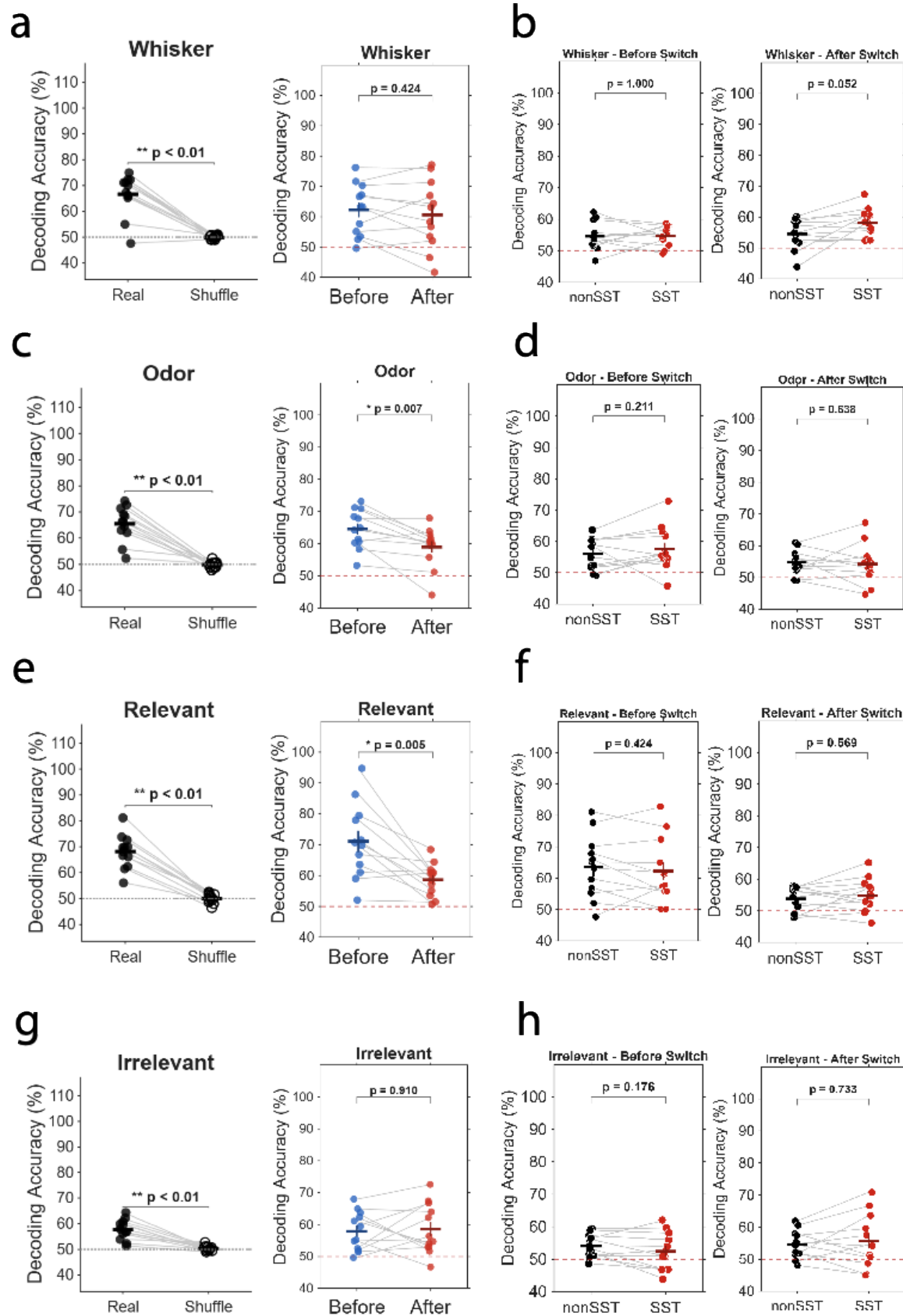


Figure 2.3.1: Decoding of stimulus information from distinct trial groups.

a) Left: Decoding of whisker frequency 1 versus whisker frequency 2 during the 1 second window after stimulus presentation as compared to the same but with stimulus presentation labels shuffled. Right: Decoding of whisker stimulus frequency when training and testing on the 32 trials before versus after a switch. **b)** Decoding performance at discriminating frequencies 1 and 2 when trained and tested using balanced numbers of Non-SST and SST neurons. **c)** As in a, but with odor. **d)** As in b, but with odor. **e)** As in a, but with relevant stimulus. **f)** As in b, but with relevant stimulus. **g)** As in a, but with irrelevant stimulus. **h)** As in b, but with irrelevant stimulus.

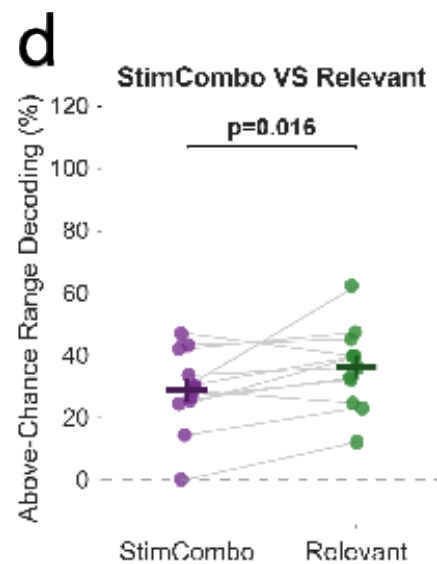
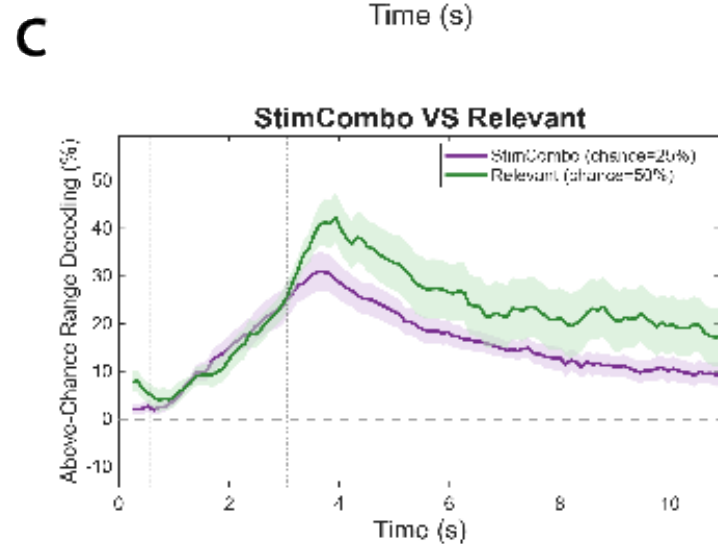
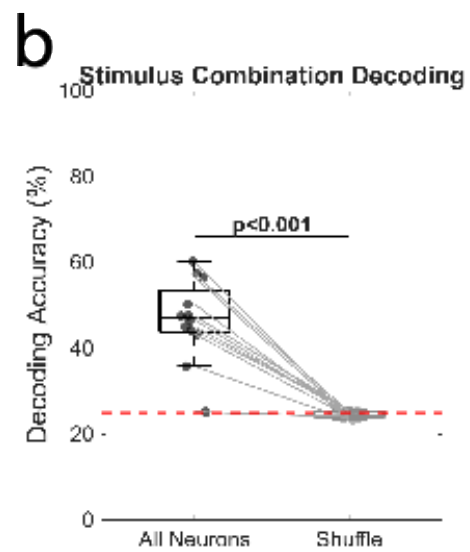
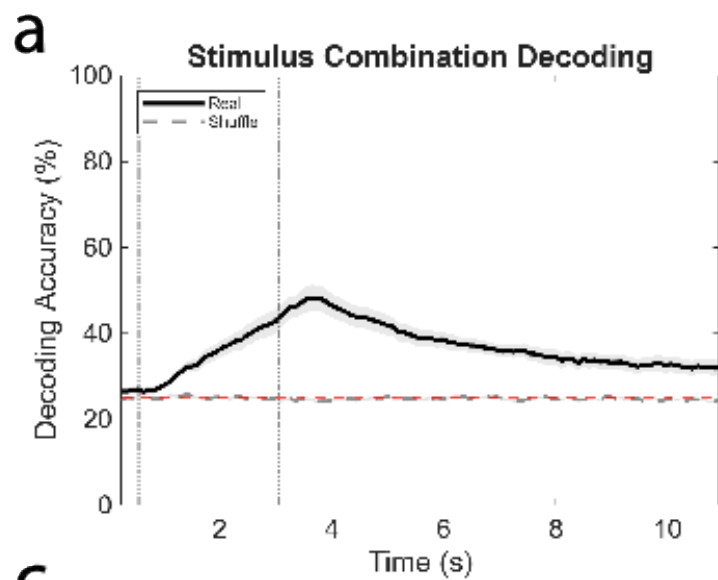


Figure 2.3.2: Stimulus combination decoding

a) Tiled decoding of stimulus presentation combination decoding across trial time bins. Chance at 25% indicates the 4 possible combinations across the two presented modalities. **b)** Stimulus combination decoding relative to trial label shuffled data. **c)** Overlay of relevant stimulus decoding and stimulus combination decoding aligned to the range from chance to 100% decoding performance. X axis represents decoding across 100 millisecond tiles with a 500 millisecond window. **d)** Comparison of relative above chance decoding in the 1 second window following stimulus presentation.

