Title
Orbitofrontal and Hippocampal Contributions to State Coding and Decision Making

Author
Ford, Celia Fernandez

Publication Date
2023

Peer reviewed|Thesis/dissertation

Orbitofrontal and Hippocampal Contributions to State Coding and Decision Making

by

Celia Fernandez Ford

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Neuroscience

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Joni Wallis, Chair
Professor Anne Collins
Professor David Foster
Professor Rich Ivry

Summer 2023

Orbitofrontal and Hippocampal Contributions to State Coding and Decision Making

Abstract

Orbitofrontal and Hippocampal Contributions to State Coding and Decision Making

by

Celia Fernandez Ford

Doctor of Philosophy in Neuroscience

University of California, Berkeley

Professor Joni Wallis, Chair

To make informed decisions, one needs to understand the world in which those decisions are being made: the current task state. Both the orbitofrontal cortex (OFC) and the hippocampus (HPC) have been implicated in representing task state, but the nature of these representations, and what role each region plays, remains unclear. Recent work from our lab showed that HPC-driven neural oscillations in OFC are critical for value-guided decision making, and that value states in OFC oscillate during the deliberation process. We hypothesized that (1) both HPC and OFC robustly and flexibly represent task state while primates learn and make decisions, and (2) value state dynamics in OFC are shaped by theta oscillations originating from HPC.

We simultaneously recorded large populations of neurons from OFC and HPC while two monkeys performed a probabilistic reversal learning task, where reward contingencies could be captured by two task states. Using population-level decoding, we found neural representations of task state in both OFC and HPC that remained stable within each trial but strengthened with learning as monkeys adapted to reversals. Subjects also appeared to use their understanding of task structure to anticipate reversals, evidenced by anticipatory neural representations of the upcoming task state. We then decoded representations of option values from OFC neurons recorded from two other monkeys performing a value-based decision-making task. Relating decoded state dynamics to theta-filtered LFP signals in OFC, we found that value states switched independent from theta cycles, suggesting that a mechanism other than HPC-driven theta oscillations shapes the dynamics of deliberation.

To Charlotte

Contents

List of Figures

List of Tables

Acknowledgements

This work is dedicated to my mentors, my community, and my chosen family.

Thank you to Joni Wallis for guiding me through this project—through multiple course-altering failures, a global pandemic, and many moments of existential despair. Thank you for believing that there's always something left to discover.

Thank you to my lab mates—Eric Knudsen and Zuzanna Balewski for more hours of mentorship than I could possibly count; Thom Elston, Eric Hu, Lauren Meckler, and Nate Munet for making the LKS basement feel less terrible. I lucked out, joining a lab filled with such good eggs. And the biggest thanks imaginable to Don and Victor, who did not choose to participate in this work, yet made the greatest contribution of all.

Thank you to the OLAC staff—nonhuman primate research is physically taxing and emotionally draining and would be impossible without your attention and care.

Thank you to my cohort, my thesis committee, and the rest of the HWNI graduate program. The friendships I made here have shaped me in such beautiful ways. I'm sorry for sending so many voice memos (especially to you, Willa).

Thank you to the communities that embraced me as a whole person beyond academia, especially as my path veered ever further off-course. To my pole siblings, teachers, and students who bring me strength and grace; to my bandmates, jam buddies, and artists who bring me magic and joy—you make this place feel like home.

Thank you to my family in Orlando, without whom I would never have been prepared to consider graduate school at all. You raised me to be confident, curious, and kind, and supported me in every interest I've passed through—thank you for investing in me.

Thank you to Roan, for entering and re-entering my life exactly when you were meant to. Thank you for your unconditional support and laughter and joy and excitement and ambition and love—I feel like more of myself when I'm with you.

And as I close this chapter and start to look ahead, I'm overwhelmed with gratitude for Siri Carpenter and Sabrina Imbler, who took a chance on me as a writer. Thank you to *Carry the One Radio* and the *Berkeley Science Review* for providing me with outlets to explore science communication—I finally know what I want to be when I grow up.

Of course, thank you to Charlotte, my most devoted research assistant. You bring so much love into my life, and make the long, uncertain days feel a little shorter.

Chapter 1

Introduction

To make informed decisions, one needs to understand the world in which those decisions are being made. For example, when standing in front of your closet in the morning, your outfit choice will depend heavily on where you're planning to be; the shorts you'd wear to excel at a dance class likely won't be as optimal for delivering an exit talk. In a reinforcement learning framework, *task states* are aspects of the environment (e.g., whether you're at a dance studio vs. a conference room) that indicate what actions are likely to be most rewarding (e.g., wearing shorts vs. pants) (Langdon et al., 2019; Niv, 2019; Sutton & Barto, 1998). Over time, one can learn the likelihood of one state leading to another, and represent transitions between states as an abstract cognitive map (Behrens et al., 2018; Knudsen & Wallis, 2022; Whittington et al., 2022).

Historically, the orbitofrontal cortex (OFC) has been linked to a wide range of cognitive functions (Stalnaker et al., 2015) including reward- and value-based decision-making (Padoa-Schioppa & Assad, 2006; Rich & Wallis, 2016), response inhibition (Izquierdo & Jentsch, 2012; Jones & Mishkin, 1972), and signaling emotion (Bechara et al., 1997), but more recent evidence suggests OFC function can be better encapsulated as task state representation (Chan et al., 2016; Schuck et al., 2016; Takahashi et al., 2011; Wilson et al., 2014). The hippocampus (HPC) is also implicated in task state representation and the process of inferring ones current state (Sanders et al., 2020), building on the long-existing theory that HPC stores a cognitive map that flexibly encodes causal relationships between internal motivations and external observations (O'keefe & Nadel, 1978; Tolman, 1948).

While the precise role of each region in representing task states is unclear, one possibility is that behavioral tasks, following rules to select actions that maximize potential reward, can be represented as state-transition maps defining how states relate to one another. One possibility is that HPC constructs and represents these transition graphs, while OFC represents the current goal and pulls information from HPC to calculate the value of states given that goal (Knudsen & Wallis, 2022). Recent work from our lab supports this prediction, demonstrating that HPC-driven theta oscillations in OFC are critically important for learning and adapting to change during value-guided decision-making (Knudsen & Wallis, 2020). We have also observed oscillations in OFC value states during the deliberation process, where populations of neurons vacillate between representations of each available option value leading up to the time of choice (Balewski et al., 2022; Rich & Wallis, 2016).

Here, we sought to understand how task states are represented in the brain, how these representations are updated when the environment changes, and how they shape the deliberation process during value-based decision-making. Given parallel bodies of research implicating OFC and HPC in task state, and the strength of their anatomical and functional connections, we first searched for evidence that both regions robustly and flexibly represent task state while primates learn and make decisions (Chapter 3). We simultaneously recorded large populations of neurons from OFC and HPC while two monkeys performed a probabilistic reversal learning task, where reward contingencies could be captured by two task states. Using population-level decoding, we found neural representations of task state in both OFC and HPC that remained stable within each trial but strengthened with learning as monkeys adapted to reversals. Subjects also appeared to use their understanding of task structure to anticipate reversals, evidenced by anticipatory neural representations of the upcoming task state.

We then tested a prediction about the interaction between OFC and HPC (Chapter 4): do theta oscillations originating from HPC shape value state dynamics in OFC? We ran new analyses on previously reported data (Balewski et al., 2022) to decode representations of option values from large populations of OFC neurons recorded from monkeys performing a value-based decision-making task. Relating decoded state dynamics to theta-filtered LFP signals in OFC, we found that value states switched independent from theta cycles, suggesting that a mechanism other than HPC-driven theta oscillations shapes the dynamics of deliberation.

Outline

Chapter 2, *Model-based acute electrophysiology in behaving nonhuman primates,* presents an overview of our methods for high-throughput recording from deep brain regions. Our decoding approach to understanding decision-making requires simultaneously recording from a large number (~100) of neurons across two regions each session. We design, 3D print, and assemble unique recording platforms for each session to densely and precisely arrange multiple multi-channel probes (Plexon: V- and K-probes) to target our regions of interest.

Chapter 3, *Orbitofrontal-hippocampal state coding dynamics during reversal learning,* explores the role of the orbitofrontal cortex (OFC) and the hippocampus (HPC) in contextualizing information and identifying links between stimuli, actions, and outcomes. We recorded simultaneously from both regions while primates performed a probabilistic reversal learning task and found that both OFC and HPC stably represent task state within trials. These representations strengthened with learning and switched in anticipation of upcoming reversals, implicating both regions in state inference.

Chapter 4, *Theta oscillations do not explain dynamic activation of value states in the orbitofrontal cortex,* presents results from a new analysis of data previously published as Balewski, Knudsen, and Wallis (*Neuron* 2022). We tested the hypothesis that OFC value dynamics, like those first reported by Rich and Wallis (*Nature Neuroscience* 2016), are driven by theta oscillations. We found that while value states were more likely to be decoded at certain phases of the theta cycle, there was no relationship between OFC value dynamics and theta oscillations.

Chapter 2

Model-based acute electrophysiology in behaving nonhuman primates

Ford, Celia F., Knudsen, Eric B., Balewski, Zuzanna Z., & Wallis, Joni D.

2.1 Introduction

This thesis aims to investigate how neural representations of task states are formed, how these representations evolve during learning, and how task states inform value-based decision making. To study dynamic cognitive processes like this, where averaging neural activity across trials would wash out the effects we are looking for, we set out to record from many single neurons simultaneously. That activity can then be used to train decoders to classify the information represented by recorded neural populations. Decoder accuracy depends largely on the quantity and quality of its training data (Wallis, 2018), so we used a system, described in depth in Knudsen et al. (2019), that enabled us to lower up to three linear multichannel electrodes in each of two target regions—the orbitofrontal cortex (OFC) and the hippocampus (HPC)—per session. This approach not only allowed us to record from a large number (~100) neurons at once, but also to broadly sample neurons from each region while minimizing tissue damage, improving both the quality and quantity of our data relative to traditional approaches.

Here, we describe our approach for using anatomical models to target deep brain structures in awake, behaving monkeys. Methods for preprocessing and analyzing behavioral and neuronal data are described in detail in Chapters 2 and 3.

2.2 Anatomical modeling

We created models of each subject's skull, brain regions of interest, and future recording chamber to guide their chamber implant surgery and craniotomy and inform recording probe placement. Subjects were anesthetized, fitted with head positioners, and imaged in a 3T MRI scanner. We manually traced regions of interest (Fig. 2.1A)—areas 11, 12, and 14 of the orbitofrontal cortex (OFC), and CA1 and CA2/3 of the hippocampus (HPC)—in Slicer3D (Fedorov et al., 2012), referencing coronal slices from a rhesus macaque brain atlas in stereotaxic coordinates (Paxinos et al., 2000). We then derived cranial models (Fig. 2.1B) from these anatomical images to inform our placement of custom-fit polyether-ether-ketone (PEEK) unilateral acute recording chambers (Gray Matter

Research, Bozeman, MT). Chambers were milled using traditional approaches and surgically implanted over the left hemisphere of one subject (D) and the right hemisphere of another subject (V). Both craniotomies were roughly 25 mm long (anterior-posterior) and 15 mm wide (medial-lateral), granting access to both the OFC and the HPC via vertical trajectories from the chamber surface (Fig. 2.1C).

2.3 Calculating electrode trajectories

In the absence of tools for visually tracking electrodes to deep brain regions as they are lowered, we used labeled MRI images and 3D models of the chamber and recording platform to determine the starting position and vertical trajectory of each probe. To identify points where an electrode could reach OFC or HPC via a vertical path through the craniotomy, we superimposed a grid (points evenly spaced 1 mm apart) on a model of the skull, chamber, and brain regions of interest (Fig. 2.2A). After verifying that lines passing through each point would intersect with our targets (Fig. 2.2B), we translated the coordinates of each site on the recording platform to MRI space, pulling brain slices aligning with the anterior-posterior position of each electrode. We then plotted a vertical trajectory at the medial-lateral probe position corresponding to each slice to visualize expected gray-to-white matter transitions between the dura and the target depth (Fig. 2.2C). This map was used as a reference during recording. To minimize recording from damaged tissue, we avoided repeating target coordinates to the best of our ability. Balancing this goal with the need to maximize efficiency in the prototyping and platform-building process, we designed chamber attachments ("collars," see Fig. 2.3B) that allowed us to shift the entire recording system 1 mm in the anterior, posterior, medial, or lateral direction. So, each recording system design could be reused up to four times by shifting the entire assembly 1 mm in any cardinal direction.

2.4 Building a modular recording system

During each recording session, we simultaneously lowered up to three multisite linear probes (V/K-probes, Plexon Inc., Dallas, TX), each configured with 32 recording sites spaced 100 μm apart (Fig. 2.3A). Unlike electrode arrays, which are designed for surface-level recordings, linear probes enabled us to reach deeper brain regions several centimeters below the dura. Using multisite probes also increases our neuronal yield relative to the number of insertion sites, representing a strong advantage over conventional tungsten microelectrodes. However, while tungsten probes are pliable, multisite probes are hollow and thus more fragile. The multisite probes also interface directly with the headstage (tungsten probes do not). To account for the size and rigidity of our electrodes, we aimed to design a recording system that would enable us to fit as many electrodes into our regions of interest as possible, while ensuring that each

electrode, and its corresponding headstage, could follow a path orthogonal to the chamber surface to prevent breakage.

We created CAD models (Onshape, Boston, MA) of our recording structure, including a platform fitted to the chamber, stabilizing towers to protect each probe, and shuttles enabling probe movement. Each designed component was printed in-house using the Form 2 or Form 3 stereolithographic (SLA) printer (Form Labs, Cambridge, MA). Probes were lowered into the brain using a shuttle with a long arm supporting the length of the hypodermic tubing protecting the probe shank (Fig. 2.3B). The shuttle is maneuvered by a 0-80 drive screw that lowers it at a rate of 330 μm per turn. The drive screw is encased in a stabilizing tower. With this configuration, we were able to fit up to six probes on the recording platform at once.

On the day of recording, we sterilized the electrodes and recording surface, cleaned the chamber interior, and administered topical (lidocaine) and/or oral (meloxicam) analgesics prior to attaching the recording system to the chamber (Fig. 2.3B depicts the chamber, collar, and recording system together). Probes were lowered approximately 1.5 mm at a time, in turn, to minimize tissue compression. Neuronal activity was monitored by displaying waveforms on a monitor in real time (OmniPlex Neural Recording Data Acquisition System, Plexon Inc., Dallas, TX) and playing waveforms from selected channels as sound. White and gray matter transitions, audible as transitions between noise and spiking activity, were used to track progress towards the target region. After all probes were lowered to their final positions, we allowed tissue to settle for roughly one hour[1] before making final adjustments and initiating the behavioral task.