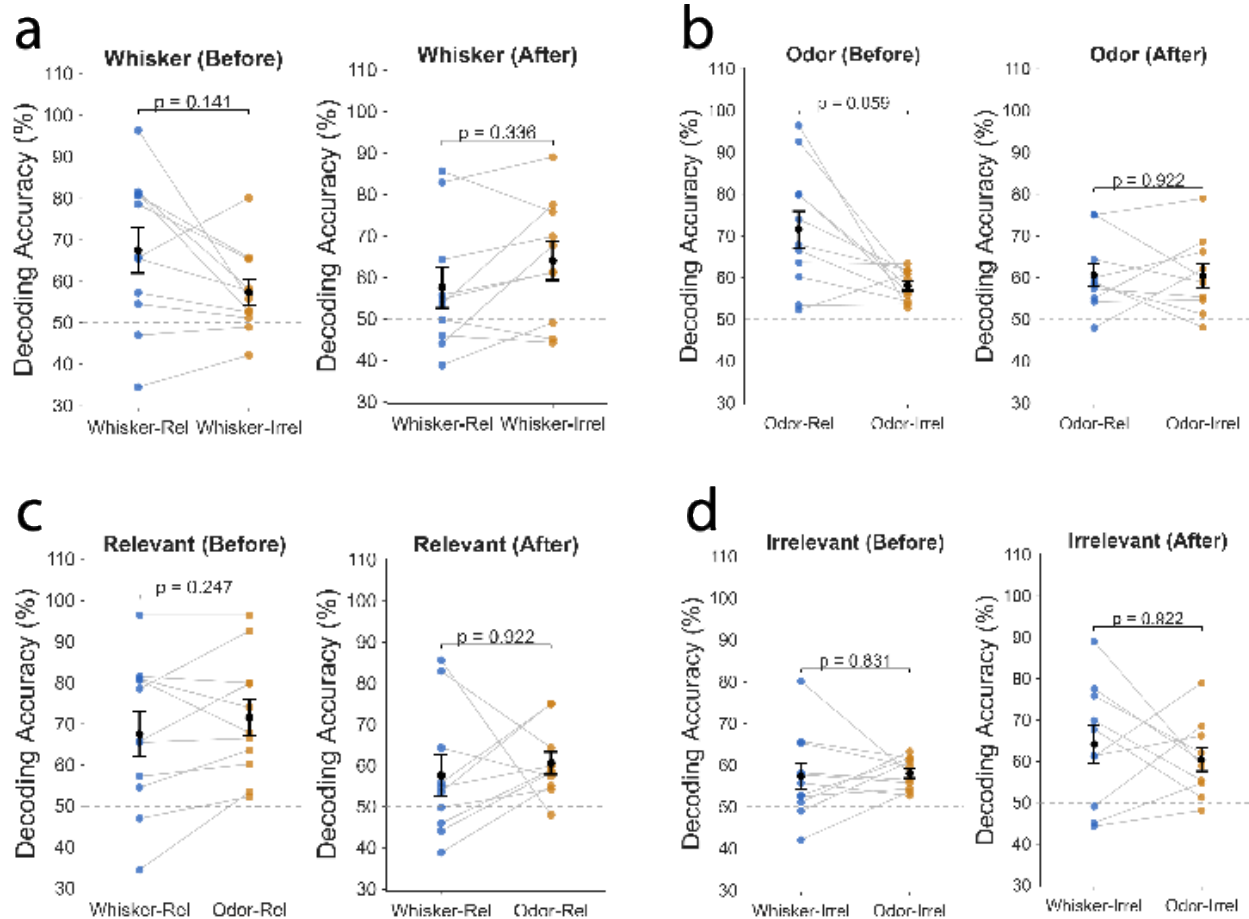


Figure 2.3.3: Whisker and odor decoding in rule-related contexts

a) Comparison of whisker decoding when it is the relevant stimulus versus when it is the irrelevant stimulus. Left: Trained and tested on 32 trials before and after switches. Right: Trained and tested on 32 trials before and after switches. **b)** As in a, but for odor decoding. **c)** Comparison of whisker and odor decoding when they are the relevant stimulus. Left: Trained and tested on 32 trials before and after switches. Right: Trained and tested on 32 trials before and after switches. **d)** As in c, but for irrelevant stimulus.

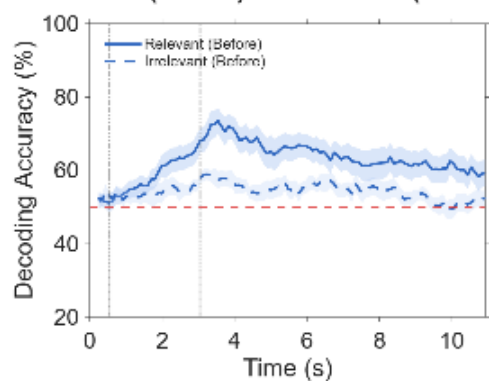
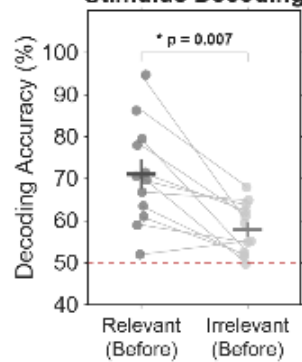
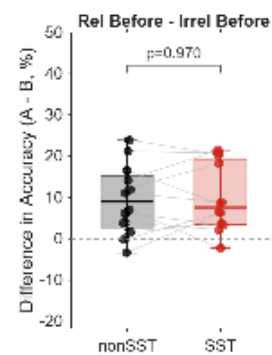
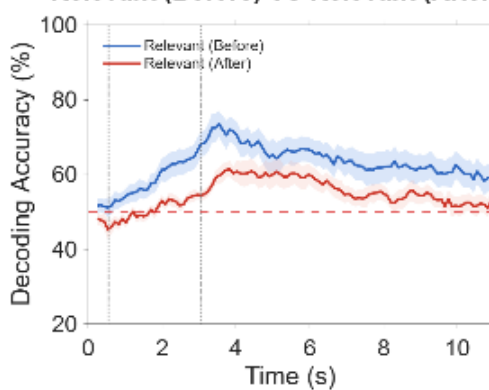
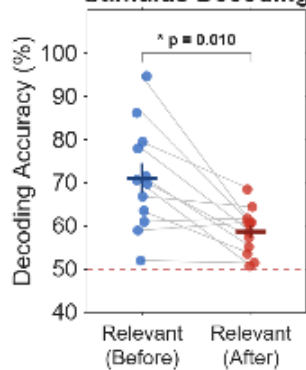
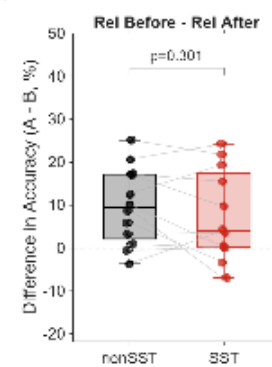
a**Relevant (Before) VS Irrelevant (Before)****Stimulus Decoding****b****c****Relevant (Before) VS Relevant (After)****Stimulus Decoding****d**

Figure 2.3.4: Relevant stimulus decoding across a switch with an eye to contributions of discrete neuronal groups

- a)** Left: Sliding window of 500ms decoding tiled every 100ms across trial structure for relevant and irrelevant stimulus from the 32 trials before a switch. Right: Peak decoding comparison for each taken from the 1 second after stimulus presentation. **b)** Comparison of the gap between relevant and irrelevant stimulus decoding before a switch when trained and tested using equalized counts of Non-SST neurons versus SST neurons. **c)** As in a, but with relevant stimulus decoding before and after a switch. **d)** As in c, but with relevant stimulus decoding before and after a switch.

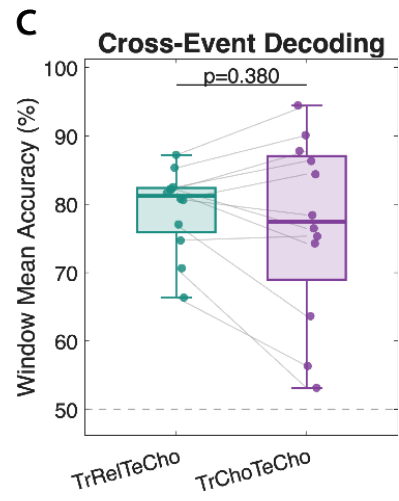
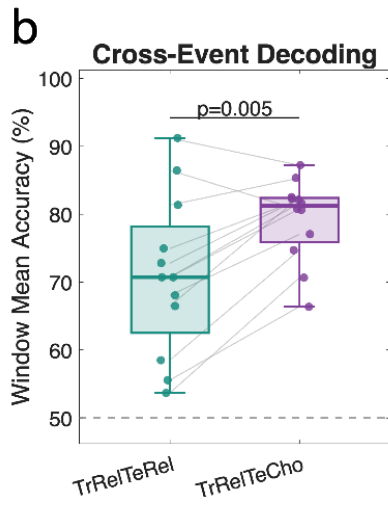
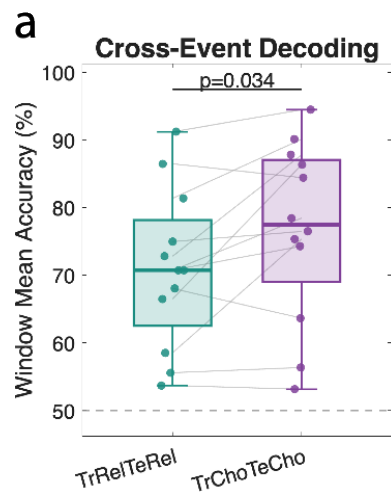


Figure 2.3.5: Confounds in before switch decoding of co-varying event-types

a) A comparison between the performance of two decoding paradigms on trials before a switch. Here decoding performance is compared between a decoder trained on relevant stimulus and tested on relevant stimulus, versus a decoder trained on choice and tested on choice. **b)** As in a, but the data on the right is from decoders trained on relevant trials and tested for discriminating choice. **c)** As in a, but the data on the left is from decoders trained on relevant trials and tested for discriminating choice.

Chapter 3

Changes In Task Event Decoding Around Rule Switches

Chapter 3.1: Investigating informational changes immediately after a switch

Previous analyses compared performance using decoders to classify within event type trained on trials within the trial group being tested. However, this does not identify the extent to which neural activity patterns truly shift across a switch. By training decoders on a subset of trials before a switch, then testing on the held-out trials before a switch as well as on an equal number of trials after a switch, we can identify the extent to which neural activity patterns are still able to classify event information as new rule-related feedback is presented to the animal (Figure 3.1.1a).

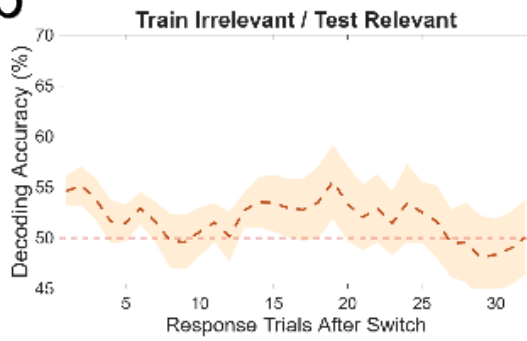
We first assessed the extent to which decoding performance drops for non-rule-related event types. While there was some decay in decoding ability, choice, outcome, whisker, and odor decoding did not drop significantly between testing on trials before versus after a shift (Figure 3.1.2a-d). We then explored how decoding rule related information changed across a shift by training on irrelevant trials before a shift, then testing on irrelevant trials before the shift and relevant trials after the shift. This allowed us to test decoding of the same stimulus modality before and after a switch under different rule conditions. Decoding of previously irrelevant stimulus remained low when testing after a switch for the now relevant stimulus (Figure 3.1.1b). However, there was still a significant decrease from decoding of that stimulus before a switch (All neurons, train irrel-before: test-before 60.68 vs test-after 50.20% mean window accuracy, median drop 11.83, $p_{\text{raw}} = 0.027$; paired Wilcoxon, $n = 12$) (Figure 3.1.1c). In the case of decoding the previously relevant stimulus, decoding of the now irrelevant stimulus decayed over

the course of the 32 trials after the switch (Figure 3.1.1e). This led to an overall decrease in decoding performance when testing before versus after a shift (All neurons, train rel-before: test-before 77.06 vs test-after 60.65% mean window accuracy, median drop 18.64, $p_{\text{raw}} = 0.00049$; paired Wilcoxon, $n = 12$) (Figure 3.1.1f). In both cases, decoding with Non-SST neurons trended towards a greater magnitude drop in performance across a switch relative to SST neurons, but these effects did not reach significance threshold (Figure 3.1.1d&g).

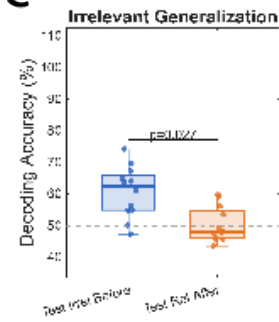
a

Trial	-32	...	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	...	32
Group	Train	...	Test 1	Train	Train	Test 1	Train	Train	-	Test 2	-	-	-	-	...	Test 2

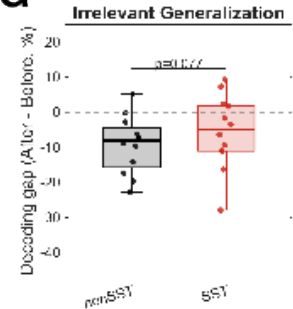
b



c



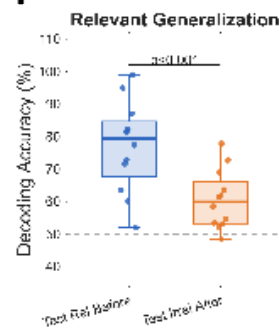
d



e



f



g

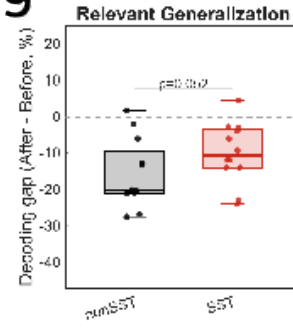
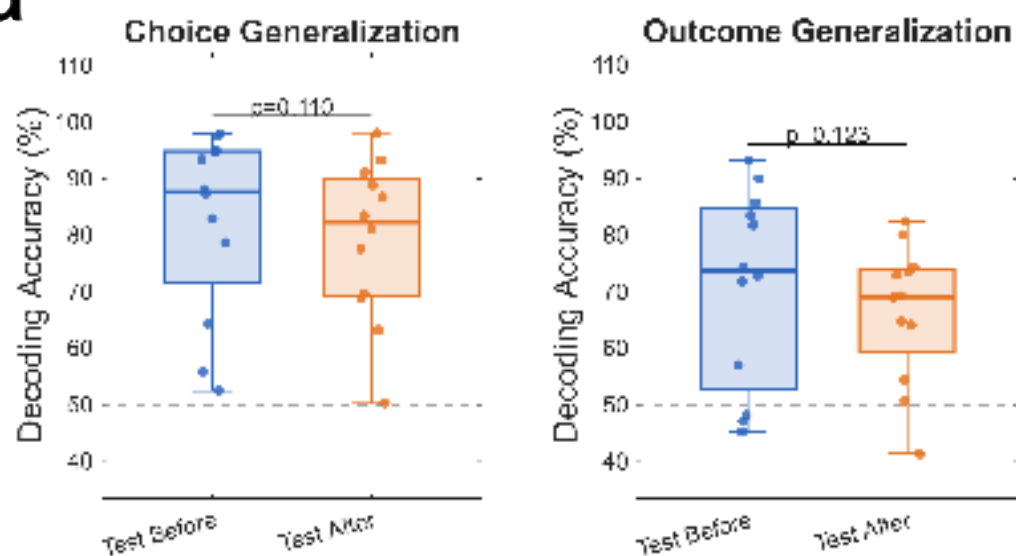


Figure 3.1.1: Generalization of decoding ability when trained on trials before a switch and tested before and after a switch

a) Schematic showing the training and testing paradigm of the generalization decoding. Decoders are trained on subsets of trials from the 32 trials before a switch, then tested on equal numbers of held out trials from the 32 trials before and after a switch. **b)** Sliding window example of relevant stimulus decoding on 5 trial window tiled across the 32 trials after a switch when trained on the 32 irrelevant trials before a switch. **c)** Decoding when testing on trial subsets from before versus after a switch using the same decoder trained on irrelevant trials before a switch. **d)** The difference in the gap between testing performance from trials before and after a switch when training on Non-SST versus SST neuronal groups. **e)** As in b, but trained on relevant trials before a switch and tested on irrelevant trials after a switch. **f)** As in c, but trained on relevant trials before a switch and tested on relevant trials before a switch as well as irrelevant trials after a switch. **g)** as in d, but using the gaps shown in f.

a



b

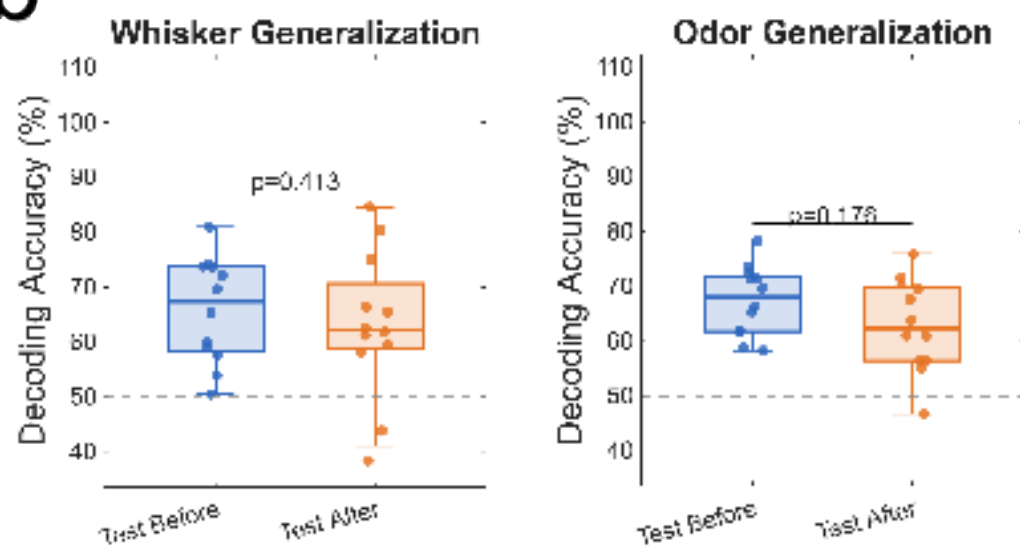


Figure 3.1.2: Generalization decoding from non-rule-related task variables

a) Left: Choice decoder performance when training on the 32 trials before a switch then testing on the held-out trials from before and after a switch. Right: As in left but with outcome decoding. **b)** As in a, but with odor and whisker stimulus decoding.

Chapter 3.2: Investigating informational changes over many trials

To assess how the brain updates its encoding of task-related information over the course of learning a new rule, we updated our SVM decoder train-test strategy. Here we trained SVM decoders on the trials before a switch, then tested on groups of 10 trials, tiling across the block of trials after the switch up until the next switch (Figure 3.2.1a). To test if decoding accuracy of the previously irrelevant, now relevant, stimulus had increased performance over learning, we trained the decoder on irrelevant trials then tested its ability to decode relevant trials after the switch. We found that there was no increase in decoder performance over the course of the block (Figure 3.2.1b). This lack of change was reflected in the decoding when using SST, as well as Non-SST neuron groups (Figure 3.2.1c).

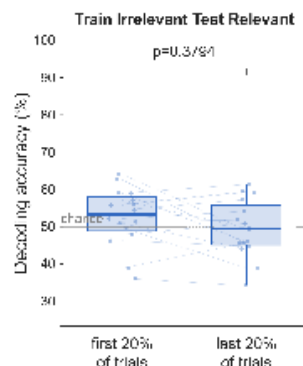
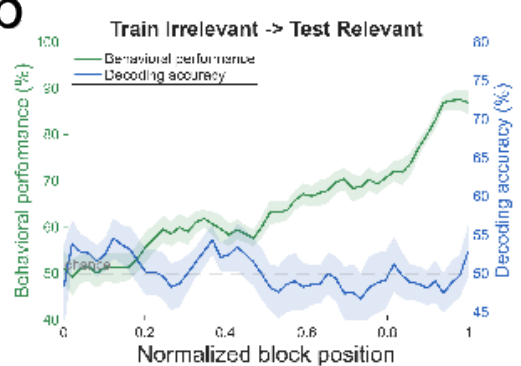
We then tested if newly relevant stimulus information occupied similar neural subspace as previously relevant stimulus. To do so, we trained the decoder on the previously relevant stimulus, then tested its ability to decode the new relevant stimulus. We observed that over the course of learning decoding performance of the new relevant stimulus improved to be higher at the end of the block relative to the beginning of the block, despite the current and previous relevant stimuli being of different stimulus modalities (All neurons, Relevant->Relevant: block accuracy vs normalized block position median Pearson $r = 0.327$, $p_{\text{raw}} = 0.0023$; mean accuracy first-vs-last block-tile 52.92 vs 60.98%; sign-rank across 17 blocks) (Figure 3.2.1d). There was no difference in magnitude of this change when decoding using SST versus Non-SST populations (Figure 3.2.1e).

We investigated other rule-related stimulus transitions as well. Training on the previously relevant stimulus and testing on the same modality, the now irrelevant stimulus, showed a decay in decoding performance that was roughly equal when decoding with individual neuronal populations (Relevant->Irrelevant, block accuracy vs normalized block position median Pearson r : All -0.356, $p_{\text{raw}} = 0.044$; nonSST -0.227, $p_{\text{raw}} = 0.022$; SST -0.264, $p_{\text{raw}} = 0.044$; sign-rank across 17 blocks) (Figure 3.2.2a). To discover whether or not irrelevant stimulus information is encoded in a shared neuronal subspace with previously irrelevant stimulus information of the opposite modality we trained and tested our decoder to classify trials as having a left or right indicating irrelevant stimulus presentation. Unlike the relevant stimulus, irrelevant stimulus information did not return to populate a shared neuronal subspace (Figure 3.2.2c). This was true for both SST and Non-SST populations (Figure 3.2.2b&d).

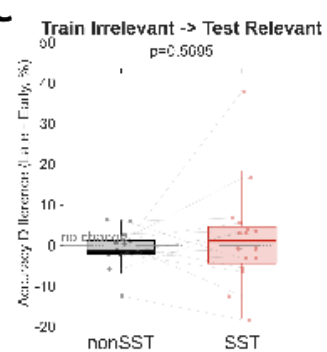
a

Trial	-32	...	-6	-5	-4	-3	-2	-1	1 - 10	2 - 11	3 - 12	4 - 13	5 - 14	...
Group	Train	Train	Train	Train	Train	Train	Train	Train	Test 1	Test 2	Test 3	Test 4	Test 5	...