[1] While probes were being lowered, and during this one-hour waiting period, monkeys watched their favorite shows on Netflix. Subject D preferred nature documentaries, especially films following the social drama of rhesus macaques in the Himalayas. Subject V was a sci-fi buff who loved action and vivid colors and powered through the Star Wars: Clone Wars series during his tenure.

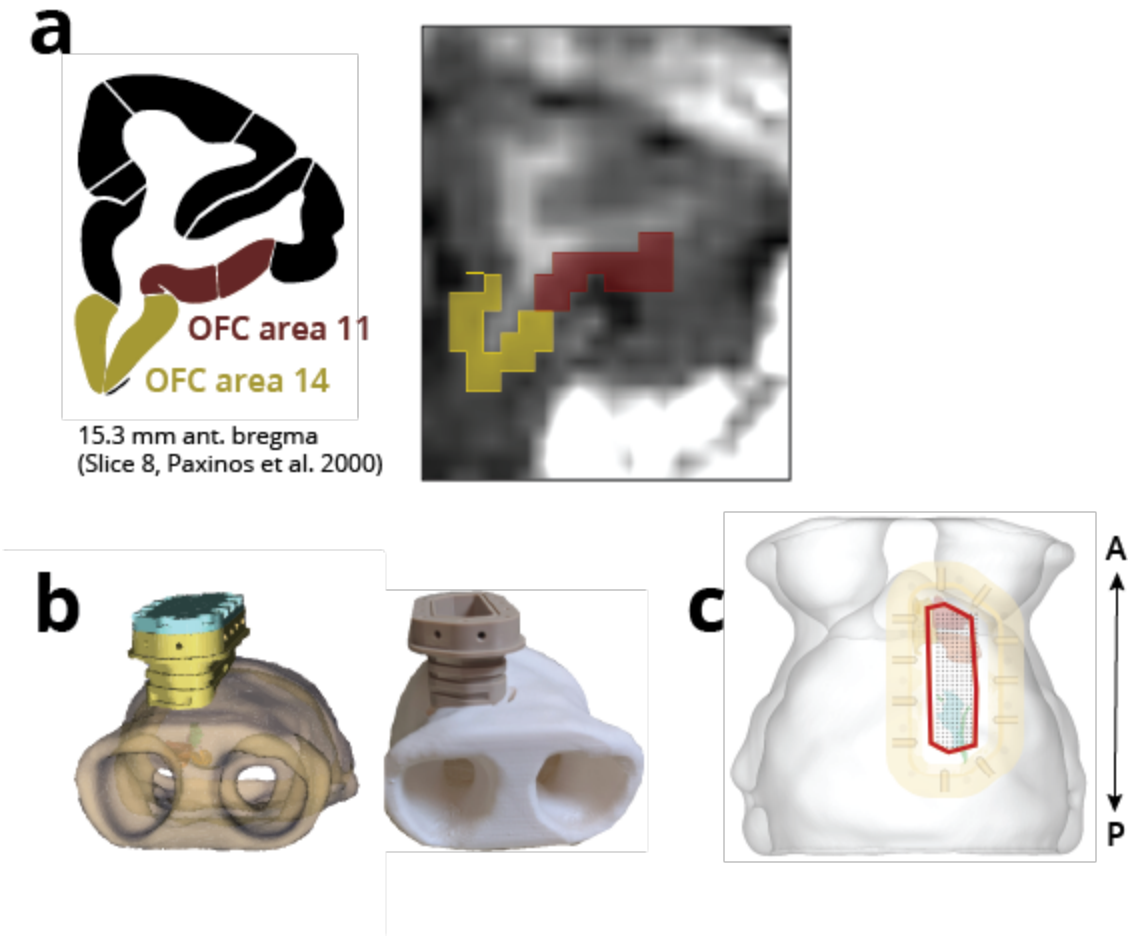

Figure 2.1. MRI-guided anatomical modeling and chamber placement. (A) A sample coronal brain slice from the macaque brain atlas (left) used to locate and trace two regions of interest (red = OFC area 11, gold = OFC area 14) in the corresponding coronal slice from subject V's MRI scan (right). (B) Cranial model with a custom-fit acute recording chamber (left = model in Slicer3D, right = photograph of subject V's PEEK chamber placed on a 3D-printed skull model). (C) Transparent cranial model (subject V) including chamber and regions of interest (warm colors = OFC, cool colors = HPC). The craniotomy site is outlined in red - both OFC and HPC are accessible via vertical electrode trajectories.

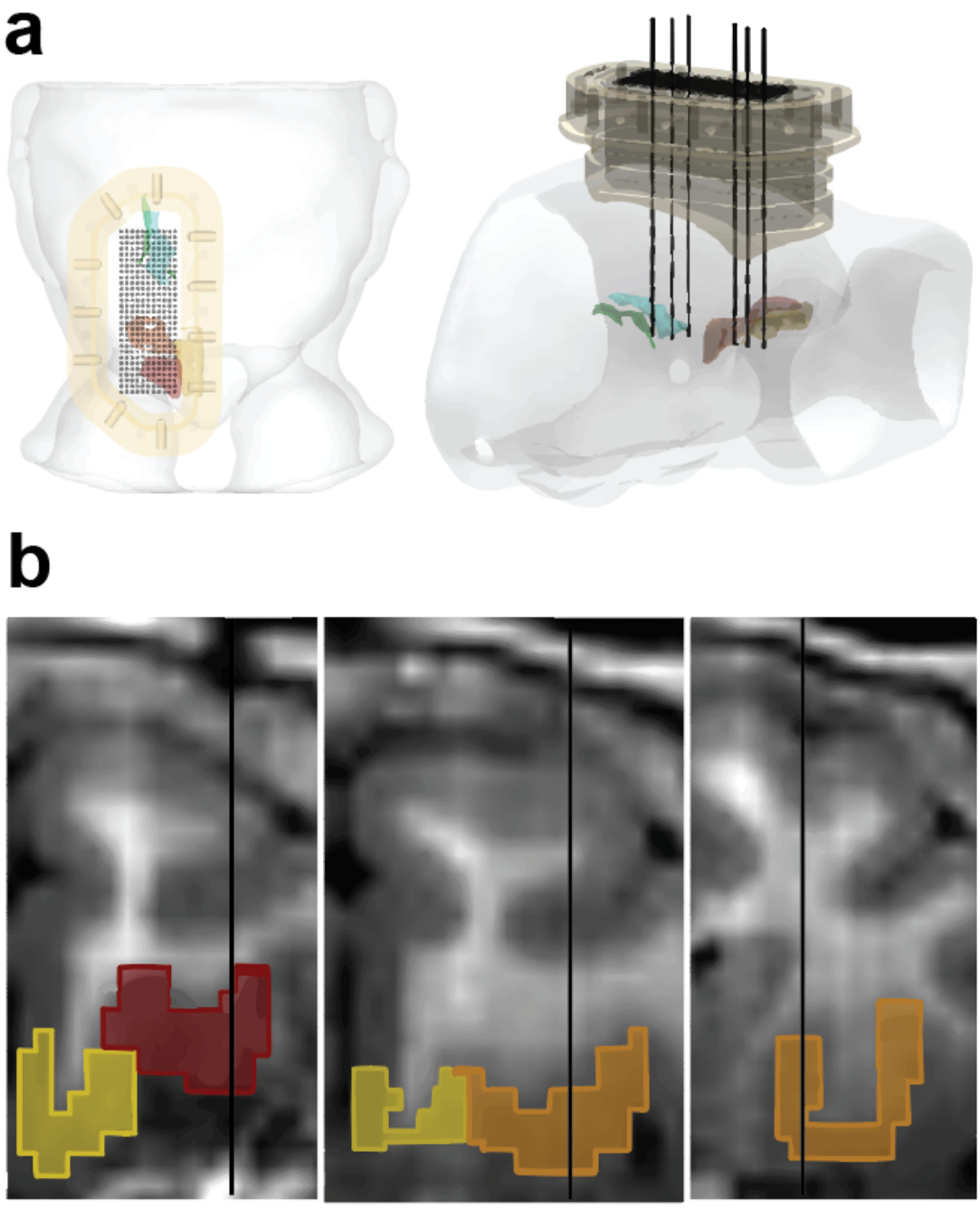

Figure 2.2. Planning electrode trajectories. (a) 3D cranial model with superimposed chamber implant (beige) and grid layout (black dots, spaced 1 mm apart) over MRI-derived models of CA1 and CA2/3 (cool colors) and cytoarchitectonic areas 11, 13, and 14 of the OFC (warm colors). The left panel illustrates electrode trajectories as black lines orthogonal to the recording platform, each intersecting with a point on the simulated grid. (b) Probe trajectories superimposed on labeled orbitofrontal MRI slices, from anterior (left) to posterior (right). Expected distance from the dura and transitions between white and gray matter were used to verify electrode placement during recording.

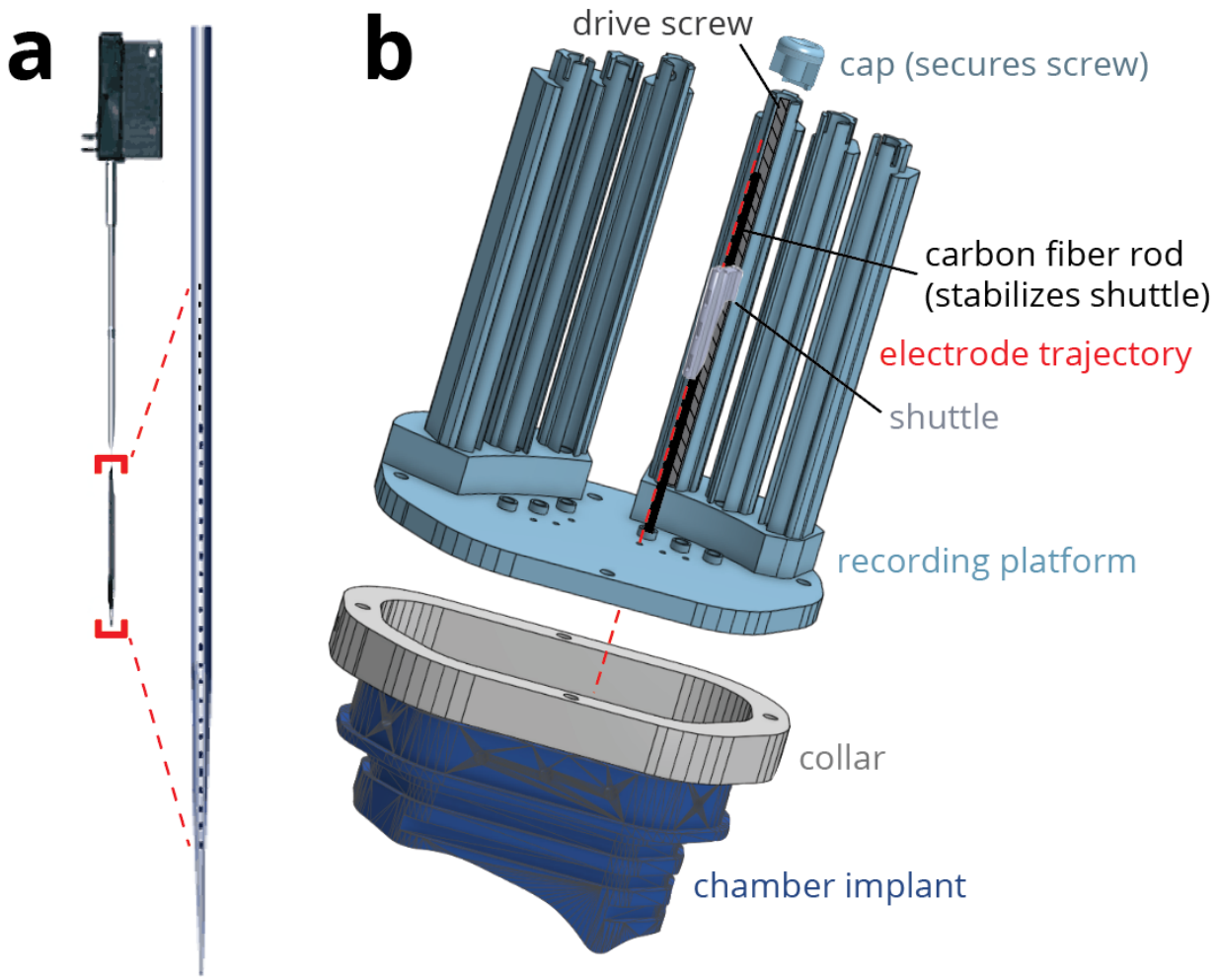

Figure 2.3. Building a modular recording system. (A) Illustration of a 32-channel V-probe from Plexon Inc., Dallas, TX. (B) Labeled CAD design of a recording system used with subject V. The red dashed line represents where a single electrode would be placed. A guide tube cut from a hypodermic needle was placed in each of the small holes illustrated on the recording platform, to pierce the dura and protect the delicate probe tip from damage prior to recording.

Chapter 3

Orbitofrontal-hippocampal state coding dynamics during reversal learning

Ford, Celia F., Balewski, Zuzanna Z., & Wallis, Joni D.

To build an understanding of our world, we make inferences about the connections between our actions, experiences, and the environment. This process, *state inference*, requires an agent to guess the current state of the world given a set of observations. During value-based decision-making, a growing body of evidence implicates the orbitofrontal cortex (OFC) and the hippocampus (HPC) in the process of contextualizing information and identifying links between stimuli, actions, and outcomes. However, the neural mechanisms driving these processes in primates remain unknown. To investigate how OFC and HPC contribute to state inference, we recorded simultaneously from both regions while monkeys performed a probabilistic reversal learning task, where reward contingencies could be captured by two task states. Using population-level decoding, we found neural representations of task state in both OFC and HPC that remained stable within each trial but strengthened with learning as monkeys adapted to reversals. Subjects also appeared to use their understanding of task structure to anticipate reversals, evidenced by anticipatory neural representations of the upcoming task state.

3.1 Introduction

In a reinforcement learning framework, task state is defined as the set of landmarks and rules that characterize an agent's current environment (Sutton & Barto, 1998); (Niv, 2019). To make optimal decisions in a dynamic world, animals must make inferences about their environment, and identify task states that can explain recent observations and inform future decisions (Gershman, Norman, & Niv, 2015). These states can then be linked together to form an abstract cognitive map that specifies the likelihood of one state leading to another state (Behrens et al., 2018; Knudsen & Wallis, 2022; Whittington et al., 2022). By constructing distinct state representations that capture features of the environment that dictate whether actions will be rewarding, we can learn to effectively behave and adapt to change (Langdon et al., 2019). Recent evidence suggests that the cognitive map may be constructed, housed, and updated through interactions between the orbitofrontal cortex and the hippocampus (Wikenheiser & Schoenbaum, 2016; Wilson et al., 2014; Schuck et al., 2016). However, the nature of these representations, and what role each region plays in representing task structure, remains unclear.

The OFC has been linked to a wide range of cognitive functions (Stalnaker et al., 2015), including reward- and value-based decision-making (Padoa-Schioppa & Assad, 2006; Rich & Wallis, 2016), response inhibition (Jones & Mishkin, 1972; Izquierdo & Jentsch, 2012), and signaling emotion (Bechara et al., 1997). In the past decade, studies in rodents (Takahashi et al., 2011), humans (Chan et al., 2016; Schuck et al., 2016), and artificial agents (Wilson et al., 2014) point towards task state representation as a broad function that encapsulates all of these more parsimonious roles. The hippocampus (HPC) is also implicated in task state representation and hidden state inference (Sanders et al., 2020), and has long been theorized to store a cognitive map that flexibly encodes causal relationships between internal motivations and external observations (Tolman, 1948; O'keefe & Nadel, 1978). Many studies, especially in rats, highlight the importance of HPC in spatial navigation (Hartley et al., 2014), while lesion and functional neuroimaging studies in humans tend to focus on the role HPC plays in episodic memory (Gelbard-Sagiv et al., 2008; Chadwick et al., 2010). Framing the HPC as a flexible system for representing relationships between the internal and external worlds could unify these seemingly contradictory accounts of HPC function (Eichenbaum, 2017), and points toward a role in task state representation mirroring that proposed in the OFC.

Despite striking similarities between the proposed functions of the OFC and the HPC, few studies have tested their contributions to state inference during value-based decision-making in parallel. Furthermore, what little we do know has largely been derived from the rodent model (Zhou, Gardner, et al., 2019; Zhou, Montesinos-Cartagena, et al., 2019; Wikenheiser & Schoenbaum, 2016). There have been dramatic changes in the structure of HPC across the course of mammalian evolution, particularly in the size and complexity of anterior HPC in primates (Insausti, 1993), which shows strong anatomical and functional connectivity with OFC (Barbas & Blatt, 1995; Young & Shapiro, 2011). This makes it difficult to extrapolate findings from the rodent to the primate. To mitigate these issues, we recorded simultaneously from OFC and HPC of two non-human primates while they performed a reward-based learning task.

Here, we trained two monkeys (*Macaca mulatta*) to perform a probabilistic reversal-learning task where they were challenged to adapt their decision-making behavior to uncued changes in reward contingencies. We used this task since it is a classic test of OFC function. Rodents (Schoenbaum et al., 2003), monkeys (Dias et al., 1996; Iversen & Mishkin, 1970), and humans (Fellows & Farah, 2003) with OFC damage, have no problem with the initial learning of such contingencies, but show severe impairments when contingencies reverse, continuing to perseverate by choosing the previously rewarded outcome. Recent theoretical explanations of these findings have invoked the concept of a cognitive map, whereby reversal is not simply the unlearning of the original association and the learning of the new contingency. Instead, the subject learns that there are two

'states' of the world, one in which stimulus A is rewarded and B is not, and one in which stimulus B is rewarded and A is not (Wilson et al., 2014). We found neurons in both OFC and HPC that encoded task state stably within each trial, and then used population-level decoding to further explore the dynamics of these task state representations and examine how they interact with neural representations of reward.

3.2 Results

Behavioral analysis

We trained two monkeys (subjects D and V) on a probabilistic reversal task (Fig. 3.1). Subjects took a median of 44 (D) and 70 (V) trials to perform a single reversal and completed a median of 16 (D) and 12 (V) reversals per session (Fig. 3.2a-c). To gain a better understanding of the behavioral strategies used by each subject during task performance, we studied the effect of reward outcome, given the reward probability associated with the chosen option, on the decision made on the subsequent trial. Because pictures within each set always had the same value, we reasoned that if subjects understood the structure of the task, they would apply what they learn from choosing one picture to its pair.

To see whether subjects generalized across pictures within the same set, we split trial pairs into those where the same picture was chosen on trial t+1 as trial t, and those where the paired picture was chosen on trial t+1. We restricted this analysis to free-choice trial pairs where the same picture chosen on trial t was available on trial $t+1$, or the paired picture was available on trial $t+1$. These conditions were always mutually exclusive – all free-choice trials presented one picture from each set. If the likelihood of choosing a paired picture or the same picture on trial $t+1$ are similarly influenced by the outcome of trial t, it would point towards a model-based reinforcement learning strategy. Subjects could use their understanding of task structure to generalize across pictures within each pair and update the expected values of paired pictures in sync. On the other hand, if the likelihood of choosing a paired picture is unaffected by the result of choosing its pair on the previous trial, it would suggest subjects were tracking the values of each picture separately, without employing a more general task model.

We calculated the difference in the probability of repeating a previous choice on rewarded trials over unrewarded trials, to isolate the effect of reward on future choices across different trial conditions (Fig. 3.2d). In general, reward was more likely to drive a repeat choice on trials where the high-value option (i.e., the most likely option to be rewarded) was chosen. This suggests that after each reversal, subjects quickly learned that rewarded low-value choices were unlikely and shouldn't be repeated. In general, subjects became more likely to repeat rewarded high-value options over the course of learning, whether

the same picture (one-way ANOVA, monkey V: $F(2,21) = 11.96$, $p < 0.001$; monkey D: n.s.) or the paired picture (one-way ANOVA, monkey V: $F(2,21) = 16.08$, $p < 0.001$; monkey D: $F(2,15) = 25.76$, $p < 0.001$) was available on trial $t+1$. In later learning stages, there was also no difference in the likelihood of repeating rewarded over unrewarded low-value choices. Together, these patterns of behavior point towards the use of model-based strategies during reversal learning.

Neural representations of task state strengthen with learning and in anticipation of change

We recorded single neurons from OFC (D: N = 224, V: 346) and HPC (D: 342, V: 580), using up to 3 acute multisite probes per region per session (Fig. 3.3a). To examine the relationship between learning, choice behavior, and neuronal firing rates across and within trials, we performed linear regressions on firing rates aligned to the time of picture onset and reward outcome. We fit firing rates in sliding time windows with task state, choice direction, chosen value, chosen picture, trial type (forced or free) and reward as predictors (Fig. 3.3b, see Methods, Eq. 3.1-3.2). The firing rates of most neurons were significantly predicted by at least one model term in one task epoch in both OFC (D: 177/224 or 79%, V: 193/346 or 56%) and HPC (D: 300/342 or 88%, V: 395/580 or 68%).

Figure 3.4 illustrates examples of single neuron selectivity in OFC and HPC. The firing rate of some neurons was driven by a single factor, such as the reward-selective neuron in OFC (Fig. 3.4a). However, many neurons in OFC (D: 92/224 or 41%, V: 76/346 or 22%) and HPC (D: 166/342 or 49%, V: 146/580 or 25%) encoded more than one factor. For example, the OFC neuron in Figure 3.4b encoded both the state and the choice response, while the HPC neuron in Figure 3.4d encoded the choice response and the reward outcome. Many neurons also encoded the current state and maintained this information across all epochs of the task (e.g., Fig. 3.4c and 3.4e). In subject D, these neurons tended to be more prevalent in HPC than OFC (Fig. 3.3b), but they were equally prevalent in both areas in subject V.

To examine how state information responded to reversals, we used linear discriminant analysis (LDA) to decode trial-by-trial estimates of state from population level neuronal activity (see Methods). Mean decoder accuracy was above chance (50%) for all sessions and remained relatively constant across trial events. In general, decoders trained on HPC activity were more accurate (V: 60.5% ± 2.7%, D: 61% ± 1.6%) than those trained on OFC (V: 56.5% ± 2%, D: 55.5% ± 1.1%, two sample t-test, D: $p = 0.02$, V: n.s.). However, we also typically had larger neuronal sample sizes in HPC. When we downsampled HPC ensembles to match those in OFC, there was no consistent difference in decoder performance between the two areas across sessions ($p > 0.1$, permutation test for all epochs in both subjects, Fig. 3.8).

To determine how state representations in OFC and HPC relate to behavior, we examined how state decoding evolved over the course of learning. Blocks varied in decoder accuracy and confidence, but generally, decoder accuracy evolved in parallel with behavior (Fig. 3.5a). Immediately following a state reversal, behavioral performance usually dipped below chance as subjects continued erroneously choosing the best option for the pre-reversal state. Once subjects detected the state change, their performance improved, eventually meeting the criterion required to trigger the next reversal. Decoder confidence (defined as the normalized mean posterior probability of decoding the current state over the previous state, averaged across post-event timestamps) showed similar dynamics, gradually increasing over the first half of the block as the subject learned that the state had changed (Fig. 3.5b). More surprisingly, decoder confidence *decreased* during the second half of the block despite behavioral performance being near or above criterion by that point in learning. This suggests that subjects are beginning to represent the *next* task state in anticipation of the upcoming reversal. In other words, they recognize that a period of consistent reward (from making optimal choices) will always soon be followed by a state reversal, and this anticipation is reflected in the state represented.

Combined, these findings suggest that the population activity in OFC and HPC encodes information about task state, and that these representations strengthen as the animal becomes more certain that its observations align with that state. Once performance rises to the point where the animal begins to suspect that a reversal may happen soon, OFC and HPC begin to represent the upcoming task state in anticipation of the next block.

Neural state representations drive choice behavior

We next asked whether the speed with which the monkeys adapted to the new task state after a reversal correlated with neural state decoder dynamics. Because the number of trials subjects took to trigger reversals varied appreciably, we were able to sort blocks into one of two groups: those where choice accuracy for the first 10 trials post-reversal was above the 75th accuracy percentile across all blocks ("fast" blocks), and one where it was below the 25th percentile ("slow" blocks). In slow blocks, subjects perseverate on previously optimal choices for up to 40 trials before recognizing the reversal. In fast blocks, on the other hand, subjects adapt to the new state so quickly that their running performance accuracy never dips below chance (Fig. 3.6a).

We visualized the dynamic adaptation of population activity to the new task state by making heatmaps of decoder confidence across time at both trial epochs (picture onset, outcome), averaged over the first 50 and last 20 trials of each block for fast and slow groups. In slow blocks, posteriors took around 30 or so trials to change, whereas in fast blocks, posteriors changed in a handful of trials (Fig. 3.6b). To quantify this effect, we

correlated behavioral performance with changes in the state reported by the decoder. Specifically, we examined whether the posterior probability of decoding the current state over the new state evolved in pace with behavioral performance, such that the decoder gained confidence in the current task state more slowly during blocks where subjects were slow to adapt their choice behavior (and vice versa). We defined performance thresholds for behavior and the decoder, and identified the number of trials it took to hit that threshold within each block. For behavior, we set the threshold to the learning criterion used to trigger probabilistic reversals during recording sessions: choosing the more valuable option on 70% or more of 30 trials and maintaining this performance level for 10 trials. For the decoder, we set the threshold to 70% of the peak decoder confidence (defined as the absolute value of the difference between the probability of decoding the current state relative to the previous state).

The number of trials it took an OFC decoder to reach threshold was significantly correlated with the number of trials it took the subject to reach learning criteria for a given block (Fig. 3.6c). This suggests that the speed with which OFC updates the representation of task state to match the current block informs how quickly behavior adapts after reversals. The results from HPC were inconsistent across subjects, with a significant correlation between behavioral performance and the HPC decoder in subject V, but not subject D.

Neural representations of expected value and reward

Our final analysis of neuronal tuning focused on events at the time of the reward outcome. We examined whether representations of reward in OFC and HPC interacted with expectations of value as informed by the current task state. We grouped trials into four classes: rewarded high-value choices and unrewarded low-value choices (where the outcome should be expected if the current state is known) and rewarded low-value choices and unrewarded high-value choices (where the outcome was unexpected if the current state was known). We trained a decoder to classify trials into these four categories based on neural activity using the same procedure as for the state decoding (see Methods). Over the course of the trial decoder accuracy remained at chance (25%) until the time of the outcome, when it increased substantially (Fig. 3.7a).

To understand decoding performance in more detail, we created confusion matrices illustrating the pattern of errors typically made by our decoders. The most common mistake was to correctly identify the outcome but misclassify the value of the choice. These errors could arise because neurons only weakly encoded the value of the choice (at least relative to the outcome, Fig. 3.2b). However, they might also arise because the subject was misrepresenting the current task state, thereby mistaking a low-value choice for a high-value choice. If this were the case then we might expect these misclassifications

to be most common during the early stages of learning, when the subject is adjusting their behavior in response to the reversal. To test this, we identified "confused" trials, where over 50% of decoded samples (across time bins and LDA iterations) for that trial correctly classified reward while misclassifying expected value. We then split blocks into five equally sized bins of trials and calculated the percentage of confused trials (of all decoded free-choice trials) in each stage. For decoders trained on both OFC and HPC activity, we observed the most confused trials in the earliest learning stage, with the percentage of confused trials decreasing over the course of learning (Fig. 3.7c). Interestingly, we do not see an increase in confusion in the later learning stages, as we observed in our state decoders.

3.3 Discussion

We observed encoding of the current task state in both OFC and HPC that was evident throughout the trial and that was independent of stimulus presentation and reward outcome. The strength with which task state was represented in both regions predicted choice behavior, such that neuronal decoding of state and behavioral performance increased together in the trials following a reversal in reward contingencies. Subjects also appeared to use their understanding of task structure to anticipate reversals, evidenced by anticipatory neural representations of the upcoming task state.

Many neurons in OFC and HPC were also sensitive to the interaction of reward outcome and expected value, and the encoding of these neurons strengthened with learning of the reward contingencies. However, unlike the representations of task state, there was no evidence that value expectations were anticipating future reversals. Thus, although it appeared that neurons preemptively encoded the upcoming task state, they continued to represent the current expected values until evidence (i.e., failure to reliably receive reward for previously rewarding choices) pushed them to update those.