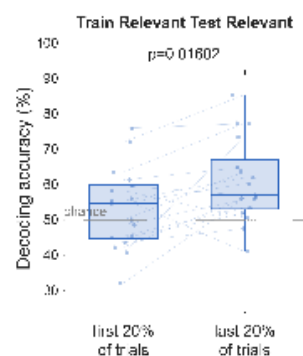
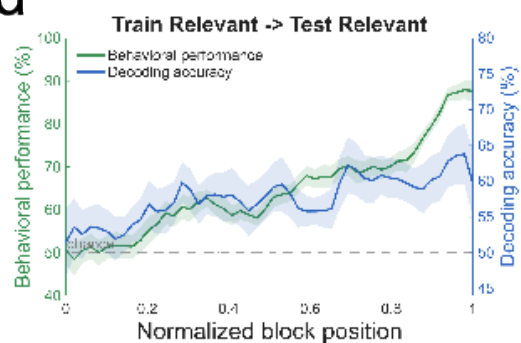
b



c



d



e

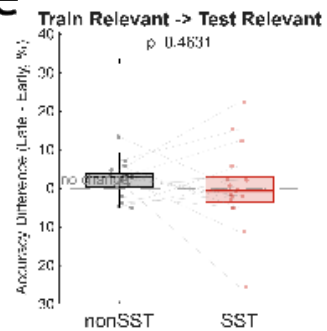


Figure 3.2.1: Decoding relevant information across block learning following decoder training

a) Schematic of training/testing paradigm. Decoder is trained on the 32 trials before a switch, then tested on groups of 10 trials following the switch tiling trial-by-trial across the block up until the next switch. **b)** Left: Results are interpolated onto a generalized x axis across sessions. Results shown are from training on a previously irrelevant stimulus, then tested on the same stimulus modality, the now relevant stimulus. Right: Comparison of decoding on the first 20% of trials versus the last 20% of trials from the post-switch block. **c)** Comparison of the gap between the first and last 20% trials decoding between Non-SST and SST populations. **d)** As in b, but training on relevant trials and testing on relevant trials. **e)** as in c, but training on relevant trials and testing on relevant trials.

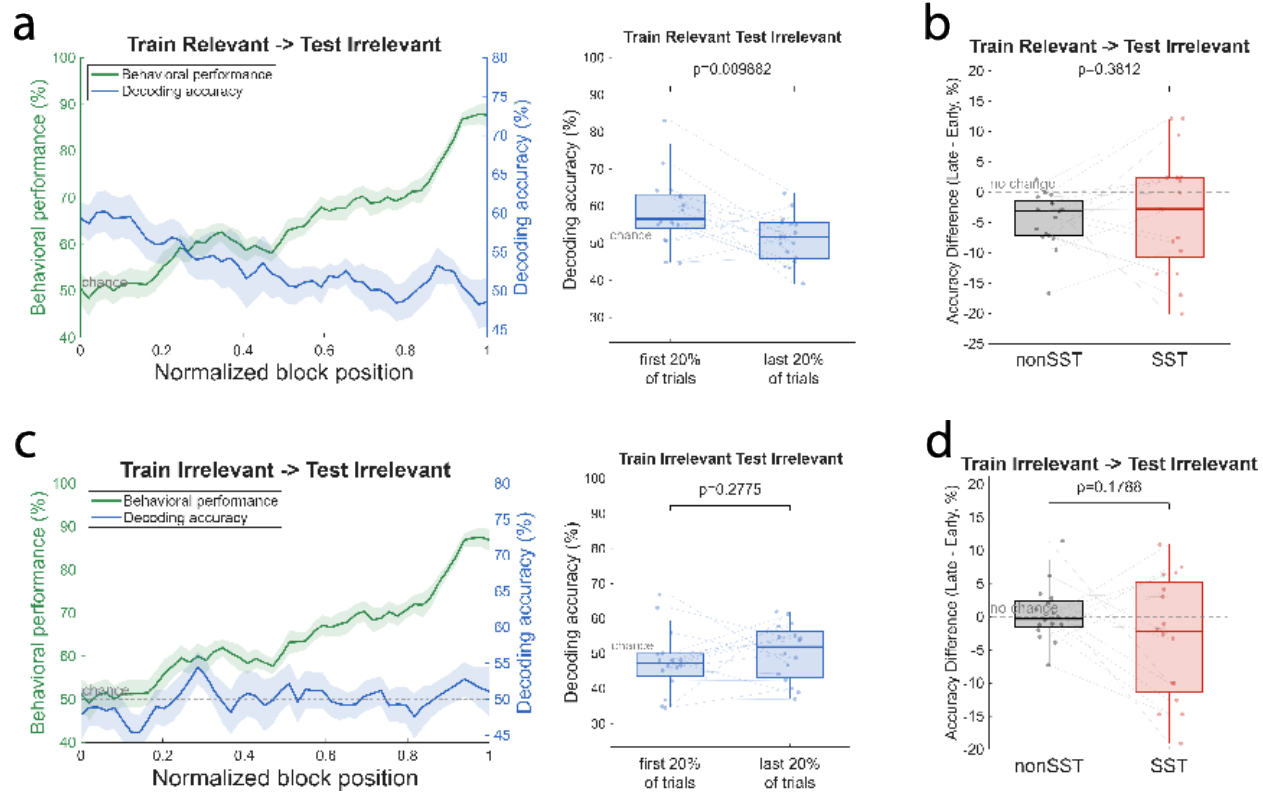


Figure 3.2.2: Decoding irrelevant information across block learning following decoder training

a) As in Figure 3.2.1b, but training on relevant trials and testing on irrelevant trials. **b)** As in Figure 3.2.1c, but training on relevant trials and testing on irrelevant trials. **c)** As in a, but training on irrelevant trials and testing on irrelevant trials. **d)** As in b, but training on irrelevant trials and testing on irrelevant trials.

DISCUSSION

Summary of findings

The experiments presented in this dissertation set out to characterize how the medial prefrontal cortex represents rule-related information during an attentional set-shifting task, whether that representation is reorganized when relevance changes at a rule switch, and whether somatostatin-expressing interneurons contribute to this process differently from the Non-SST population. Head-fixed mice performed multiple un-cued extradimensional shifts within single sessions, and two-photon calcium imaging of the mPFC, with SST neurons identified transgenically, allowed population activity to be decoded while retaining the identity of a genetically defined interneuron class. This work is organized around three main findings. First, the mPFC preferentially encodes the relevant stimulus over the irrelevant stimulus when the mouse is an expert at the current stimulus-reward contingency. Second, this preference is dynamic across a rule switch: relevant-stimulus decoding is stronger than irrelevant-stimulus decoding before a switch, when the animal is an expert on the rule. Then, in the trials immediately following, a switch, the relevant-stimulus decoding becomes weaker than when the animal was an expert on the previous rule. Third, when learning a new rule, information about the newly relevant stimulus reoccupies a population subspace previously used to encode the previously relevant stimulus, a stimulus of a different physical modality, while irrelevant-stimulus information does not undergo this same transformation.

Taken together, these observations support a view of the mPFC in which stimulus information is formatted according to its behavioral relevance rather than its sensory identity, and in which the code that carries relevant information is partially conserved

across rule changes even when the physical stimulus that occupies it changes. This discussion considers the contribution of SST interneurons, addresses the limitations of the design and the statistics, and proposes future experiments to test the geometry of stimulus representations in relevant and irrelevant contexts more directly than the decoders used in this work.

Within-session flexibility and the structure of behavioral state

Mice learned to perform multiple un-cued extradimensional shifts within a single session with no bias in performance for either stimulus modality, reaching the hidden performance criterion in comparable numbers of trials. This behavioral symmetry is important for the neural analyses, because it means that differences in decoding between rules are unlikely to reflect an asymmetry in how well the animal performed on one modality versus the other. The latent-state model showed that performance was not influenced by a fixed strategy but by a small number of discrete states that persisted for many trials at a time, comprising an engaged state in which choice depended on the current stimulus and two side-biased states in which choice was dominated by a lick bias and performance fell below chance. These states switched slowly, with high probability of persistence from trial to trial, consistent with prior descriptions of interleaved strategies in perceptual decision making⁶⁸ and reinforces the point that trials cannot be treated as independent samples of a single behavioral process.

The decision to retain disengaged trials in the decoding analyses, made in order to match trial counts before and after switches, has implications for the interpretation of this work. Because same-side choices and unrewarded outcomes cluster within biased states, decoders could exploit this to classify choice and outcome above chance earlier

in the trial than the underlying task events would explain. The results were interpreted with this confound in view, focusing on the rise in decoding that accompanies the emergence of task events. The influence of this bias on the central stimulus-relevance findings is minimal: the relevant and irrelevant stimuli are defined by the same physical modalities presented in the same trials, so a state-clustering artifact does not impact the core findings of this work.

Rule-dependent representation of stimulus information

The foundational result of this work is that the mPFC represents stimulus information preferentially when that stimulus is relevant to the current rule. At the level of single neurons, a larger proportion of cells had selective firing for the relevant stimulus than for the irrelevant stimulus, an effect that survived a nested analysis controlling for variance across mice and that was present in both the SST and non-SST populations. At the population level, a decoder blind to physical modality could read out the left-versus-right meaning of the relevant stimulus well above chance and did so more accurately than it could read out the irrelevant stimulus. The comparison between the modality-blind relevant decoder and the four-class stimulus-combination decoder further strengthens the evidence for this foundation: relevant-stimulus decoding exceeded the above-chance performance of the combination decoder, indicating that stimulus information in the mPFC is carried more in the form of rule-relevant meaning than in the form of raw stimulus identity. These findings are consistent with a long line of work distinguishing prefrontal cortex from primary sensory areas by its tendency to encode stimulus information according to task relevance rather than to represent all stimuli equally³¹.

The fact that the relevance effect appeared in both SST and non-SST populations, and that direct cell-type comparisons of stimulus decoding revealed no significant differences either before or after a switch, implies that rule-dependent formatting of stimulus information is not preferentially represented within these classes of neurons. This is a meaningful negative result given SST neurons distinctive role within the circuit: their dendrite-targeting position, their documented response plasticity during association learning, and their sensitivity to behavioral state⁵⁰. The data presented here indicate that, for the readout of stimulus relevance, whatever contribution SST neurons make is not distinct from that of the broader population under the statistical power available here.

Reorganization of the stimulus code across a rule switch

Relevant-stimulus decoding was stronger before a switch than in the trials immediately after, a pattern also seen for odor decoding, as well as in the generalization analyses, in which a decoder trained before a switch generalized poorly to the newly relevant stimulus of the opposite modality in the first trials of the new block. Irrelevant-stimulus decoding did not show the same strength in decoding before a switch, which is consistent with the idea that the pre-switch enhancement is specifically a signature of relevance.

The interpolation analysis extends these findings across the slower timescale of learning a new rule. When a decoder was trained on the previously relevant stimulus and tested on the newly relevant stimulus of a different modality, decoding improved from the beginning to the end of the post-switch block, even though the two stimuli were separate modalities. The corresponding transition for irrelevant information did not recover. A decoder trained on a previously irrelevant stimulus did not come to decode the new

irrelevant stimulus over the block. Together, these results suggest that the mPFC maintains a population subspace for relevant stimulus information into which the newly relevant stimulus is gradually mapped as the animal learns, while irrelevant information is not similarly organized into a conserved subspace. This has important implications, because it indicates that the subspace in which relevant information is represented is separable from the physical modality of the stimulus and is instead organized around rule-relevance.

There are two points to keep in mind, however, and both are based in what the decoders can and cannot support. First, cross-modality generalization shows that a decoder trained on one stimulus carries some ability to discriminate between stimuli from another modality, but it does not show that the representations share a common geometry. Furthermore, it does not quantify the extent to which a shared geometry exists. Second, the interpolation analysis normalizes blocks of differing length onto a common axis and mixes trials across engagement states; the improvement across the block should be read as a rough outline.

Contributions of somatostatin interneurons

The cell-type comparisons in this dissertation showed a largely consistent negative result for stimulus coding and a more suggestive, but unconfirmed, result for the encoding of outcome. Decoding of outcome after a switch was better from SST activity than from a neuron-balanced Non-SST subsample, and the population of SST neurons contained a higher proportion of error-selective than reward-selective cells; however, the latter effect did not survive a nested analysis controlling for the mouse, and the gap of the across-switch change in outcome decoding did not differ significantly between the two

populations. These results are compatible with the literature implicating SST neurons in feedback- and value-related signaling⁷⁸, but they do not establish a distinct SST contribution to outcome coding. They should be considered in the context of the low power driving the results of this work.

Several features of the design limit how strong any cell-type claim can be. SST neurons are far less numerous than non-SST neurons in every session, and the comparison was equalized by subsampling non-SST neurons to the SST count on each repeat. This equalizes neuron number but leaves the SST estimates resting on a small and fixed set of cells per session, so that a real but modest SST-specific effect could fail to be detected. The non-SST group pools excitatory neurons with non-SST interneuron classes such as PV and VIP, whose contributions cannot be separated in this data. A genuine division of labor among interneuron types, of the kind described in the introduction, could be present but invisible.

Within the critical context of the low power driving this study, one interesting trend that emerges in the data is that Non-SST decoding of rule-related information seems to regularly suffer greater deficits in decoding accuracy after a switch. Similarly, before a switch, SST neurons seem to have milder decoding accuracy for many event types. It may be worth investigating the extent to which SST neurons, or interneurons more broadly, are influenced by task events relative to excitatory neurons that tend to project information already sculpted within region to other brain areas. Worth noting however, is that analyses looking at the gaps between before and after differences in SST and Non-SST populations regularly revealed no significant differences. Still, with the low number

of sessions available within this dataset, this may be an area worthy of consideration in future studies.

Experimental and interpretive limitations

Beyond the state-clustering and cell-type power limitations already noted, several constraints should be kept in view when weighing the conclusions. The dataset comprises a small number of mice and sessions, and although the session was treated as the unit of replication and cross-session inference used paired nonparametric tests with FDR correction within related comparisons, the number of independent units is modest and the resulting tests have limited power to detect small effects. A non-significant cell-type difference should not be considered evidence of equivalence. The decoders throughout are linear, so all claims about represented information concern linearly decodable information; nonlinear structures would be missed. Calcium imaging reports activity relatively slowly, which limits the temporal precision with which fast spiking can be resolved. Furthermore, decoding measures the information that can be read from a population, not the information the animal actually uses. A task variable that decodes well may not actually be reflective of neural activity truly driving behavior, and demonstrating a causal role would require perturbation experiments. Finally, co-variance of event-types, such as choice and relevant stimulus during expert performance, provides challenges in some contexts for assigning relationships between decoder performance and singular task variables.

Future directions: the geometry of relevant and irrelevant stimulus representations

The most direct extension of this work is to move from decoding, which asks whether information is present, to geometry, which asks how that information is organized in the population space. The central finding, that relevant-stimulus information is preferentially encoded and that the newly relevant stimulus comes to reoccupy a conserved subspace across a switch, relies heavily on the assumption of shared geometry, and the cross-modality generalization decoders only test this indirectly. A geometric analysis would ask whether population activity occupies an axis for relevance into which stimuli of either modality are projected when they become relevant.

A first step is to define, within a shared low-dimensional subspace across all trials of a session, a vector for each stimulus in each rule context, defined as the difference between the population centroids of the two stimulus identities for all trials. Then the metric that can be used to assess similarity is the angle between coding vectors. Tracking these angles across the trials of a block would show whether the newly relevant stimulus ends up aligning with the relevance axis as the animal learns. This would provide evidence at the level of the population geometry to support the decoding seen in the interpolation analysis.

Beyond geometry, the natural follow ups to this work are causal analyses. Targeted optogenetic or chemogenetic manipulation of SST neurons during the trials surrounding a switch would test whether the interneuron class contributes to task performance, moving past the correlational analysis of imaging.

Conclusion

This dissertation shows that the mPFC encodes stimulus information according to its behavioral relevance, that this rule-dependent code is reshaped across un-cued rule switches, and that information about a newly relevant stimulus comes to occupy a population subspace independent of stimulus modality. These properties were expressed across the population rather than confined to SST or Non-SST neuron groups, whose distinct contribution, though suggested for outcome coding, was not established under the statistical power available. Overall, this work provides foundational evidence for mPFC's innate organization of cue information being rule-driven as opposed to being organized by physical modality.

METHODS

Animals

All of the experiments were conducted according to the National Institutes of Health (NIH) guidelines and with the approval of the Chancellor's Animal Research Committee of the University of California Los Angeles. Four adult transgenic mice (14-20 weeks old) were used for *in-vivo* mPFC calcium imaging experiments. Mice were a cross between SST-Cre (Jackson Laboratory Strain #:013044) and Ai14 (Jackson Laboratory Strain #:007914).

Surgical Procedure

Mice were administered pre-operative drugs (carprofen 5mg/kg, dexamethasone 0.2mg/kg, lidocaine 5mg/kg) 30 minutes before surgery. Mice were then anaesthetized with isoflurane (5% induction, 1-2% for maintenance) and fitted into a stereotactic frame with blunt earbars (Kopf Instruments). Body temperature was maintained on a feedback-controlled heating pad (Harvard Apparatus) set to 37 degrees celcius. Eyes were protected from desiccation using artificial tear ointment. The scalp was cleaned three times with 10% povidone-iodine and 70% ethanol. Fascia was removed by applying hydrogen peroxide. After removing the scalp and clearing the skull of connective tissues, the skull was scored to help secure adhesives at the end of surgery a hole was drilled centered at 1.8mm anterior, and 0.5mm lateral of bregma above the left hemisphere. This was expanded 0.5mm in all directions to make room for the GRIN lens. Using a nanoject injector, 500nL of virus (AAV9-hSyn-GCaMP8f-WPRE) was injected 1.8mm anterior, 0.4mm lateral, and 2.0mm ventral to bregma. After waiting 30 minutes, the capillary from the nanoject was removed from the brain. Using a blunt needle tip (28g) on a syringe attached to a vacuum brain tissue was aspirated out of the cortex down 1.8mm ventral from bregma to make room for the GRIN lens. Cortex buffer (NaCl = 7.88g/L, KCl = 0.372g/L, HEPES = 1.192g/L, CaCl₂ = 0.264g/L, MgCl₂ = 0.204g/L, at a pH of 7.4) was continuously flushed during aspiration until bleeding stopped. A 1mm diameter GRIN lens with proview cap was lowered 1.8mm ventral from bregma, centered at 1.8mm anterior and 0.5mm lateral from bregma. This was secured in place using vetbond adhesive. Then, custom made lightweight stainless steel metal headbars were surgically attached to the back of the skull using cyanoacrylate adhesive. Mice were allowed one week to recover before water deprivation and habituation training.

Experimental Setup

The behavioral setup is as described in Spellman et al., 2021¹⁹. Mice were head fixed above an open top aluminum tube with two lick spouts secured several millimeters below the mouth on either side of the skull. A continuous stream of air, passing through a vial of mineral oil control, was delivered under the animal's nose. Airflow was diverted to alternative odorant-containing vials depending on odor stimulus to be presented. This was achieved using a series of 3-way valves controlled by an arduino Mega 2560 microcontroller with a custom shield⁷⁹. Odorants were purchased from Sigma Aldrich: Cinnamon leaf oil (W229210), Citral (C83007), Pinene (147524), and Limonene (W504505). Odorants were diluted 1:10 in mineral oil. Water release was controlled by two, two-way valves which were also controlled by the arduino microcontroller. Water was held in a 10mL syringe with the back removed roughly a foot above the apparatus. Tubing attached to the end of the syringe allowed water to flow as directed by gravity upon the opening of one of the valves. Whisker stimulus was delivered via two covered speakers with pipette tip funnels placed onto the whiskers. Whisker stimulus was driven by the arduino microcontroller, with signal sent to a Digital Analog Converter before being sent to a preamp connected to the speakers. Behavioral data was collected and returned to the computer by recording binary values at 1kHz from the arduino and recording them in Matlab. The behavior rig was operated through custom Matlab and arduino code from Spellman et al., 2021¹⁹.

Behavioral Training

Habituation:

Habituation 1: Male and Female SST-Cre::Ai14 mice on a hybrid background were habituated to the behavioral rig without head fixation where they were guided towards reward spouts by the experimenter. Reward spouts released a 3uL water reward at alternating spouts every 5 seconds. Mice were removed upon displaying signs of agitation as repetitive grooming. Pass criteria were consuming 500uL of water

Habituation 2: Mice were habituated to head fixation over a period of 10 minutes in which they could lick for water rewards at alternating spouts for 2-4 licks per side. The mouse was removed after 10 minutes or after obtaining 500uL water

Habituation 3: Mice were habituated to alternating sides during head fixation by allowing mice to lick freely during which a single spout was armed for 20 licks after which the armed spout switched to the opposite spout. Mice were removed after 10min or after reaching performance criteria or 500uL consumed

Shaping:

Shaping 1: Mice were introduced to the trial structure on a single stimulus modality, either whisker or odor. A trial start cue of static noise was presented for 500ms, after which one of two stimuli were presented for 2.5 seconds. Reward was dispensed at the correct spout in response to the first lick in the 1.5 second response period after the

stimulus presentation period, independent of the spout chosen. Individual spouts were armed in blocks of 20 trials, during which an associated stimulus cue was presented during the stimulus period, with the cue alternating along with the associated spout side every block. Mice were removed after 10 consecutive trials without licking during the response window. Pass criteria was 500uL consumed.

Shaping 2: Mice were presented with a similar structure as Shaping 1, except that water reward was only presented if the mouse licked the correct spout. Pass criteria was >95% correct in a 100 trial window.

Shaping 3: Mice were presented with a similar structure as Shaping 2, except that correct side blocks were randomized to be between 1-4 trials before alternating. Pass criteria was >95% correct in a 100 trial window.

Training Stages:

Simple Discrimination (SD): Mice were presented with a randomized and equal likelihood of left and right trials. Lick bias to a particular side was countered with an increase in presentation of opposite side trials. Pass criteria was >95% correct in a 30 trial window.

Compound Discrimination (CD): The opposite stimulus modality was presented in parallel as a distractor during the stimulus presentation period. Pass criteria was >95% correct in a 30 trial window.

Intradimensional Shift 1 (IDS1): The reward-indicating stimuli were replaced by two new stimuli of the same modality. Pass criteria was >95% correct in a 30 trial window.