Representing rules, states, and beliefs in the frontal cortex

Our results are consistent with recent ideas arguing that OFC is important for representing task states (Schuck et al., 2018; Stalnaker et al., 2015; Wikenheiser & Schoenbaum, 2016). For example, Wilson and colleagues tested RL models on an array of classic benchmark tests of OFC function, and found that preventing their models from representing task state made them perform similarly to animals with OFC lesions (Wilson et al., 2014). Indeed, lesioning or temporarily deactivating medial OFC prevents rats from succeeding at tasks like outcome devaluation, where animals must rely on partially hidden information to make optimal decisions (Bradfield et al., 2015).

However, one challenge to these ideas is that there are many semantically related ideas to states, including rules (Wallis, Anderson, et al., 2001), categories (Freedman et al., 2001; Miller et al., 2002), contexts (Rainer et al., 1998), and beliefs (Zhu et al., 2012). Implementation of these functions has typically not been ascribed to OFC, but rather lateral prefrontal cortex (Miller et al., 2003). One distinction that has been proposed argues that states reflect abstract representations of the observable and unobservable properties of the current environment, while rules specify mappings from conditions to actions (Wilson et al., 2014). However, it can be difficult to determine how to apply such distinctions to behavioral tasks to make predictions. For example, in visual discrimination learning, lesions of the lateral prefrontal cortex impair switching the attentional set, while lesions of OFC impair switching of reward contingencies (Wallis, Dias, et al., 2001). It is unclear why an attentional set would be more clearly related to actions than reward.

More recently, we have proposed that OFC may have a more restricted role than representing task states (Knudsen & Wallis, 2022). We argue that behavioral tasks can be represented as state-transition graphs, which define how one state relates to another. HPC may play an important role in constructing and representing such graphs, although other prefrontal areas, such as lateral and medial prefrontal cortex, may also play an important role in the process. OFC is responsible for representing the current goal and using the state-transition graph to calculate the value of states in relation to the goal. This allows the subject to choose optimally in any specific state and to flexibly respond when goals change. Lesion studies are consistent with this proposed distinction. For example, Buckley and colleagues trained monkeys to perform a simplified version of the Wisconsin Card Sorting Task, which required subjects to decide whether to retain or change rules, given the consequences of their choices (Buckley et al., 2009). They observed that OFC lesions selectively impaired reward-based updating of rule values (i.e., the extent to which a specific task state was associated with reward reward), while lateral PFC lesions (specifically in the principal sulcus) selectively impaired subjects' ability to hold the current rule in working memory.

Hippocampal and orbitofrontal contributions and potential interactions

Previous studies investigating the roles of OFC and HPC in representing task states during reversal learning have largely focused on each region individually. When OFC and HPC are compared, it was often across animal cohorts and limited to rodent models (Wikenheiser & Schoenbaum, 2016; Zhou, Gardner, et al., 2019; Zhou, Montesinos-Cartagena, et al., 2019). Here, we recorded from OFC and HPC simultaneously during reversal learning in primates, enabling us to study how both regions encoded task states in parallel, within the same subject and recording session. We found neural representations of task state in both regions, which remained stable within trials while

strengthening with learning. The trial-by-trial dynamics of state encoding appeared similar across OFC and HPC, and we did not see consistent differences between regions in terms of representation strength or timing. Precisely how and when OFC and HPC interact during learning to use models of task state to guide behavior remains an open question.

Studies in humans (Backus et al., 2016) and nonhuman primates (Brincat & Miller, 2015) have shown that synchrony between prefrontal cortex and HPC, quantified by the coupling strength of oscillations in the theta frequency band, predict learning in value-based and memory-based decision-making tasks. Previous work in our lab causally demonstrated that HPC-driven theta oscillations in OFC are crucial for updating choice preferences in the face of changing reward contingencies (Knudsen & Wallis, 2020), raising the possibility that OFC-HPC interactions are important for creating, updating, or employing representations of task state to guide action selection. One intriguing direction for future research could be examining interactions between OFC and HPC during structure learning to better understand their individual contributions to constructing task state representations.

Interpretational issues

The current study did not aim to compare areas within OFC or HPC, so we did not sample units evenly across each region. Nontrivial functional differences across subregions are well-reported in both OFC and HPC, although the homology of these functional distinctions between rodents and primates is still debated (Wallis, 2012). In humans and nonhuman primates, central OFC (areas 11 and 13) from where we recorded in the current study is distinct from ventromedial PFC (area 14) are distinct. Lesion studies in rats and monkeys (Baxter et al., 2000; Gallagher et al., 1999) and fMRI studies in humans (Gottfried et al., 2003; Valentin et al., 2007) implicate central OFC in value updating when stimulus-response-outcome mappings change. The vmPFC, on the other hand, is more closely tied to encoding expected values and rewards, without playing as much of a role in updating those expectations (Noonan et al., 2011). The neuronal properties of this area remain less understood than those of the central OFC (Bouret & Richmond, 2010; Rich & Wallis, 2014; Smith et al., 2010; Wallis & Rich, 2011), in part because the neurons in this region are typically less well driven by typical laboratory decision-making tasks (Wallis, 2012).

With respect to HPC, prominent theories of HPC organization have argued for a dorsal-ventral (in rodents) or posterior-anterior (in primates) gradient of organization, with the dorsal/posterior HPC involved in cognitive processes, such as representing spatial information for navigation and memory, whereas the ventral/anterior HPC is more important for processing affective information (Fanselow & Dong, 2010; Moser & Moser,

1998; Strange et al., 2014). For example, there is a gradient within hippocampus with respect to the resolution of place cells, with the dorsal/posterior HPC encoding finer-grained spatial representations than the ventral/anterior HPC (Kjelstrup et al., 2008). In contrast, neurons in ventral/anterior HPC are more likely to differentiate locations with respect to reward contingencies (Royer et al., 2010). There is also a gradient in HPC with respect to its connections to the frontolimbic cortex, including OFC, with an increased density of connections towards the ventral/anterior HPC (Barbas & Blatt, 1995). However, more recent studies have shown that reward encoding is evident throughout HPC (Yun et al., 2023). Future studies specifically comparing subregions of OFC and HPC during state inference could more precisely locate task state representations in the brain and guide further research on how state representations are created, communicated, and updated.

An additional challenge is determining the behavioral strategy that the animals use to solve the task. Although reversal learning can be modeled as a two-state inference task (Bartolo & Averbeck, 2020; Wilson et al., 2014), there is considerable debate regarding how subjects actually learn the task (Izquierdo et al., 2017). Furthermore, even when a fully model-based behavioral strategy is theoretically available, animals (including humans) tend to employ a blend of model-based, model-free, and other processes (Collins & Cockburn, 2020; Daw et al., 2011; Gläscher et al., 2010). In our study, we found evidence that subjects used a model of the task to generalize across paired stimuli, where the outcome of choosing one picture informed the likelihood of choosing the paired picture. However, expected value, or the probability that selecting a given picture will lead to reward, is difficult to pull apart from task state in a reversal learning task where 'state' is defined by the current reward contingencies. Future studies in rodents and nonhuman primates could pilot more complex, but still consistently achievable, tasks that more clearly separate task state from the factors that define it.

Conclusion

Modeling tasks as a series of transitions through specific states is a computationally tractable and concise way to describe goal-directed behavior, but its precise implementation in the brain remains unclear. We showed that neural activity in OFC and HPC may be important, not only for representing current states, but also potentially anticipating future states. Understanding these processes could shed light on OFC damage and dysfunction which typically results in deficits in anticipating the consequences of one's actions.

3.4 Methods

Experimental model and subject details

All procedures were carried out as specified in the National Research Council guidelines and approved by the Animal Care and Use Committee and the University of California, Berkeley. Two male rhesus macaques (subject D and V, respectively) aged 8 and 11 years old, weighing 9.8 and 10 kg at the time of recording were used in the current study. Subjects sat head-fixed in a primate chair (Crist Instrument, Hagerstown, MD) while eye movements were tracked with an infrared eye-tracking system (SR Research, Ottawa, Ontario, CN). Stimulus presentation and behavioral conditions were controlled using the MonkeyLogic toolbox (Hwang et al., 2019). Subjects had unilateral recording chambers implanted, centered over the posterior frontal lobe and parietal lobe.

Task Design

Subjects performed a probabilistic reversal learning task where they were required to choose between two differently valued pictures (free choice, 85% of trials) or a single picture (forced choice, 15% of trials). Forced choice trials were included to prevent side bias and ensure that subjects experienced outcomes associated with all picture options. All pictures were isoluminant, presented against a dark gray background. To initiate the trial subjects had to maintain fixation on a central cue for 500 ms. After a 400 ms delay, pictures (2° x 2°) appeared 8° on either side of the fixation cue. Subjects used saccades to choose the left or right picture, indicating their choice by maintaining fixation for 500 ms. After making a choice, the unchosen picture disappeared, and an apple juice reward was delivered probabilistically after a 400 ms delay. Trials were separated by a 1500 ms intertrial interval.

There were two sets of two pictures each, where one set was associated with a 40% probability of reward, and the other set was associated with a 90% probability of reward. Reversals were triggered once the animal reached a learning criterion where they chose the more valuable option on 70% or more of 30 trials and they had to maintain this performance level for 10 trials. Once this criterion was met, subjects entered a phase where an uncued pseudo-probabilistic reversal could occur within 10 trials so long as accuracy remained above 65% throughout this phase. These post-criterion trials enabled the subject to exploit what they had learned, ensuring they did not become frustrated with constant reversals. Subject D completed 8638 trials (102 reversals) over 6 sessions, and subject V completed 14484 trials (120 reversals) over 9 sessions.

Neurophysiological recordings

Subjects were fitted with head positioners and imaged in a 3T MRI scanner. From the MR images, we constructed 3D models of each subjects' skull and target brain areas (Paxinos et al., 2000). Subjects were implanted with custom radiolucent recording chambers fabricated from polyether ether ketone (PEEK). During each recording session, up to 3 multisite linear probes (32-channel V probes with 100μm contact spacing, Plexon, Dallas, TX) were lowered into OFC (areas 11, 12, and 13, as determined from gray-white matter transitions) and HPC (CA1 and CA2/3) simultaneously (up to 6 electrodes total). Unique electrode trajectories were defined for each session in custom software, and the appropriate microdrives were 3D printed (Form 2 and 3, Formlabs, Cambridge, MA) (Knudsen et al., 2019). Lowering depths were derived from the MR images and verified from neurophysiological signals via gray/white matter transitions. Neural signals were digitized using a Plexon OmniPlex system, with continuous spike-filtered signals (200 Hz - 6 kHz) acquired at 40 kHz and local field-filtered signals acquired at 1 kHz.

We recorded neuronal activity over the course of 6 sessions for subject D and 9 sessions for subject V (Table 3.1). We restricted our analysis to neurons with a mean firing rate across the session > 1 Hz. To ensure adequate isolation of individual units, we excluded neurons where >0.2% of spikes were separated by <1100 ms. We transformed single neuron activity into a binary time series at 1 ms resolution, where 1 indicated the presence of a spike and 0 the absence.

Single neuron regression analysis

For each neuron, in overlapping 100 ms windows shifted by 25 ms, we performed the following linear regression on firing rates (FR) aligned to picture onset:

$$FR = \beta_0 + \beta_1 * state + \beta_2 * chosen\ side + \beta_3 * chosen\ value + \beta_4 * chosen\ picture + \beta_5 * trial\ type + \beta_6 * trial\ number \quad \text{Eq. 3.1}$$

We added a parameter for reward when performing linear regression on firing rates aligned to outcome:

$$FR = \beta_0 + \beta_1 * reward + \beta_2 * state + \beta_3 * chosen\ side + \beta_4 * chosen\ value + \beta_5 * chosen\ picture + \beta_6 * trial\ type + \beta_7 * trial\ number \quad \text{Eq. 3.2}$$

with binary variables for *reward* (+1 reward, -1 no reward), *state* (+1 state 1, -1 state 2), *chosen side* (+1 left, -1 right), *chosen value* (+1 high, -1 low), and *trial type* (+1 free, -1 forced). *Chosen picture* was dummy coded (+1 for chosen picture, -1 for all three unchosen options). We included *trial number* as a noise parameter to absorb potential variance due

to neuronal drift over the recording session. Significance was defined as maintaining $p < 0.001$ for 100 ms (four consecutive time bins). We validated this threshold by finding the proportion of units labeled as significantly predicted by chosen side during the 750 ms before and 200 ms after fixation, before pictures appeared on the screen. Only 20 of 1,492 units, or 1%, met these criteria during this window, so we proceeded to use this cutoff to detect units significantly predicted by all factors in our regression model.

Task state decoding with single trial resolution

For each session, we used an LDA decoding algorithm to derive trial-by-trial estimates of task state from population-level activity. We trained the decoder on trials where behavioral performance was high, since these would be more likely to correspond with neural representations of the current task state. Thus, we restricted training trials to those where running behavioral accuracy exceeded 60% across 30 trials. To account for nonstationary firing rates caused by neuronal drift, we split trials into overlapping bins for training, where each bin contained approximately 25% of all available trials. Bins were then stepped by 50 trials and the decoder training repeated. Within each bin, training trials were balanced to include an equal number corresponding to each state. This involved downsampling trials to match the state with the fewer number of trials. For each bin, decoders were trained on sliding windows of 200 ms of neural activity, stepped by 50 ms. In summary, training data for each bin consisted of an [n training trials x n timestamps x n units] matrix of firing rates. To reduce the dimensionality of the input features in each window, we then performed principal components analysis across trials and restricted LDA inputs to the top principal components that explained 95% of variance.

To assess the performance of our decoder, we used a leave-one-out (LOO) cross-validation procedure for each bin. Iterating over trials within each time bin, we held out one trial, trained the decoder on the remaining trials, then tested on the held-out trial. Because trials were randomly removed in the downsampling procedure, to ensure that as many trials from each session contributed to the analysis as possible, we repeated the entire procedure 25 times and averaged posteriors across all instances where a trial was tested. We defined decoder confidence as the difference between the posterior probabilities of decoding the current state relative to the previous state, averaged across decoder iterations (N):

$$confidence = \Sigma\ (p\ (\text{decode}\ S_{current}) - p\ (\text{decode}\ S_{previous}))\ /\ N \qquad \text{Eq. 3.3}$$

Very occasionally (D: 1 of 133 windows, V: 6 of 196 windows), there was an extended sequence of trials where only one state label was present—since these periods tended to

indicate stretches of low task motivation and poor performance (thereby preventing a reversal from occurring), we excluded these trials from our training sets.

Reward-value class decoding with single trial resolution

The pipeline for decoding the interaction between value and reward outcome was the same as described for state decoders above. Trials were assigned to one of four classes (labeled 1-4): (1) high-value choice + reward, (2) high-value choice + no reward, (3) low-value choice + reward, (4) low-value choice + no reward. During training, trials were balanced to include an equal number from each class.