Reversal (Rev): Reward indicating stimuli were reversed to indicate reward at the opposite spout. Pass criteria was >95% correct in a 30 trial window.

Extradimensional Shift (EDS): Reward indicating stimulus modality was switched with distractor stimulus modality. Pass criteria was >95% correct in a 30 trial window.

Intradimensional Shift 2 (IDS2): As in IDS1, the new reward indicating stimuli were replaced with two new stimuli of the same modality. Pass criteria was >95% correct in a 30 trial window.

Serial Extradimensional Shift (SEDS): In the testing phase, mice were presented with stimuli as in IDS2. Upon reaching 80% performance on the last 40 trials, 50% performance on the last 10 right and left trials, while maintaining an overall response rate of 80% on the last 10 trials, an uncued extradimensional shift occurred. Only sessions with 2 or more shifts were included in final analyses.

In-vivo two-photon imaging

All imaging was conducted using a resonant scanning two-photon microscope (Scientifica) fitted with a 10x 0.5 NNA objective (Thorlabs Olympus). Recordings were 512x512 pixel frames at 30.9Hz. Recordings were taken at 1.4x zoom using excitation light delivered with a Ti:sapphire excitation laser (Chameleon Ultra II, Coherent),

operated at 920nm. Fluorescence was recorded with two photomultiplier tube s(GaAsP PMT; Hamamatsu) with one per channel. Light was split through dichroic mirrors to isolate out emission spectra thought to originate from tdTomato fluorophores and GCaMP8f fluorescence. Microscope control was performed using the LabView-based software SciScan. Behavioral data was synchronized to imaging data through the recording of TTL pulses generated at the start of each imaging frame. This was aligned to the 1kHz outputs from the arduino microcontroller driving the behavioral apparatus. Imaging sessions were performed 6 days a week, taking roughly 1-2 hours per session. Only imaging sessions that had behavioral performance of 2 or more switches were included with others being excluded. Fields of view were aligned across days by eye. Laser power was adjusted in an animal-by-animal fashion (50-80mW) to maximize the signal retrieved while limiting bleaching over long recording sessions. PMT sensitivity was kept constant across mice/sessions.

Histology

Mice were deeply anaesthetized with isoflurane and transcardially perfused with 10-20mL of 1x PBS. This was followed by 10mL of 4% paraformaldehyde in 1x PBS. After perfusion, brains were removed and post-fixed for 20 hours in 4% paraformaldehyde solution. Using a vibratome (Leica VT 1200S), brains were coronally sectioned at 40um 24 hours after perfusion. Sections were mounted onto microscope slides, provided a DAPI-containing mounting medium, and covered with a glass cover slip. Images were acquired on a tiling microscope to confirm proper lens placement.

Quantification and Statistical Analysis

Spectral unmixing and tiling

Raw interleaved two-photon frames (512 × 512 pixels, 16-bit, alternating green GCaMP and red reference channels) were separated by channel. To correct spectral bleed-through between channels, a 2 × 2 linear unmixing matrix with fixed off-diagonal coefficients was applied to each frame pair. Unmixed frames were rewritten as block-wise interleaved TIFF stacks for cell extraction, and global mean images of both the raw and unmixed data were retained for quality control.

ROI detection

Regions of interest and neuropil signals were obtained with Suite2p⁸⁰. Neuropil-corrected z-scored fluorescence was used as the primary activity measure. ROIs

classified as cells by the Suite2p classifier were sorted manually to separate neurons from noise.

SST interneuron identification

Somatostatin-expressing (SST) interneurons were identified manually. For each session, ROIs whose location coincided with red-labelled somata in the mean red-channel image had their ROI numbers recorded and for isolation in downstream analysis. A neuron was classified as SST if it was identified as red-labelled in at least one session; all remaining cells were classified as non-SST.

Cross-session cell registration

Neurons were tracked across sessions within each animal using CellReg⁸¹, which produced a cell-to-index map assigning each across-session cell identity its per-session ROI index (or a null entry where the cell was not detected in a given session). All per-session activity tensors were reformatted into this common across-session cell space so that a given row referred to the same tracked neuron in every session in which it appeared. No analyses used neural identity across sessions, and statistical inference across sessions used the session as the unit of replication.

Trial-tensor assembly

For each session the activity was organized into a neurons $\times$ timepoints $\times$ trials tensor, with the time axis locked to trial onset and truncated to a common trial length (the minimum across sessions) so that tensors were directly comparable. Rule-switch trial indices were identified from the behavioral record, and each session was partitioned into pre-switch and post-switch epochs. The post-switch epoch was defined as the 32 trials immediately following each switch, matching the pre-switch window in duration. These per-session tensors, together with the behavioral tables and event-frame indices, were assembled into a single master data structure per animal for downstream analysis.

Behavioral engagement state labelling

To characterize task engagement, each responded trial was assigned a latent behavioral state using a Bernoulli generalized-linear-model hidden Markov model (GLM-HMM) derived from Ashwood et al., 2022⁶⁸. A three-state model was pre-fit on a larger pooled set of sessions and then applied to each imaging session. The per-trial GLM inputs were the current relevant-stimulus side, the previous choice, and the previous relevant-stimulus side, as well as a lick bias constant. The three inferred states corresponded to an engaged state and two response-biased (left- and right-biased)

states; the two biased states were merged into a single disengaged category for downstream analysis. States were inferred only on responded trials and were forward- and backward-filled to intervening no-response trials. Posterior state probabilities were retained for each trial. Engagement-split analyses used these labels to partition trials into engaged and disengaged subsets. For each GLM-HMM state, we computed choice entropy as the Shannon entropy (in bits) of the animal's binary choice distribution,

$$H = -p \cdot \log_2 p - (1-p) \cdot \log_2 (1-p),$$

where p is the posterior-weighted proportion of rightward choices assigned to that state. Entropy ranges from 0 (fully deterministic responding) to 1 bit (choices at chance, $p = 0.5$), providing a state-level index of behavioral stochasticity.

General-linear-model (GLM) for single neuron event specificity

In parallel, one ordinary-least-squares GLM was fit per neuron with the response variable defined as mean activity in the 1 s post-outcome window (anchored to reward delivery on rewarded trials and to the last lick on unrewarded trials). The design matrix contained an intercept and predictors for context, choice, outcome, relevant stimulus, and irrelevant stimulus. All predictors were fit simultaneously, so each coefficient reflects that variable's contribution with the others held constant. Significance of a given predictor was assessed by permuting only that predictor's column (1,000 permutations) while holding the remainder of the design matrix fixed, refitting, and forming a two-tailed permutation p-value from the resulting null coefficient distribution; permutation again preserved label proportions. Neurons were subject to a minimum of 10 valid trials, and rank-deficient design matrices were excluded. Standardized (activity-z-scored) coefficients and per-neuron goodness-of-fit (R^2 and adjusted R^2) were additionally computed for effect-size reporting and model-quality screening.

Population decoding

Population-level information about task variables was assessed with linear support-vector-machine (SVM) decoders applied to the trial tensors. Decoders were run for six targets — whisker side, odor side, relevant-stimulus side, irrelevant-stimulus side, choice side, and outcome — within three trial groups (all trials, pre-switch, post-switch) and three neuron groups (all neurons, non-SST, SST). The physical-modality targets (whisker, odor) and the rule-relevance targets (relevant, irrelevant) were decoded separately, since the same physical stimulus contributes to a relevant target in one

block and an irrelevant target in the next; contrasting these targets isolates rule-dependent reorganization from stimulus-driven responses.

The base decoder was a linear SVM with standardized features and unit box constraint. On every train/test repeat the classes were balanced by subsampling to the minority-class count, enforcing a 50% chance level. Runs with fewer than eight trials per class were skipped

Because the feature space (hundreds of neurons $\times$ time) greatly exceeds the number of training trials, decoders were also fit after dimensionality reduction by principal component analysis (PCA). To prevent data leakage, the reduction was fit on the training trials only within each repeat and then applied to the held-out test trials. The number of retained components, R , was chosen from a diagnostic sweep over candidate values on a representative session: R was set to the smallest value achieving at least 95% of the peak decoding accuracy.

For pre- versus post-switch comparisons, the number of balanced trials was capped at the minimum of the two epochs so that the two epochs were decoded with matched statistical power within each session.

A within-session chance floor was established by refitting each decoder on label-permuted data, reusing the identical train/test split and dimensionality-reduction projection so that the null differed from the real decoder only in the label assignment. Decoder performance above chance was summarized per session as the gap between the real and shuffled accuracy over a fixed post-stimulus window, and this gap was tested against zero across sessions.

Cell-type comparison decoder

A direct comparison of decoding by SST and non-SST populations is confounded by group size, since non-SST neurons vastly outnumber SST neurons in every session. To control for this, on every repeat all SST neurons were used while an equal-sized random subsample was drawn from the non-SST pool, equalizing neuron count per group. Consequently, the SST variance reflects trial subsampling alone, whereas the non-SST variance additionally reflects neuron subsampling, and the two are matched in the number of neurons contributing to each decode.

For paired statistical testing, each time-resolved decoding curve was collapsed to a single scalar, the area under the decoding curve (Mean Window Accuracy), defined as the mean Accuracy over a fixed post-stimulus window of 3.05–4.05 s relative to trial onset. The same window was applied to every decoder in the pipeline so that scalars were directly comparable, and pre- versus post-switch mean window accuracy values were compared with paired Wilcoxon signed-rank tests across sessions.

Multi-class stimulus-combination decoder

To assess joint coding of the two stimulus dimensions, a four-class decoder classified the combined odor-side × whisker-side identity of each trial using one-versus-rest linear SVM decoding. All four classes were subsampled to the minority-class count on every repeat, giving a 25% chance level, and PCA was again fit on training trials only. Multi-class performance was summarized by the macro-averaged one-versus-rest accuracy across the four classes.

To compare decoding of the four-way stimulus combination against the binary relevant- and irrelevant-stimulus decoders, all decoders were placed on a common scale by expressing accuracy as the proportional distance from chance to ceiling: $\text{prop} = (\text{accuracy} - \text{chance}) / (100 - \text{chance}) \times 100$, using chance = 25% for the 4-class decoder and 50% for each binary decoder. This maps chance to 0 and perfect classification to 100 for every decoder. For each session, mean window accuracy was determined for a one second window during the trial (3.05–4.05 s). Decoders were compared pairwise across sessions (StimCombo vs. Relevant) using paired Wilcoxon signed-rank tests, with the session as the unit of replication.

Cross-epoch and cross-modality generalization

The stability of the population code across the rule switch was probed with generalization decoders. A decoder was trained on pre-switch trials and tested both on held-out pre-switch trials and on post-switch trials projected into the same (training-derived) latent space; the drop from within-epoch to cross-epoch performance indexes how much the coding geometry reorganized across the switch. In every case the dimensionality reduction was fit on the training set alone and applied unchanged to the test set, so all values were evaluated in a common latent space, and balanced-trial caps matched those used in the standard decoders.

A cross-modality variant was trained and tested across different task variables to ask whether the geometry that encodes stimulus in a rule-related way changes around a switch. Each train/test pairing carried its own label-permuted null.