Statistics

All statistical tests are described in the main text or the corresponding figure legends. Error bars and shading indicate standard error of the mean (s.e.m.) unless otherwise specified. All terms in regression models were normalized and had maximum variance inflation factors of 1.4. The coefficient of partial determination (CPD) was used to quantify the percentage of overall variance uniquely explained by each term. All comparisons were two-tailed.

Resource availability

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Joni Wallis (wallis@berkeley.edu).

3.5 Acknowledgements

We thank Thomas Elston, Eric Hu, and Nathan Munet for useful feedback on the manuscript. This work was funded by NIMH R01-MH117763 and NIMH R01-MH121448 to JDW. CFF and JDW designed the experiments. CFF wrote the manuscript. JDW edited the manuscript. CFF and ZZB collected the data. CFF analyzed the data. JDW supervised the project.

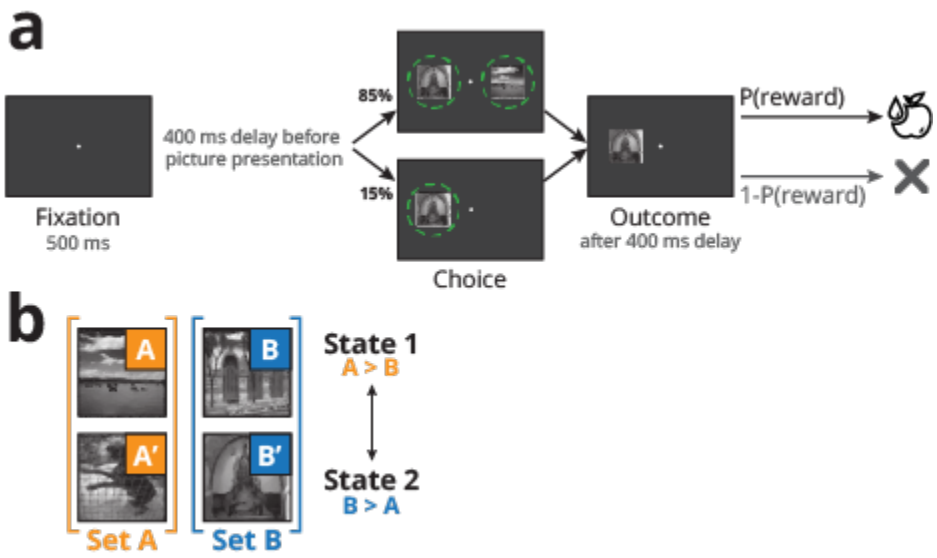

Figure 3.1: Task design. (a) Subjects fixated on a central cue to initiate each trial. After acquiring fixation, the central cue disappeared. After a 400 ms delay, they were presented with one or two pictures (forced and free choice trials, respectively). To select a picture and receive a probabilistic juice reward, monkeys made a saccade and fixated the picture for 500 ms. The outcome of the choice (juice or no juice) was revealed after 400 ms. (b) Subjects learned the values of two sets of pictures, where pictures within each set had the same value. After each reversal, the values corresponding to each set switched. Value was defined here as reward probability (40% vs. 90%).

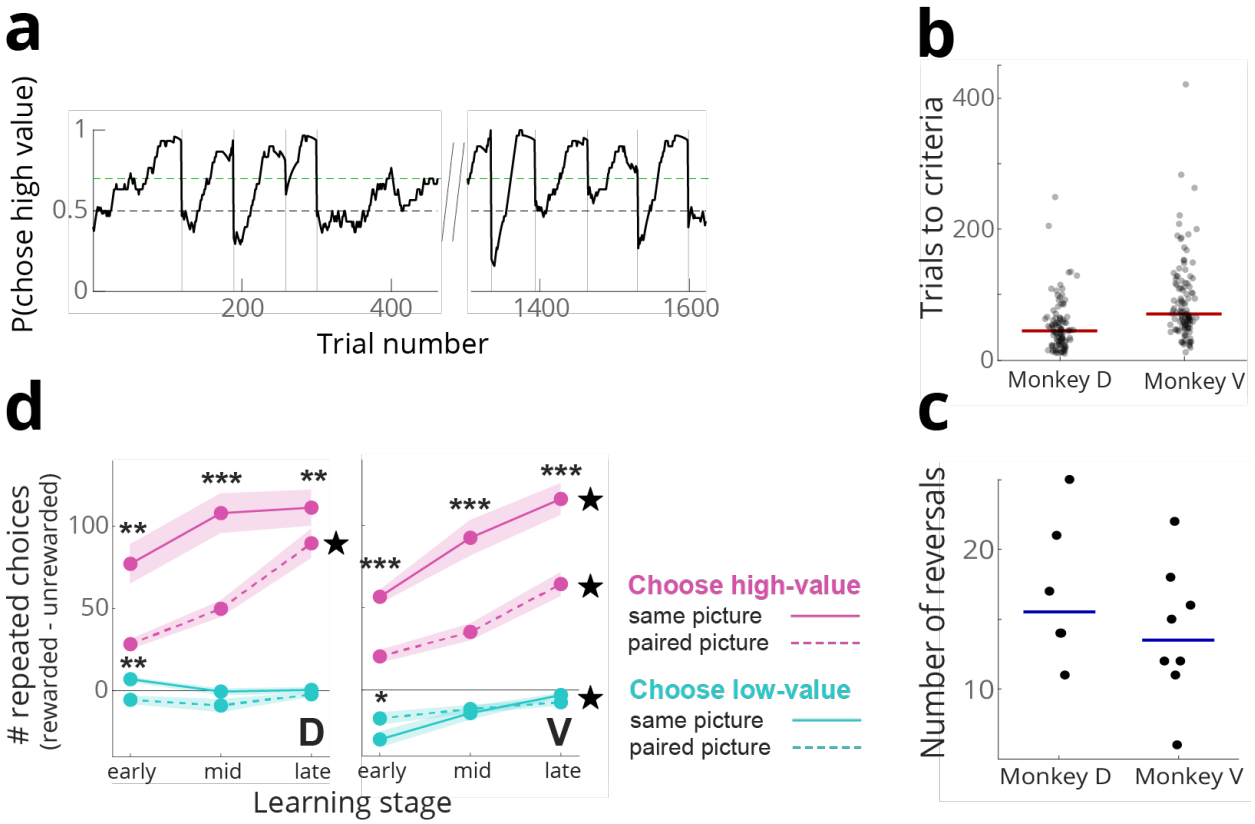

Figure 3.2: Behavior. (a) Probability of choosing the highest-value option across an example session (subject V). Vertical gray lines mark reversal points. The dashed black line is chance (50%), and the dashed green line is the learning criteria accuracy threshold (maintaining at least 70% running accuracy for 10 consecutive trials). (b) Median number of trials to reach learning criteria for each reversal block (red lines), where dots represent individual blocks. (c) Median number of reversals for each session (blue lines), where dots represent individual sessions. (d) Difference in the number of trials where a choice was repeated following reward, relative to trials where no reward was delivered. Trials where the same picture was available on trial t+1 are represented by solid lines, and trials where the paired picture was available on trial t+1 are dashed lines. Trials were further split by whether the choice on trial t was high value (pink) or low value (blue). Blocks (trials between reversals) were split into three learning stages: early, middle, and late (x axis). Monkey D is on the left, and monkey V is on the right.

Subject	Session	OFC neurons	HPC neurons
D	210201	88	71
	210212	14	54
	210215	25	34
	210218	28	55
	210222	29	83
	210225	40	45
V	211018	65	35
	211111	31	125
	211113	61	58
	211115	83	84
	211122	20	89
	211129	46	39
	211204	22	46

Table 3.1: Neural data. Number of neurons recorded per session in OFC and HPC for subjects D and V.

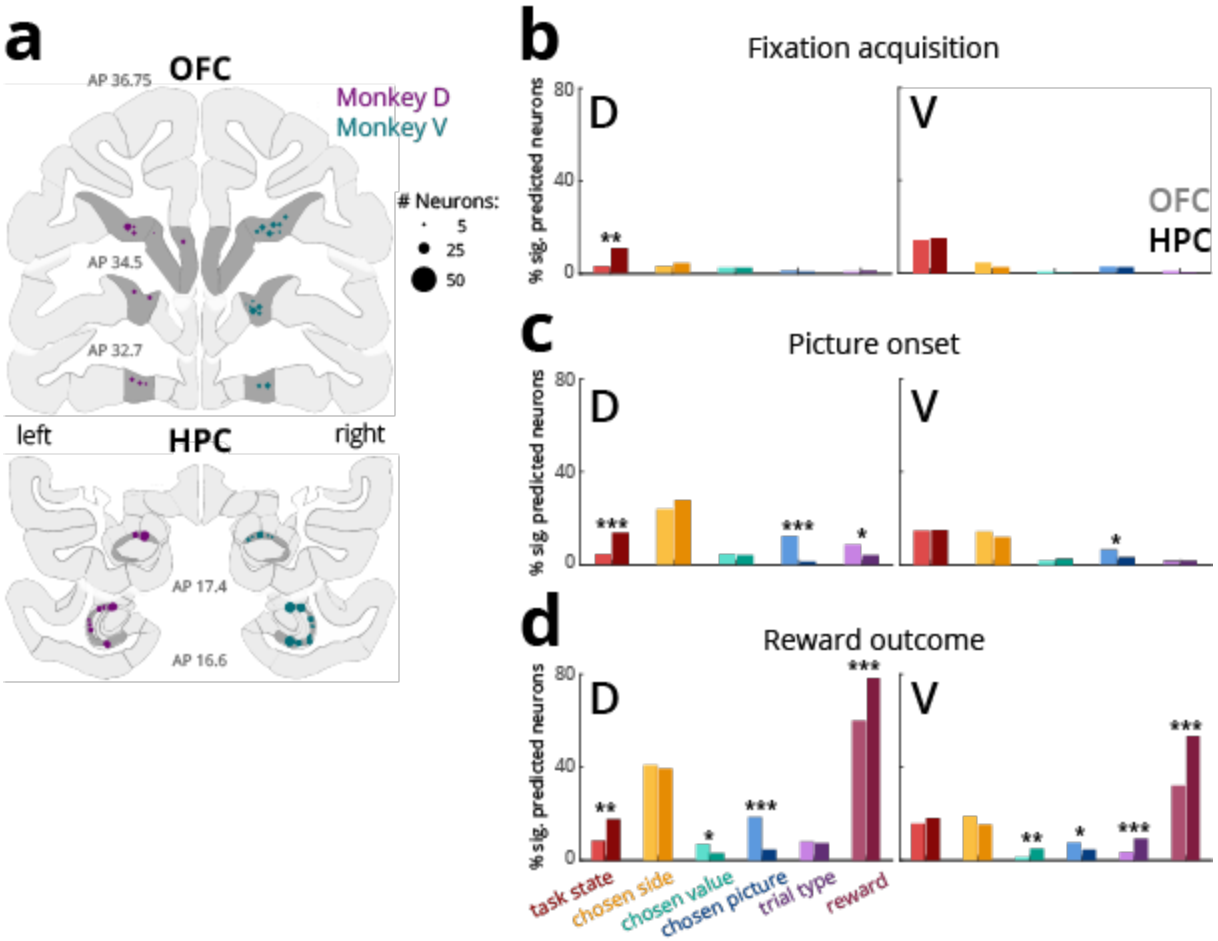

Figure 3.3: Task parameters encoded in OFC and HPC. (a) Reconstruction of OFC and HPC recording sites on coronal slices. The size of each circle represents the approximate number of neurons recorded at that location. AP (anterior-posterior) locations are in millimeters, relative to the interaural line. (b) Percentage of neurons encoding different task parameters during different task epochs. Selectivity was defined as a significant beta weight ($p < 0.01$) for at least four consecutive 100 ms time bins. Lighter bars represent OFC neurons and darker bars represent HPC neurons. Asterisks indicate that there was a significant difference in the prevalence of selective neurons between the areas (chi-squared test, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$).

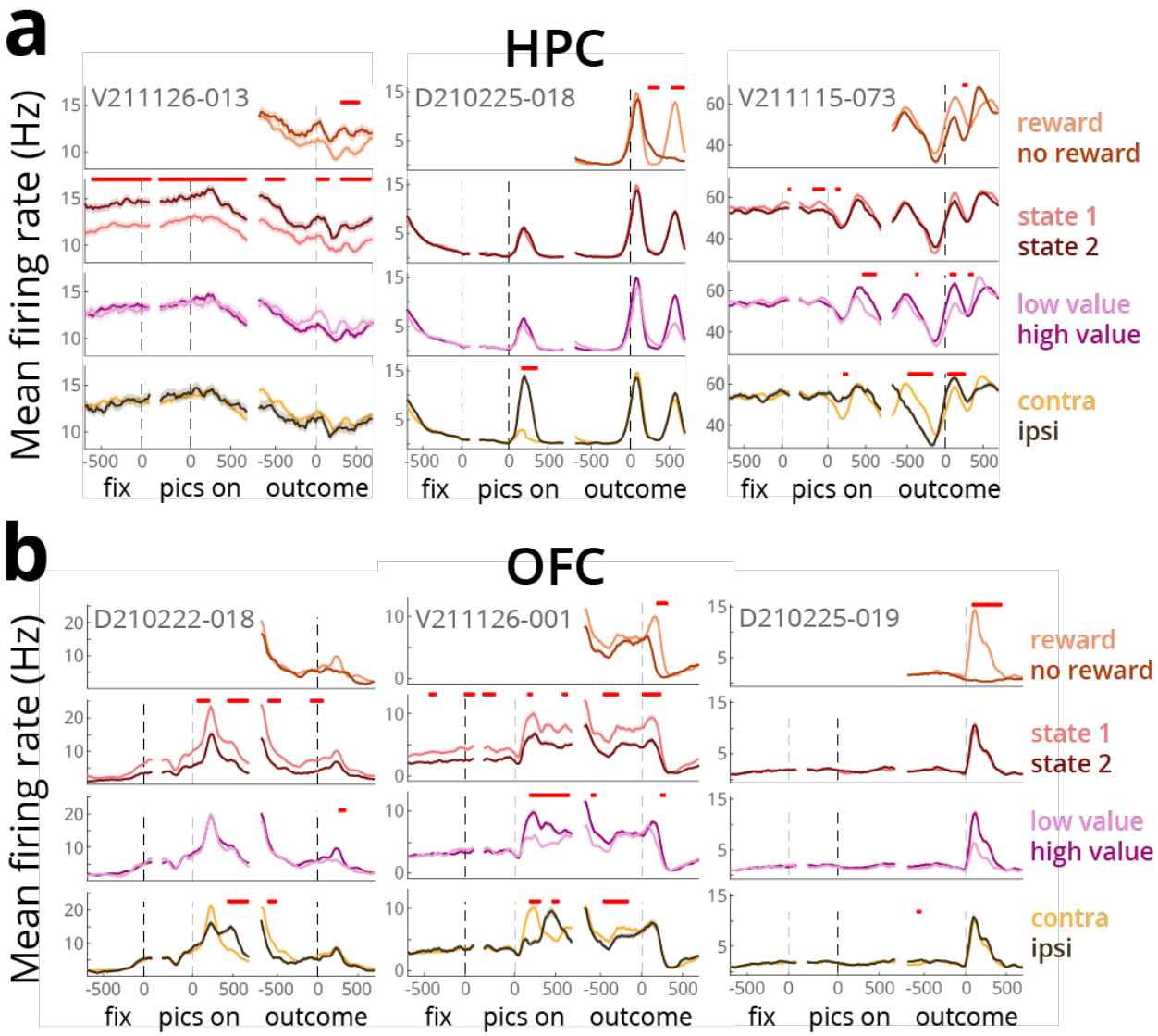

Figure 3.4: Single neuron examples. Spike density histograms recorded from single neurons in OFC (a-c) and HPC (d-f), including significant encoding of (from top to bottom) reward outcome, task state, chosen value, and chosen side, across trial epochs (fixation, picture onset, and reward outcome). Lines depict firing rates (mean ± SEM) averaged over conditions within each factor. Red bars at the top of each panel indicate periods of significance determined by the regression.

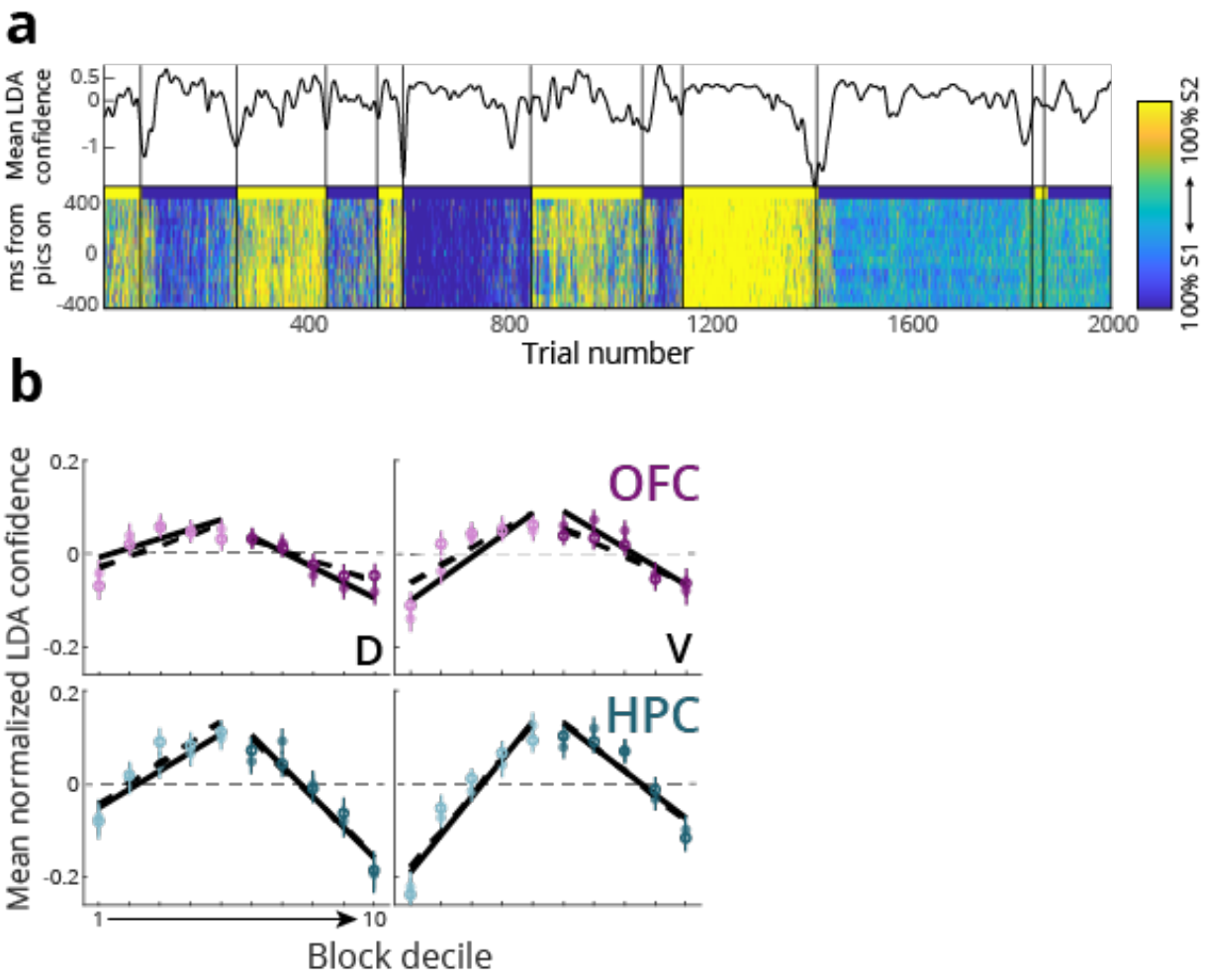

Figure 3.5: Decoding of state across learning. (a) Top: Normalized (z-scored) mean posterior probability of decoding the current state over the previous state, averaged across timestamps. Bottom: Posterior probability of decoding state 1 over state 2, where 100% state 1 is gold and 100% state 2 is blue. Solid bars above the heatmap show the true state label for each block. Here, we show all trials from one example session (subject V) where decoders were trained on HPC population activity synced to picture onset. (b) To equate learning across blocks with different numbers of trials, we split each block into bins containing 10% of the trials (trial decile). We then plotted state decoding as a function of trial decile and fit generalized linear regression models to the first five deciles and second five deciles. Data points show the mean (± s.e.m.) confidence for each decile, averaged across blocks (pooled across sessions). Black lines illustrate best-fit lines. Solid lines and filled points represent the picture onset epoch, and dashed lines and open circles represent the outcome epoch. Fits were significant in both subjects for both brain areas and task epochs (linear regression, $p < 0.01$).

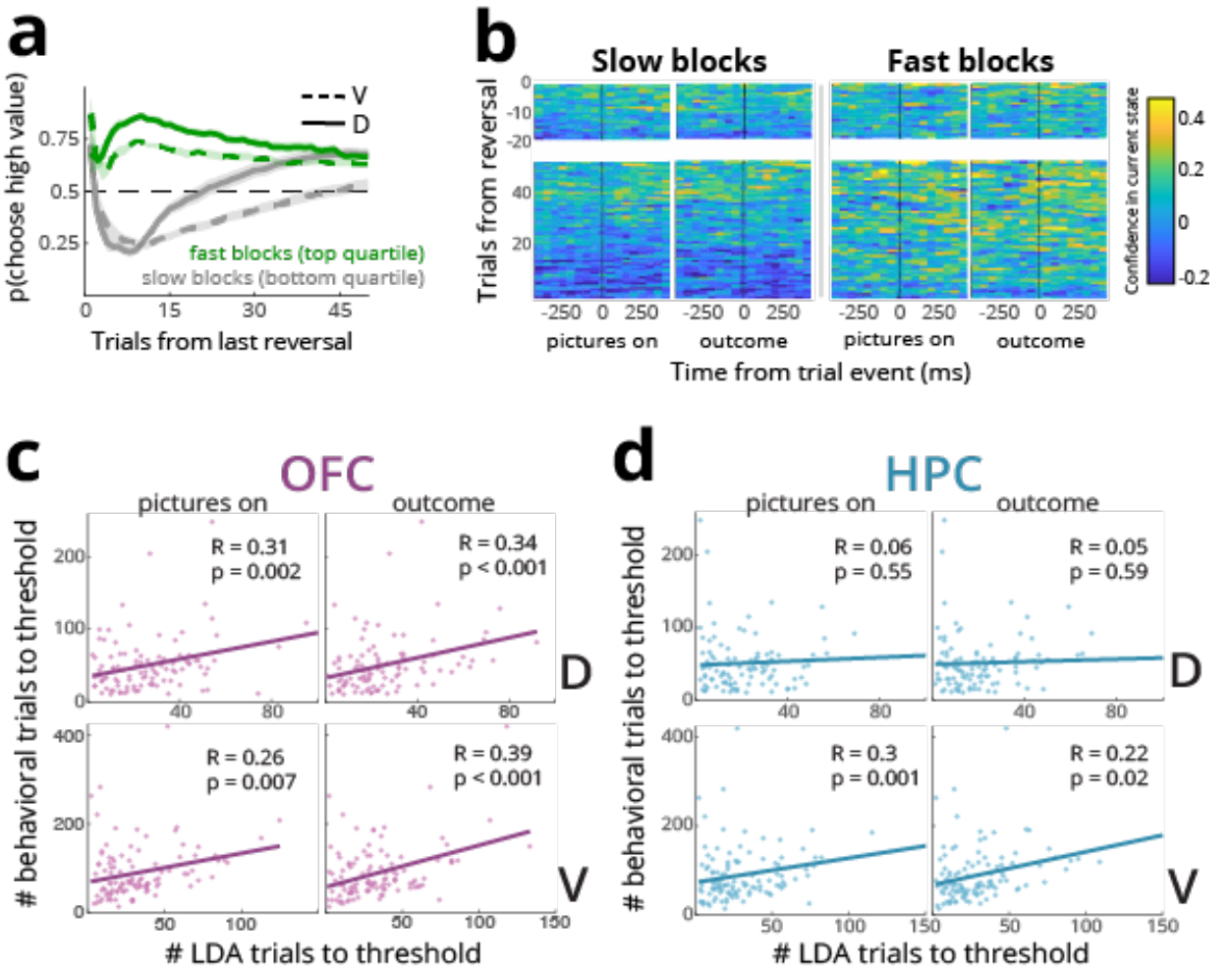

Figure 3.6: State decoding correlates with learning dynamics. (a) Mean probability of choosing the high-value option in the first 50 trials following reversal. Fast blocks (75th percentile choice accuracy in the first 10 trials of the block) are depicted by green lines, and slow blocks (25th percentile) are gray. The horizontal dotted black line shows chance performance (50%). (b) Mean decoder confidence for slow blocks (left) and fast blocks (right) across trials (rows) and time within each trial (columns), with picture onset and outcome marked by black vertical lines. Dark blue indicates greater confidence in the old, now-incorrect state than the current state, while yellow indicates confidence in the new, correct state. These heat maps illustrate results from decoders trained on HPC activity in monkey V. Similar patterns were observed in HPC in monkey D but were less prominent in OFC for both subjects. (c-d) Correlation between learning and (c) OFC and (d) HPC decoder performance. Behavioral performance was measured as the number of trials to reach the learning criterion (maintaining 70% accuracy over 30 trials for 10 consecutive trials). Decoder performance was measured as the number of trials required to exceed 70% of the peak posterior probability of decoding for the correct task state in the current block. Datapoints show individual blocks combined across sessions for each subject.

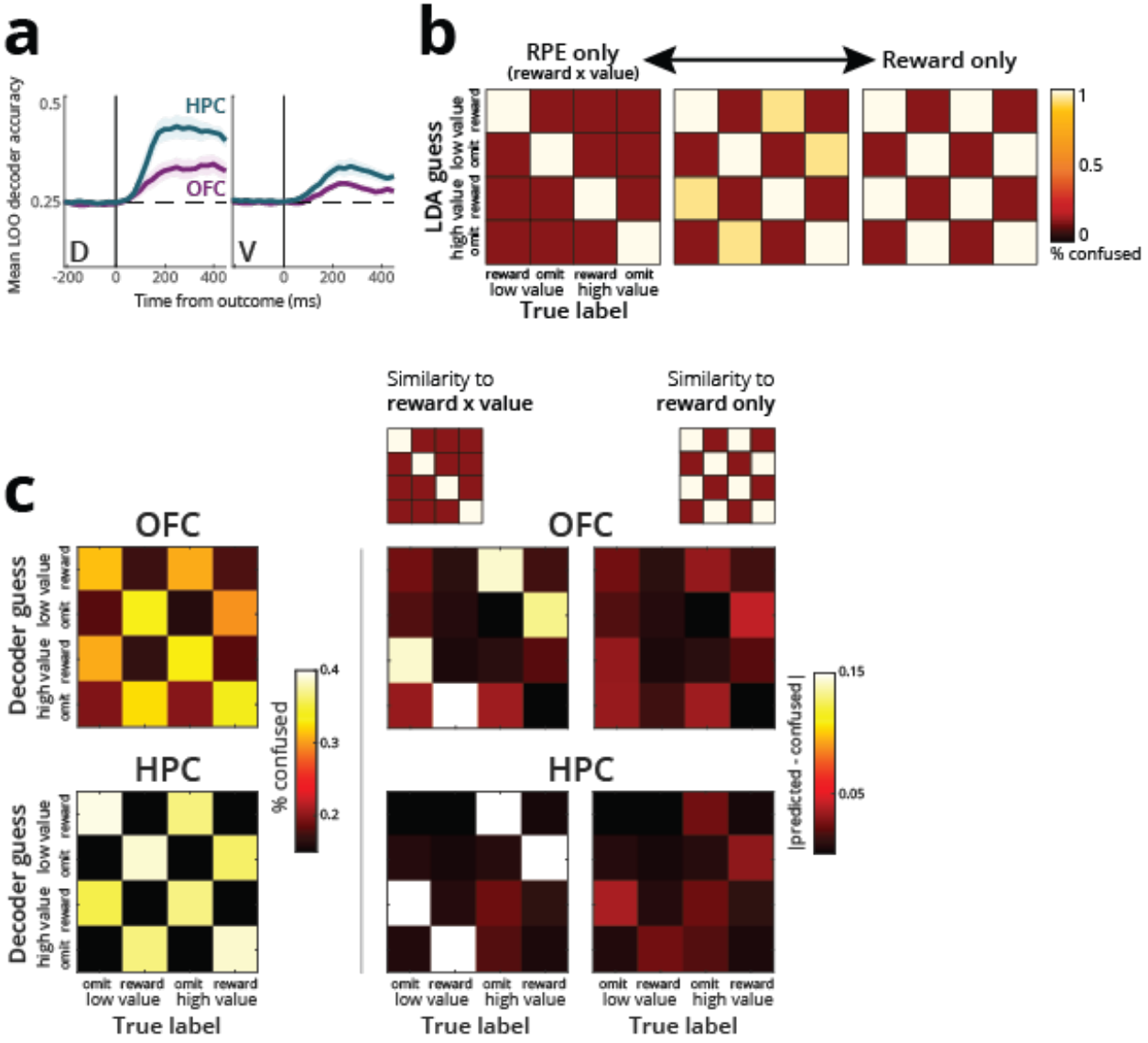

Figure 3.7: Decoding expectancies and outcomes. (a) Population decoding of chosen value and reward in OFC and HPC aligned to reward outcome (vertical black lines) for each subject (left =. Each line represents mean LOO accuracy across sessions (± SEM), where chance was 25% (four classes: rewarded high value, rewarded low value, unrewarded high value, unrewarded low value). (b) Confusion matrices for decoders trained on OFC (top) and HPC (bottom) population activity (left column). Results are averaged across subjects and sessions 100-500 ms after reward outcome. (c) The percentage of confused trials, where reward outcome was correctly labeled but expected value was misclassified, significantly decreased over the course of learning within the block (one-way ANOVA, $p < 0.001$ for both regions and subjects). Each bar reflects the mean (± s.e.m.) across all blocks, combined across recording sessions.