Block-normalized interpolation decoder

To relate the time course of neural decoding to the animal's behavioral adaptation within a block, a decoder was trained on the 32 response trials immediately preceding a block (the tail of the previous block) and applied to a centered, edge-shrinking sliding window advanced across the trials of the current block. Training and testing trial sets were disjoint by construction, so there was no leakage even at the final window. At each

window position both a decoding scalar (mean accuracy over the activity window) and behavioral performance (fraction correct) were computed. Each window's native trial position was mapped onto a normalized [0, 1] block axis and both curves were interpolated onto a common grid, allowing blocks of differing length to be averaged and the neural decoding trajectory to be correlated with the behavioral learning trajectory across the block. For non-SST groups, neuron subsets were redrawn on each repeat to variance-match the smaller SST population rather than merely count-matching it, and a label-shuffle null for the normalized early-versus-late contrast was generated by permuting the test labels once per permutation.

Statistical analysis and reproducibility

The session was the unit of replication for all cross-session inference. Paired comparisons used the two-sided Wilcoxon signed-rank test across sessions, and unpaired comparisons used the Wilcoxon rank-sum test. Where a family of related tests was performed — for example, per-time-bin comparisons along a decoding curve, or per-neuron tests within an event — the Benjamini–Hochberg procedure was applied to control the false-discovery rate within that family.

All resampling, permutation, and subsampling procedures used explicit, reproducible random-number streams. Because a single global seed does not propagate to parallel workers, each parallelized permutation or repeat drew from an independent counter-based (Threefry) substream keyed to a hash of the animal, session, epoch, and analysis type, so that results were reproducible across re-runs and independent of parallel execution order. Permutation nulls were built with 1,000 iterations for single-neuron selectivity and with analysis-specific iteration counts for the decoders and geometry, as noted above. Dimensionality reduction was always fit on training data only. Decoder configurations, retained-component counts, and balanced-trial counts were logged per session and verified by dedicated reporting routines.

Software

The behavior rigs were operated using Arduino code implemented via the Arduino IDE. The analysis pipeline was implemented in MATLAB (including the Parallel Computing Toolbox for the permutation and repeat loops), with the behavioral-state model fit and applied in Python using the SSM package from Ashwood et al., 2022⁶⁸. Cell extraction used Suite2p⁸⁰ and across-session registration used CellReg⁸¹. Per-neuron and per-session summary statistics were exported for figure generation and confirmatory statistical visualization. Single neuron event specificity analysis statistics and plotting was performed with Prism Graphpad 10.

REFERENCES

1. Wang JX, Kurth-Nelson Z, Kumaran D, et al. Prefrontal cortex as a meta-reinforcement learning system. *Nat Neurosci*. 2018;21(6):860-868. doi:10.1038/s41593-018-0147-8
2. Cho KKA, Davidson TJ, Bouvier G, Marshall JD, Schnitzer MJ, Sohal VS. Cross-hemispheric gamma synchrony between prefrontal parvalbumin interneurons supports behavioral adaptation during rule shift learning. *Nat Neurosci*. 2020;23(7):892-902. doi:10.1038/s41593-020-0647-1
3. Mante V, Sussillo D, Shenoy KV, Newsome WT. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*. 2013;503(7474):78-84. doi:10.1038/nature12742
4. Minxha J, Adolphs R, Fusi S, Mamelak AN, Rutishauser U. Flexible recruitment of memory-based choice representations by the human medial frontal cortex. *Science*. 2020;368(6498). doi:10.1126/science.aba3313
5. Rikhye RV, Gilra A, Halassa MM. Thalamic regulation of switching between cortical representations enables cognitive flexibility. *Nat Neurosci*. 2018;21(12):1753-1763. doi:10.1038/s41593-018-0269-z
6. Tait D, Chase E, Brown V. Attentional Set-Shifting in Rodents: A Review of Behavioural Methods and Pharmacological Results. *Curr Pharm Des*. 2014;20(31):5046-5059. doi:10.2174/1381612819666131216115802
7. Reinert S, Hübener M, Bonhoeffer T, Goltstein PM. Mouse prefrontal cortex represents learned rules for categorization. *Nature*. 2020;(November 2019). doi:10.1038/s41586-021-03452-z
8. Maggi S, Peyrache A, Humphries MD. An ensemble code in medial prefrontal cortex links prior events to outcomes during learning. *Nat Commun*. 2018;9(1). doi:10.1038/s41467-018-04638-2
9. Lam NH, Mukherjee A, Wimmer RD, Nassar MR, Chen ZS, Halassa MM. *Prefrontal Transthalamic Uncertainty Processing Drives Flexible Switching*. Vol 637. Springer US; 2024. doi:10.1038/s41586-024-08180-8
10. Colacicco G, Welzl H, Lipp HP, Würbel H. Attentional set-shifting in mice: Modification of a rat paradigm, and evidence for strain-dependent variation. *Behav Brain Res*. 2002;132(1):95-102. doi:10.1016/S0166-4328(01)00391-6
11. Lebedeva A, Wang Y, Funnell L, et al. Dorsal prefrontal cortex drives perseverative behavior in mice. *Nat Commun*. 2026;17(1):1-22. doi:10.1038/s41467-026-71664-w

12. Cho KKA, Hoch R, Lee AT, Patel T, Rubenstein JLR, Sohal VS. Gamma rhythms link prefrontal interneuron dysfunction with cognitive inflexibility in *dlx5/6+/-* mice. *Neuron*. 2015;85(6):1332-1343. doi:10.1016/j.neuron.2015.02.019
13. Weinberger DR, Lipska BK. Cortical maldevelopment, anti-psychotic drugs, and schizophrenia: a search for common ground. *Schizophr Res*. 1995;16(2):87-110. doi:10.1016/0920-9964(95)00013-C
14. Lesh TA, Niendam TA, Minzenberg MJ, Carter CS. Cognitive control deficits in schizophrenia: Mechanisms and meaning. *Neuropsychopharmacology*. 2011;36(1):316-338. doi:10.1038/npp.2010.156
15. Miller JA, Tambini A, Kiyonaga A, D'Esposito M. Long-term learning transforms prefrontal cortex representations during working memory. *Neuron*. 2022;110(22):3805-3819.e6. doi:10.1016/j.neuron.2022.09.019
16. Lui JH, Nguyen ND, Grutzner SM, et al. Differential encoding in prefrontal cortex projection neuron classes across cognitive tasks. *Cell*. 2021;184(2):489-506.e26. doi:10.1016/j.cell.2020.11.046
17. Klein-Flügge MC, Bongioanni A, Rushworth MFS. Medial and orbital frontal cortex in decision-making and flexible behavior. *Neuron*. 2022;110(17):2743-2770. doi:10.1016/j.neuron.2022.05.022
18. Wójcik MJ, Stroud JP, Wasmuht D, et al. *Learning Shapes Neural Geometry in the Primate Prefrontal Cortex*. Vol 29. Springer US; 2026. doi:10.1038/s41593-026-02333-w
19. Spellman T, Svei M, Kaminsky J, Manzano-Nieves G, Liston C. Prefrontal deep projection neurons enable cognitive flexibility via persistent feedback monitoring. *Cell*. 2021;184(10):2750-2766.e17. doi:10.1016/j.cell.2021.03.047
20. Le Merre P, Heining K, Slashcheva M, et al. A prefrontal cortex map based on single-neuron activity. *Nat Neurosci*. 2026;29(March). doi:10.1038/s41593-025-02190-z
21. Morgan AA, Alves ND, Stevens GS, et al. Medial Prefrontal Cortex Serotonin Input Regulates Cognitive Flexibility in Mice. *BioRxiv Prepr Serv Biol*. 2023;6216(646):1-34. doi:10.1101/2023.03.30.534775
22. Hjort MM, Garrett ZQ, Gordon AG, et al. Prefrontal to ventral tegmental area dynamics drive contingency degradation. *Nature*. 2026;655(December 2024). doi:10.1038/s41586-026-10443-5
23. Nigro M, Silva Tortorelli L, Garad M, Zlebnik N, Yang H. Locus coeruleus modulation of prefrontal dynamics during attentional switching in mice. *eLife*. 2026;14:1-18. doi:10.7554/elife.105911.4

24. Klune CB, Jin B, Denardo LA. Linking mpfc circuit maturation to the developmental regulation of emotional memory and cognitive flexibility. *eLife*. 2021;10:1-33. doi:10.7554/eLife.64567
25. Musall S, Sun XR, Mohan H, et al. Pyramidal cell types drive functionally distinct cortical activity patterns during decision-making. *Nat Neurosci*. 2023;26(3):495-505. doi:10.1038/s41593-022-01245-9
26. Finkelstein A, Fontolan L, Economo MN, Li N, Romani S, Svoboda K. Attractor dynamics gate cortical information flow during decision-making. *Nat Neurosci*. 2021;24(6):843-850. doi:10.1038/s41593-021-00840-6
27. Hajnal MA, Tran D, Einstein M, et al. Continuous multiplexed population representations of task context in the mouse primary visual cortex. *Nat Commun*. 2023;14(1). doi:10.1038/s41467-023-42441-w
28. Langdon C, Engel TA. Latent circuit inference from heterogeneous neural responses during cognitive tasks. *bioRxiv*. Published online 2022:2022.01.23.477431-2022.01.23.477431. doi:10.1038/s41593-025-01869-7
29. Tye KM, Miller EK, Taschbach FH, Benna MK, Rigotti M, Fusi S. Mixed selectivity: Cellular computations for complexity. *Neuron*. 2024;112(14):2289-2303. doi:10.1016/j.neuron.2024.04.017
30. Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J. Dynamic coding for cognitive control in prefrontal cortex. *Neuron*. 2013;78(2):364-375. doi:10.1016/j.neuron.2013.01.039
31. Chelaru MI, Eagleman S, Andrei AR, Milton R, Kharas N, Dragoi V. High-Order Correlations Explain the Collective Behavior of Cortical Populations in Executive, But Not Sensory Areas. *Neuron*. Published online 2021:1-8. doi:10.2139/ssrn.3803611
32. Aoi MC, Mante V, Pillow JW. Prefrontal cortex exhibits multidimensional dynamic encoding during decision-making. *Nat Neurosci*. 2020;23(11):1410-1420. doi:10.1038/s41593-020-0696-5
33. Rigotti M, Barak O, Warden MR, et al. The importance of mixed selectivity in complex cognitive tasks. *Nature*. 2013;497(7451):585-590. doi:10.1038/nature12160
34. Williams AH, Kim TH, Wang F, et al. Unsupervised Discovery of Demixed, Low-Dimensional Neural Dynamics across Multiple Timescales through Tensor Component Analysis. *Neuron*. 2018;98(6):1099-1115.e8. doi:10.1016/j.neuron.2018.05.015