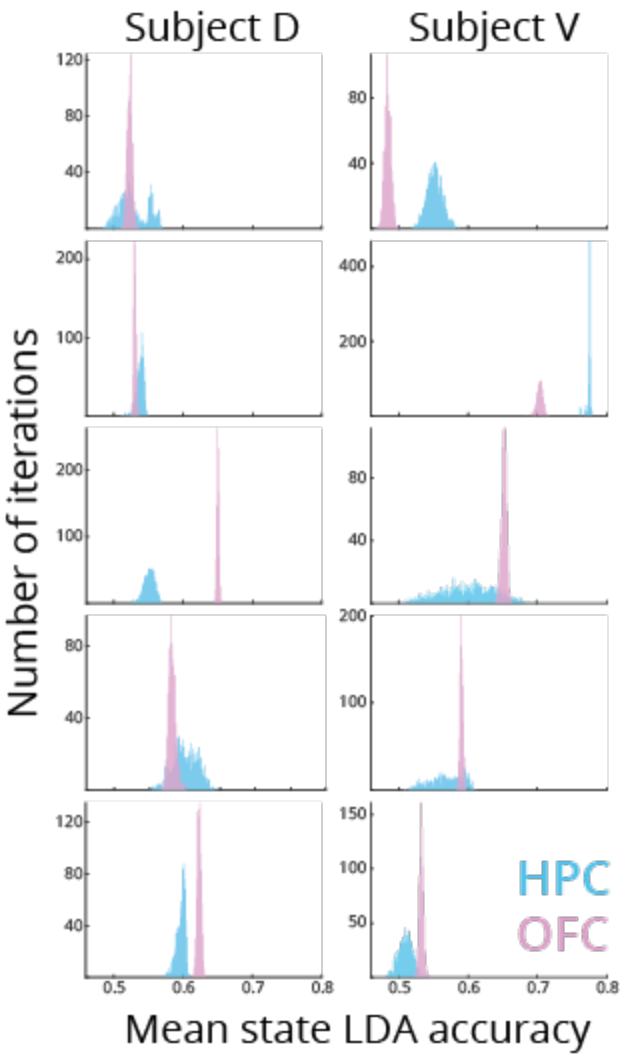

Figure 3.8: Unit-matched state decoder accuracy. For each session (subplots) where the number of recorded OFC neurons was smaller than the number of recorded HPC neurons, we randomly dropped HPC neurons to match the OFC count, then trained our state LDA decoder as described in the methods. We repeated this 1000 times, then ran a paired-samples t-test for each session. In general, decoders were more accurate when trained on data from one region over the other (all sessions for both subjects, $p < 0.001$), but the more-accurate region was inconsistent across sessions (paired-samples t-test for means across iterations, $p > 0.1$).

Chapter 4

Theta oscillations do not explain dynamic activation of value states in the orbitofrontal cortex

Ford, Celia F., Balewski, Zuzanna Z., & Wallis, Joni D.

Subjective decision-making involves the dynamic activation of alternating representations of option values in the OFC, but the mechanism determining the timing of these representational switches remains unknown. To address this, we recorded from the OFC while monkeys performed a value-based decision-making task. Using population-level decoding, we replicated previous work and found sequential representations of the values of each available option during deliberation, with idiosyncratic dynamics that varied from trial to trial. We related the timing of these decoded states to theta oscillations in the OFC and found that value state representations switched independent from theta cycles, suggesting that a mechanism other than theta oscillations shapes value-guided decision-making dynamics.

4.1 Introduction

When deciding between two options (chocolate ice cream, or vanilla? Red wine, or white?), neural (Padoa-Schioppa, 2011) and behavioral (Krajbich et al., 2010) evidence suggests that we compute the value of each available option, given context and experience, then deliberate between them. The orbitofrontal cortex (OFC) has been implicated in value-based decision-making for decades (Padoa-Schioppa & Conen, 2017), and is likely important for this deliberation process. When deciding between two options, OFC flickers between representations of each option's value, or *value states* (Balewski et al., 2022; Rich & Wallis, 2016). The dynamics of this value state flickering are highly idiosyncratic, with the timing and number of representation switches varying greatly from trial to trial. Stronger encoding of the ultimately chosen option value appears to predict faster, more decisive choice behavior, but the underlying mechanism driving and shaping the flickering dynamics remains unknown.

This pattern of neural encoding during decision-making is not unique to the OFC. There is evidence that neurons in the hippocampus (HPC) flicker between representations of past and present places during a foraging task where spatial cues defining the reward

location are switched back and forth between two configurations (Jezek et al., 2011). In a self-paced task where rats had to choose between two arms of a T-maze, HPC neurons alternated between representing two different hypothetical future scenarios (Kay et al., 2020). In both studies, HPC representations switched roughly every 125 ms, aligning with theta oscillations (4-8 Hz) in the local field potential (LFP).

HPC and OFC are anatomically (Barbas & Blatt, 1995; Carmichael & Price, 1995) and functionally (Young & Shapiro, 2011) connected, and neurons in both regions phase-lock to local theta LFPs (Buzsáki, 2002; van Wingerden et al., 2010). LFP coherence between HPC and the prefrontal cortex, especially in the theta frequency band, is positively correlated with free exploration (Siapas et al., 2005), decision-making (Tang et al., 2021), error detection during associative learning (Brincat & Miller, 2015), and context-driven memory retrieval (Benchenane et al., 2010; Morici et al., 2022; Place et al., 2016). Recent work in our lab demonstrated that value-encoding neurons in OFC phase-lock with theta oscillations, and that HPC-driven theta oscillations in OFC are critical for learning during value-based decision-making (Knudsen & Wallis, 2020). Similar results have been observed in rats performing a spatial reversal learning task, where OFC spikes phase locked to HPC theta oscillations (Riceberg et al., 2022). In both cases, OFC activity was predicted by HPC activity, suggesting that HPC may play a role in shaping OFC encoding dynamics.

One intriguing possibility is that value state representations in OFC may switch back and forth between available options in time with HPC-driven theta oscillations. While there is evidence suggesting separately that HPC-driven theta oscillations in OFC are important for learning (Knudsen & Wallis, 2020; van Wingerden et al., 2010), and that HPC representations may be synced to theta cycles, no studies to our knowledge have directly explored whether theta cycles also shape encoding dynamics in OFC. Here, we analyzed OFC activity previously recorded in our lab while monkeys performed a simple value-based decision-making task (Balewski et al., 2022). We trained decoders to identify OFC value states during deliberation periods, then asked whether the frequency, onset, or duration of those states could be explained by their position relative to OFC theta cycles. We found no consistent relationship between value state dynamics in OFC and theta oscillations, suggesting that HPC oscillations do not explain the idiosyncratic flickering we see in OFC representations during decision-making.

4.2 Results

Two macaques (monkeys C and G) learned to use a bidirectional lever to select between one (forced choice trials) or two (free choice trials) available pictures on each trial (Fig. 4.1a), as described previously (Balewski et al., 2022). They learned the values of 16

pictures, each associated with a unique juice reward defined by juice quantity and probability of delivery (Fig. 4.1b). Pictures were grouped into four value bins (where 1 is the lowest value and 4 is the highest) based on the subjective value for each subject, defined as expected value. Pictures were assigned to the same value bin for both subjects. As expected, subjects preferred larger juice amounts and more certain rewards, and responded faster as the maximum available value increased (for detailed methods and results, see Balewski et al., 2022).

We recorded single neuron firing activity and LFP from OFC during task performance, using up to four acute multisite electrodes per session (for detailed methods, see Balewski et al., 2022). After filtering noise, LFPs were bandpassed in the θ frequency band (4-8 Hz). We obtained analytic amplitudes from the Hilbert transform, then extracted phase information from the resulting signals using the angle of the Hilbert transform. We observed a consistent theta phase reset within 100 ms of picture onset across trials (Fig. 4.2). Increases in cross-trial OFC theta phase alignment in response to external task events, like pictures appearing on screen or the delivery of a juice reward, have been reported previously (Knudsen & Wallis, 2020). However, it is unknown whether theta oscillations in OFC drive behavior by facilitating internal representations of task variables.

Value states in OFC oscillate between the chosen and unchosen pictures values during deliberation, with idiosyncratic dynamics that vary with behavioral choice from trial to trial (Balewski et al., 2022; Rich & Wallis, 2016). While the timing and duration of these state representations do not appear to be predicted by any specific external stimulus (Rich & Wallis, 2016), it is possible that the OFC state "flip-flops" observed during deliberation are not random, but instead driven by neural oscillations. Such representation switches have been seen, synchronized to theta oscillations, in the hippocampus in rodents (Jezek et al., 2011; Kay et al., 2020). Given that HPC-driven theta oscillations in OFC are critical for reward-based learning (Knudsen & Wallis, 2020), we hypothesized that value state representations may be linked to theta cycles.

First, we asked whether value states were disproportionately encoded during a given phase of the theta cycle, or if their timing was more uniformly distributed. For each free-choice trial, we isolated theta cycles falling between picture onset and choice, then split each theta cycle into 20-degree phase bins and counted the number of chosen value and unchosen value states that were encoded while the theta phase angle fell in each bin. There was a significant effect of phase angle on state decodability in one subject, whether the decoded state was chosen or unchosen (one-way ANOVA, subject G, chosen: $F(17,144) = 25.24$, $p < 0.001$; unchosen: $F(17,144) = 21.97$, $p < 0.001$), where most states appear to be decoded during peaks and troughs of the theta cycle (Fig. 4.3). We saw the same qualitative trend in the remaining subject, but the overall number of decoded states was

too low in the remaining sessions to see a large effect. We also observed a statistically significant difference in the distributions of chosen and unchosen value states across the theta cycle (two-sample Kolmogorov-Smirnov test, subject C: $T = 0.035$, $p < 0.001$; subject G: $T = 0.027$, $p < 0.001$), such that chosen states appear to be slightly more likely to appear during the falling phase of the theta cycle than unchosen states.

While LFP activity originates from a complex combination of sources (Herreras, 2016), increased firing rates and increased spike synchrony both contribute to increased spectral power (Buzsáki et al., 2012). So, it is possible that the relationship we observed between theta phase and state decodability may be confounded by spike activity. To assess whether fluctuations in value state decodability are predicted by firing rate, we performed logistic regression on states (represented as a binary vector, where 1 or 0 indicated that a value state (chosen or unchosen) was or was not decoded in each time bin, respectively). We fit binary states with firing rates and theta phase angle, separated into sine and cosine components, as predictors (Fig. 4.4a**,** see Methods, Eq. 4.1). We observed that beta coefficients for firing rates were markedly larger than those for theta phase angle, and significantly predicted state decodability across the entire picture epoch, up to two seconds after picture onset (Fig. 4.4a**,** red lines). Theta phase, on the other hand, only significantly predicted decodability in the roughly 500 ms following picture onset (Fig. 4.4a, blue and purple lines).

To more clearly evaluate the relationship between firing rate and state decodability, we split firing rates, boxcar smoothed with a 20 ms window stepped at 5 ms, into 0.5 Hz bins and counted the number of value states (chosen or unchosen) that fell in each bin (Fig. 4.4b). We observed a statistically significant difference in the distributions of decoded and undecoded states across firing rate bins in one subject (two-sample Kolmogorov-Smirnov test, subject C: n.s.; subject G: $p = 0.02$, $T = 0.35$). Given the large sample size of each population of decoded states, we also found a significant difference in the mean firing rate associated with decoded and undecoded states in both subjects (two-sample t-test, subject C: $p < 0.001$, $T = 56.09$; subject G: $p < 0.001$, $T = -3.4336$), where decoded states appear to be associated with slightly higher firing rates in one subject (C), and slightly lower firing rates in the other (G). Taken together, these results suggest that firing rate predicts the presence or absence of decoded states during deliberation more than theta phase, but the specific relationship between firing rate and decodability is less clear.

While firing rates seem to impact the strength of value state encoding, it is still possible that the dynamics of value representations are guided by theta oscillations. We hypothesized that, if theta oscillations drive OFC to switch between chosen and unchosen value representational states, then the classification accuracy of a decoder trained on OFC activity would improve if we synced training data to the theta reset following picture onset, rather than to picture onset itself. For each trial, we identified the theta phase reset

time, defined here as the first timestamp to exceed 2 radians, at least 125 ms from the next local peak in phase angle. We then shifted all LFP and firing rate samples so that zero was set to the theta phase reset, relative to the picture presentation time. For almost all sessions across both subjects, decoders trained on theta-aligned data had lower classification accuracy than those trained on picture-aligned data (paired samples t-test, subject C: $d.f. = 9$, $T = 2.71$, $p = 0.02$; subject G: $d.f. = 8$, $T = 8.59$, $p < 0.001$) (Fig. 4.5a). Peak accuracy across the trial (500 ms before to 2000 ms after picture onset) for all but one session was highest with picture-aligned data (Fig. 4.5b). These results suggest that, at least for forced-choice trials, value encoding dynamics in OFC were more closely linked to task events than HPC-driven theta oscillations.