35. Bari BA, Grossman CD, Lubin EE, Rajagopalan AE, Cressy JI, Cohen JY. Stable Representations of Decision Variables for Flexible Behavior. *Neuron*. 2019;103(5):922-933.e7. doi:10.1016/j.neuron.2019.06.001
36. Panzeri S, Moroni M, Safaai H, Harvey CD. The structures and functions of correlations in neural population codes. *Nat Rev Neurosci*. 2022;23(9):551-567. doi:10.1038/s41583-022-00606-4
37. Koay SA, Charles AS, Thiberge SY, Brody CD, Tank DW. Sequential and efficient neural-population coding of complex task information. *Neuron*. 2022;110(2):328-349.e11. doi:10.1016/j.neuron.2021.10.020
38. Cunningham JP, Yu BM. Dimensionality reduction for large-scale neural recordings. *Nat Neurosci*. 2014;17(11):1500-1509. doi:10.1038/nn.3776
39. Barbosa J, Proville R, Rodgers CC, DeWeese MR, Ostojic S, Boubenec Y. Early selection of task-relevant features through population gating. *Nat Commun*. 2023;14(1). doi:10.1038/s41467-023-42519-5
40. Abs E, Poorthuis RB, Apelblat D, et al. Learning-Related Plasticity in Dendrite-Targeting Layer 1 Interneurons. *Neuron*. 2018;100(3):684-699.e6. doi:10.1016/j.neuron.2018.09.001
41. Fishell G, Kepecs A. Interneuron Types as Attractors and Controllers. *Annu Rev Neurosci*. 2020;43:1-30. doi:10.1146/annurev-neuro-070918-050421
42. Kepecs A, Fishell G. Interneuron cell types are fit to function. *Nature*. 2014;505(7483):318-326. doi:10.1038/nature12983
43. Tremblay R, Lee S, Rudy B. GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits. *Neuron*. 2016;91(2):260-292. doi:10.1016/j.neuron.2016.06.033
44. Pinto L, Dan Y. Cell-Type-Specific Activity in Prefrontal Cortex during Goal-Directed Behavior. *Neuron*. 2015;87(2):437-450. doi:10.1016/j.neuron.2015.06.021
45. Jeong H, Kim D, Song M, Paik SB, Jung MW. Distinct roles of parvalbumin- and somatostatin-expressing neurons in flexible representation of task variables in the prefrontal cortex. *Prog Neurobiol*. 2020;187(January):101773-101773. doi:10.1016/j.pneurobio.2020.101773
46. Lagzi F, Bustos MC, Oswald A marie, Doiron B. Assembly formation is stabilized by Parvalbumin neurons and accelerated by Somatostatin neurons. Published online 2021:1-34.
47. Perez SM, Boley A, Lodge DJ. Region specific knockdown of Parvalbumin or Somatostatin produces neuronal and behavioral deficits consistent with those

- observed in schizophrenia. *Transl Psychiatry*. 2019;9(1). doi:10.1038/s41398-019-0603-6
48. Piet A, Ponvert N, Ollerenshaw D, et al. Behavioral strategy shapes activation of the Vip-Sst disinhibitory circuit in visual cortex. *Neuron*. Published online 2024:1-15. doi:10.1016/j.neuron.2024.02.008
 49. Boivin JR, Schmerl B, Martin KB, Lee CF, Nedivi E. Cortical Somatostatin Innervation Follows a Unique Experience-Independent Developmental Trajectory. *J Neurosci Off J Soc Neurosci*. 2026;46(6):1-13. doi:10.1523/JNEUROSCI.1870-25.2026
 50. Khoury CF, Fala NG, Runyan CA. Specialization in somatostatin subnetworks in sensory and association cortex. *bioRxiv*. 2022;(16):2022.01.18.476770-2022.01.18.476770.
 51. Mosso MB, Zhu M, Ma X, Park E, Barth AL. Long-lasting, subtype-specific regulation of somatostatin interneurons during sensory learning. *Sci Adv*. 2025;11(33):1-15. doi:10.1126/sciadv.adt8956
 52. Park E, Mosso MB, Barth AL. Neocortical somatostatin neuron diversity in cognition and learning. *Trends Neurosci*. 2025;48(2):140-155. doi:10.1016/j.tins.2024.12.004
 53. Park E, Kuljis DA, Swindell RA, et al. Somatostatin neurons detect stimulus-reward contingencies to reduce neocortical inhibition during learning. *Cell Rep*. 2025;44(5):115606-115606. doi:10.1016/j.celrep.2025.115606
 54. Khoury CF, Fala NG, Runyan CA. Arousal and Locomotion Differently Modulate Activity of Somatostatin Neurons across Cortex. *eNeuro*. 2023;10(5):1-20. doi:10.1523/ENEURO.0136-23.2023
 55. Urban-Ciecko J, Jouhanneau JS, Myal SE, Poulet JFA, Barth AL. Precisely Timed Nicotinic Activation Drives SST Inhibition in Neocortical Circuits. *Neuron*. 2018;97(3):611-625.e5. doi:10.1016/j.neuron.2018.01.037
 56. Bugeon S, Duffield J, Dipoppa M, et al. A transcriptomic axis predicts state modulation of cortical interneurons. *Nature*. 2022;607(7918):330-338. doi:10.1038/s41586-022-04915-7
 57. Furlanis E, Dai M, Leyva Garcia B, et al. An enhancer-AAV toolbox to target and manipulate distinct interneuron subtypes. *Neuron*. 2025;113(10):1525-1547.e15. doi:10.1016/j.neuron.2025.05.002
 58. Gao WJ, Wang Y, Goldman-Rakic PS. Dopamine modulation of perisomatic and peridendritic inhibition in prefrontal cortex. *J Neurosci*. 2003;23(5):1622-1630. doi:10.1523/jneurosci.23-05-01622.2003

59. Dienel SJ, Dowling KF, Barile Z, et al. Diagnostic Specificity and Association with Cognition of Molecular Alterations in Prefrontal Somatostatin Neurons in Schizophrenia. *JAMA Psychiatry*. 2023;80(12):1235-1245. doi:10.1001/jamapsychiatry.2023.2972
60. Minxha J, Adolphs R, Fusi S, Mamelak AN, Rutishauser U. Flexible recruitment of memory-based choice representations by the human medial frontal cortex. *Science*. 2020;368(6498):39-39. doi:10.1126/science.aba3313
61. Tsai MC, Teutsch J, Wybo WAM, Helmchen F, Gr P. Hierarchy of prediction errors shapes the learning of context-dependent sensory representations. Published online 2024:1-32.
62. Popescu AT, Zhou MR, Poo MM. Phasic dopamine release in the medial prefrontal cortex enhances stimulus discrimination. *Proc Natl Acad Sci U S A*. 2016;113(22):E3169-E3176. doi:10.1073/pnas.1606098113
63. Jahn CI, Markov NT, Morea B, Daw ND, Ebitz RB, Buschman TJ. Learning attentional templates for value-based decision-making. *Cell*. 2024;187(6):1476-1489.e21. doi:10.1016/j.cell.2024.01.041
64. Lak A, Okun M, Moss MM, et al. Dopaminergic and Prefrontal Basis of Learning from Sensory Confidence and Reward Value. *Neuron*. 2020;105(4):700-711.e6. doi:10.1016/j.neuron.2019.11.018
65. Cole N, Harvey M, Myers-Joseph D, Gilra A, Khan AG. Prediction error signals in anterior cingulate cortex drive task-switching. *bioRxiv*. 2022;(April):2022.11.27.518096-2022.11.27.518096. doi:10.1038/s41467-024-51368-9
66. Roach JP, Churchland AK, Engel TA. Choice selective inhibition drives stability and competition in decision circuits. *Nat Commun*. 2023;14(1):1-16. doi:10.1038/s41467-023-35822-8
67. Khan AG, Poort J, Chadwick A, et al. Distinct learning-induced changes in stimulus selectivity and interactions of GABAergic interneuron classes in visual cortex. *Nat Neurosci*. 2018;21(6):851-859. doi:10.1038/s41593-018-0143-z
68. Ashwood ZC, Roy NA, Stone IR, et al. Mice alternate between discrete strategies during perceptual decision-making. *Nat Neurosci*. 2022;25(February). doi:10.1038/s41593-021-01007-z
69. Anastasiades PG, Boada C, Carter AG. Cell-Type-Specific D1 Dopamine Receptor Modulation of Projection Neurons and Interneurons in the Prefrontal Cortex. *Cereb Cortex*. 2019;29(7):3224-3242. doi:10.1093/cercor/bhy299

70. Najafi F, Elsayed GF, Cao R, et al. Excitatory and Inhibitory Subnetworks Are Equally Selective during Decision-Making and Emerge Simultaneously during Learning. *Neuron*. 2020;105(1):165-179.e8. doi:10.1016/j.neuron.2019.09.045
71. Gasselino C, Hohl B, Vernet A, Crochet S, Petersen CCH. Cell-type-specific nicotinic input disinhibits mouse barrel cortex during active sensing. *Neuron*. 2021;109(5):778-787.e3. doi:10.1016/j.neuron.2020.12.018
72. Keller AJ, Dipoppa M, Roth MM, et al. A Disinhibitory Circuit for Contextual Modulation in Primary Visual Cortex. *Neuron*. 2020;108(6):1181-1193.e8. doi:10.1016/j.neuron.2020.11.013
73. Dana H, Sun Y, Mohar B, et al. High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nat Methods*. 2019;16(7):649-657. doi:10.1038/s41592-019-0435-6
74. Zhang Y, Rózsa M, Liang Y, et al. *Fast and Sensitive GCaMP Calcium Indicators for Imaging Neural Populations*. Vol 615. Springer US; 2023:891. doi:10.1038/s41586-023-05828-9
75. Peixoto D, Verheine JR, Kiani R, et al. Decoding and perturbing decision states in real time. *Nature*. 2021;591(7851):604-609. doi:10.1038/s41586-020-03181-9
76. Chen G, Kang B, Lindsey J, et al. Single neurons in prefrontal cortex encode abstract rules. *Cell*. 2021;184(14):3717-3730.e24. doi:10.1016/j.cell.2021.05.026
77. Nogueira R, Esteki S, Fusi S, Kiani R. The geometry of context-dependent biased decisions during learning. Published online 2026.
78. Green J, Bruno CA, Traunmüller L, et al. A cell-type-specific error-correction signal in the posterior parietal cortex. *Nature*. 2023;620(7973):366-373. doi:10.1038/s41586-023-06357-1
79. Cristino MH, Mitchell AC, Preibisz-Kamat M, Fletcher PS, Spellman TJ. An Open-Source 3D-Printable Platform for Testing Head-Fixed Cognitive Flexibility in Rodents. *eneuro*. 2025;12(1):ENEURO.0364-24.2024. doi:10.1523/ENEURO.0364-24.2024
80. Pachitariu M, Packer AM, Pettit N, Dalgleish H, Hausser M, Sahani M. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv*. Published online 2016:1-30. doi:10.1101/061507
81. Sheintuch L, Rubin A, Brande-Eilat N, et al. Tracking the Same Neurons across Multiple Days in Ca²⁺ Imaging Data. *Cell Rep*. 2017;21(4):1102-1115. doi:10.1016/j.celrep.2017.10.013