Another possibility is that, regardless of whether states are encoded at a consistent phase angle across theta cycles, that the state being represented changes across theta cycles (Fig. 4.6a). To test this, we identified theta cycles occurring in the deliberation period between picture onset and choice for each free-choice trial. For each trial, we counted each value state (if any) encoded in the time spanned by each cycle. If most states falling within that cycle represented the chosen value, we labeled that cycle as "chosen." Likewise, states consisting of mostly unchosen value states were labeled "unchosen." To determine the likelihood of decoding the same or different states across theta cycles by chance, we counted the number of states falling in sliding 125 ms (8 Hz) time bins, stepped at 20 ms, across the deliberation period, effectively jittering the start of the first theta cycle on each trial (Fig. 4.6b). At least one value state, chosen or unchosen, was decoded during deliberation in 3963 (subject C) and 2067 (subject G) free-choice trials, significantly more than expected by chance (chi-squared test, $p < 0.001$ for both subjects) (Fig. 4.6c). Of those trials, the same state was decoded across theta cycles 87.2% (subject C) and 91% (subject G) of the time, rather than switching from state to state, significantly more often than expected by chance (chi-squared test, $p < 0.001$ for both subjects) (Fig. 4.6d). Together, these results suggest that the timing of OFC value state flip-flopping is not guided by theta oscillations, and that these trial-specific dynamics are instead driven by another internal process.

4.3 Discussion

Here, we examined the relationship between the theta-filtered LFP, and value states decoded from the deliberation period of free-choice trials and found evidence against the prediction that the dynamics of OFC value representation are driven by theta oscillations. We observed that, while value states were more likely to be decoded at certain phases of the theta cycle, alternations between states were not synchronized to, or explained by, theta oscillations. When training our decoders, we found that realigning neural data to the theta reset following picture onset, rather than to stimulus presentation itself,

decreased decoder accuracy. Together, our results suggest that, while OFC neurons encode value in phase with theta, the idiosyncratic dynamics of OFC value state fluctuations cannot be explained by the oscillatory dynamics.

These findings may initially appear surprising, given that selectively disrupting theta oscillations in OFC, while leaving single-neuron firing rates intact, severely impairs reward-guided learning (Knudsen & Wallis, 2020). How can we reconcile our results suggesting that OFC state fluctuations—a potential mechanism driving value-based deliberation—are unaffected by theta oscillations, with evidence that OFC theta oscillations are critically important for learning and decision making? One possibility is that the role of theta oscillations is not to guide the dynamics of sequential state representation in OFC, but to increase the likelihood that hippocampal neurons can share information with OFC neurons.

LFP phase coupling helps to coordinate neuronal populations across brain regions, which may be one means by which the OFC exchanges information with the hippocampus during decision making (Canolty et al., 2010). Orbitofrontal populations have heterogeneous response properties (Padoa-Schioppa, 2013; Sadacca et al., 2018; Strait et al., 2016; Zhou, Gardner, et al., 2019), and neuronal response properties do not appear to inform anatomical organization (Rich & Wallis, 2014). By preferentially firing at the same phase of the theta oscillation, hippocampal neurons are better able to coordinate with anatomically dispersed, but functionally similar, OFC neurons (Buzsáki et al., 2013). Recent work from our lab supports this hypothesis, demonstrating that OFC neurons encoding reward predictions preferentially spike at specific phases of the theta oscillation (Knudsen & Wallis, 2020). Similarly, we found that OFC value states were preferentially decoded at specific phases. This provides further evidence that hippocampal theta oscillations are involved in OFC value coding but does not necessarily mean that the timing of OFC state switches are aligned with the timing of theta oscillations.

Periodic theta-paced 'flickering' of competing states has been observed in the hippocampus (Jezek et al., 2011; Kay et al., 2020; Mark et al., 2017). However, despite growing bodies of research highlighting the role of theta oscillations in communication (Karakaş, 2020) and the functional connection between OFC and the hippocampus (Wikenheiser & Schoenbaum, 2016), it is possible that OFC value dynamics differ enough from hippocampal place dynamics to be driven by a different mechanism altogether. First, the temporal characteristics of OFC state representations and theta cycles don't quite align. The average duration of a value state in OFC is roughly 80 milliseconds, while theta in primates is defined as the frequency band spanning 4-8 Hz, or 125-250 milliseconds. Additionally, the periodicity of neural oscillations contrasts with OFC value dynamics—the duration of individual value state representations, and timing with which states switch, varies idiosyncratically from trial to trial (Balewski et al., 2022; Rich & Wallis,

2016). This raises the question: if not theta oscillations, what causes OFC state alternations?

One possibility is that OFC fluctuations are driven by shifts in attention during deliberation. Several accounts argue that the decision-making process involves sequentially allocating overt attention, via eye movements, to different choice options and accumulating evidence for the ultimately chosen option (Hare et al., 2011; Krajbich et al., 2010). While saccades and fixations do not appear to affect value state representations in the OFC, (Rich & Wallis, 2016), covert attention may still play a role in shaping decision-making dynamics. Several studies have found evidence consistent with the idea that value-encoding brain regions encode the value of an option while it is being attended to, regardless of whether that option is being fixated on (Blanchard et al., 2015; Lim et al., 2011; McGinty et al., 2016; Rich & Wallis, 2016; Rudebeck et al., 2013; Strait et al., 2014; Xie et al., 2018). Our findings suggest that a single population of neurons, rather than multiple competing populations, alternately represent both option values during deliberation (Balewski et al., 2022; Rich & Wallis, 2016). It is possible that attention, covert or overt, enables a single set of neurons to flexibly encode multiple option values leading up to the time of choice (Hayden & Moreno-Bote, 2018). This framework is aligned with recent research demonstrating that attention and reinforcement learning are deeply intertwined, such that attention modulates neural representations of value and vice versa (Hunt et al., 2018; Leong et al., 2017). Our experiment design did not explicitly probe attention, limiting our ability to study the role of attention in shaping value representations beyond course measurements of overt attention like gaze. Future studies can take this step by adding attention-pulling stimuli or visual modifications to value-guided decision-making tasks, independent of reward contingencies, while recording from OFC populations.

Another open question concerns the relationship between OFC and the hippocampus: why do theta oscillations appear to inform representational dynamics in the hippocampus, but not in OFC? Neurons in the rodent hippocampus have been shown to flicker between representations of past and present environments (Jezek et al., 2011; Mark et al., 2017) and between different hypothetical future scenarios (Kay et al., 2020) in pace with theta oscillations. These studies investigated hippocampal task state encoding in the spatial domain, while rats completed reward-driven spatial navigation tasks. However, recent evidence suggests that, in addition to encoding relationships between environmental features, place cells in the primate hippocampus also encode relationships between options in an abstract value space (Knudsen & Wallis, 2021). To our knowledge, theta-paced place cell flickering has not yet been observed in primates, perhaps due to the constraints implicit in designing spatial navigation tasks for larger animals undergoing neurophysiological recording. One intriguing direction for future research could be to record from hippocampal place cells in primates during value-based

decision-making and explore whether value representations flicker at the same theta frequency as has been observed in the spatial domain. Broadly, understanding whether the role of theta oscillations in decision-making varies by task domain, by anatomical region, or by another factor (or combination of factors) will help shed light on the functional role of neural oscillations.

OFC dysfunction is implicated in a wide range of neuropsychiatric disorders, including obsessive-compulsive disorder and addiction (Fernando & Robbins, 2011). By studying the underpinnings of decision making, including sequential representations of option values in OFC, treatments based on closed loop microstimulation (i.e., decoding brain states in real time and selectively disrupting maladaptive patterns of activity) will become increasingly tenable. Here, we presented evidence against theta oscillations as a putative driver of OFC value dynamics, narrowing the search for viable explanations. As Thomas Edison said about his search for incandescent electric light, "I have not failed 700 times. I have succeeded in proving that those 700 ways do not work."

4.4 Methods

Experimental model and subject details

All procedures were carried out as specified in the National Research Council guidelines and approved by the Animal Care and Use Committee and the University of California, Berkeley. Two male rhesus macaques (subject C and G, respectively) aged 6 and 4 years old, weighing 10 and 7 kg at the time of recording were used in the current study. Subjects sat head-fixed in a primate chair (Crist Instrument, Hagerstown, MD) while eye movements were tracked with an infrared eye-tracking system (SR Research, Ottawa, Ontario, CN). Stimulus presentation and behavioral conditions were controlled using the MonkeyLogic toolbox (Hwang et al., 2019), and subjects were trained to make choices by manipulating a bidirectional lever located on the front of the chair. Subjects had unilateral (subject C), or bilateral (subject G) recording chambers implanted, centered over the frontal lobe.

Task Design

Subjects were trained to perform a value-based decision-making task where, on each trial, they chose between pairs of pictures or a single picture ((Balewski et al., 2022)). To initiate picture presentation, subjects fixated continuously on a 0.5° fixation cue at the center of the screen. On free choice trials (67% of trials), two 2.5° x 2.5° pictures appeared on either side of the fixation cue. On forced choice trials (33%), one picture appeared on either side. Subjects moved a bidirectional lever left or right to indicate their choice. They

learned to associate 16 pictures with four juice reward levels (subject C: 0.15, 0.3, 0.45, 0.6 mL; subject G: 0.1, 0.2, 0.3, 0.4 mL) and four reward probability levels (C: 0.15, 0.4, 0.65, 0.9; G: 0.1, 0.37, 0.63, 0.9) (Fig. 4.1). Amounts and probabilities were uniquely titrated for each subject so both dimensions were considered roughly equally during decision-making. The chosen picture stayed on screen during reward delivery. Trials were separated by a 1000 ms intertrial interval. Gaze position was sampled at 500 Hz. Subject C completed 10,358 free choice trials (5,397 forced) over 11 sessions, and subject G completed 10,793 free choice trials (5,873 forced) over 12 sessions).

Neurophysiological recordings

Subjects were fitted with head positioners and imaged in a 3T MRI scanner. From the MR images, we constructed 3D models of each subjects' skull and target brain areas (Paxinos et al., 2000). Subjects were implanted with custom radiolucent recording chambers fabricated from polyether ether ketone (PEEK).

During each recording session, we lowered up to 4 multisite linear probes (16- or 32-channel V probes with 75, 100, or 200 μm contact spacing, Plexon, Dallas, TX) into OFC (areas 11 and 13, AP +29 to +39 mm) in each chamber. We used custom software to define electrode trajectories, then designed and 3D printed appropriate microdrives (Form 2 and 3, Formlabs, Cambridge, MA) (Knudsen et al., 2019). We derived lowering depths from MR images and verified them from neurophysiological signals via gray/white matter transitions. Neural signals were digitized by a Plexon OmniPlex system, with continuous spike-filtered signals (200 Hz - 6 kHz) acquired at 40 kHz and local field-filtered signals acquired at 1 kHz.

Here, we analyzed neuronal activity from 12 sessions recorded in subject C and 9 sessions recorded in subject G (Table 4.1). We manually isolated units using 1400 ms waveforms, thresholded at 4 standard deviations above noise (Offline Sorter, Plexon). We restricted our analysis to neurons with a mean firing rate greater than 1 Hz across the session, and where over 0.2% of spikes were separated by at least 1100 ms. We excluded LFP channels without at least one unit. In total, we recorded 616 and 1874 neurons and 453 and 1401 LFP channels from OFC in subjects C and G, respectively.

We notch filtered raw LFP signals at 60, 120, and 180 Hz, bandpassed using finite impulse response filters in the θ (4-8 Hz) frequency band, then calculated the Hilbert transform to obtain analytic amplitudes.

Value decoding with single trial resolution

We decoded picture value from single neuron and theta filtered LFP activity in OFC, synced to picture onset. For each session, we trained a linear discriminant analysis (LDA) decoder to predict the value (1-4, where 1 was the lowest value bin and 4 was the highest) from neural activity on forced-choice trials, where a single picture appeared on screen. To reduce the dimensionality of the input features in each window, we performed PCA across trials separately for normalized neuron firing rates and normalized magnitudes from the θ LFP band, then restricted inputs to the top PCs that explained 95% of variance.

To evaluate classification accuracy, we trained a decoder on neural activity between 100-500 ms after picture onset, for forced-choice trials balanced for picture identity and location. We randomly split trials into training sets and held-out sets, where the training sets were balanced to include an equal number of each picture identity and location. We used a k-fold validation procedure where we trained the decoder on all but one set and tested on the held-out set. We repeated this procedure with 50 random trial samples, averaged classification accuracy across all instances, then applied these decoder weights to each free-choice trial in sliding windows of 20 ms stepped by 5 ms. We computed the posterior probability of decoding each value bin for every free-choice trial, for each of 50 random forced-choice trial samples and averaged posterior probabilities across instances.

We defined a unique baseline posterior probability for each value level by averaging posterior probabilities corresponding to that level across trials where pictures of that value did not appear on screen. We defined *decoding strength* as percent change in posterior probability from baseline, for each value level. For each trial, we relabeled decoder classes as *chosen value*, *unchosen value,* and *unavailable*. We defined *states* as periods where decoding strength for a given class exceeded 200% (i.e., doubled baseline) for at least four consecutive time bins.

Value state regression analysis

For each overlapping 100 ms time bin, shifted by 25 ms, we performed the following logistic regression on value states decoded from neural activity during free-choice trials:

$$state = \beta_0 + \beta_1 * firing\ rate + \beta_2 * sin(\theta)\ \beta_3 * cos(\theta) \quad \text{Eq. 4.1}$$

where state was binary (0 = no value state decoded, 1 = decoded either chosen or unchosen value). Significance was defined as $p < 0.01$ for any given time bin.

Identifying theta cycle boundaries

To identify the boundaries of theta cycles, we used the MATLAB function 'findpeaks' to identify theta filtered LFP phase angle samples with a minimum peak of 2 radians and a

minimum distance from other putative peaks of 125 ms. Since we were interested in the dynamics of value state representation during the deliberation period between picture onset and choice, we only included theta cycles falling between those two trial events in our analyses. To determine whether the consistency of state decoding across theta cycles was significantly different from what we would expect by chance, we counted the number of states falling in sliding 125 ms time bins (8 Hz, simulating theta frequency), stepped at 20 ms.

Statistics

All statistical tests are described in the main text or in the corresponding figure legends. Error bars and shading indicate standard error of the mean (s.e.m.) unless otherwise specified. All comparisons were two-tailed.

Resource availability

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Joni Wallis (wallis@berkeley.edu).

4.5 Acknowledgments

We thank Thom Elston, Eric Hu, and Nate Munet for useful feedback on the manuscript. This work was funded by NIMH R01-MH117763 and NIMH R01-MH121448 to JDW. ZZB and JDW designed the experiments. CFF wrote the manuscript. JDW edited the manuscript. ZZB collected the data. CFF analyzed the data. JDW supervised the project.

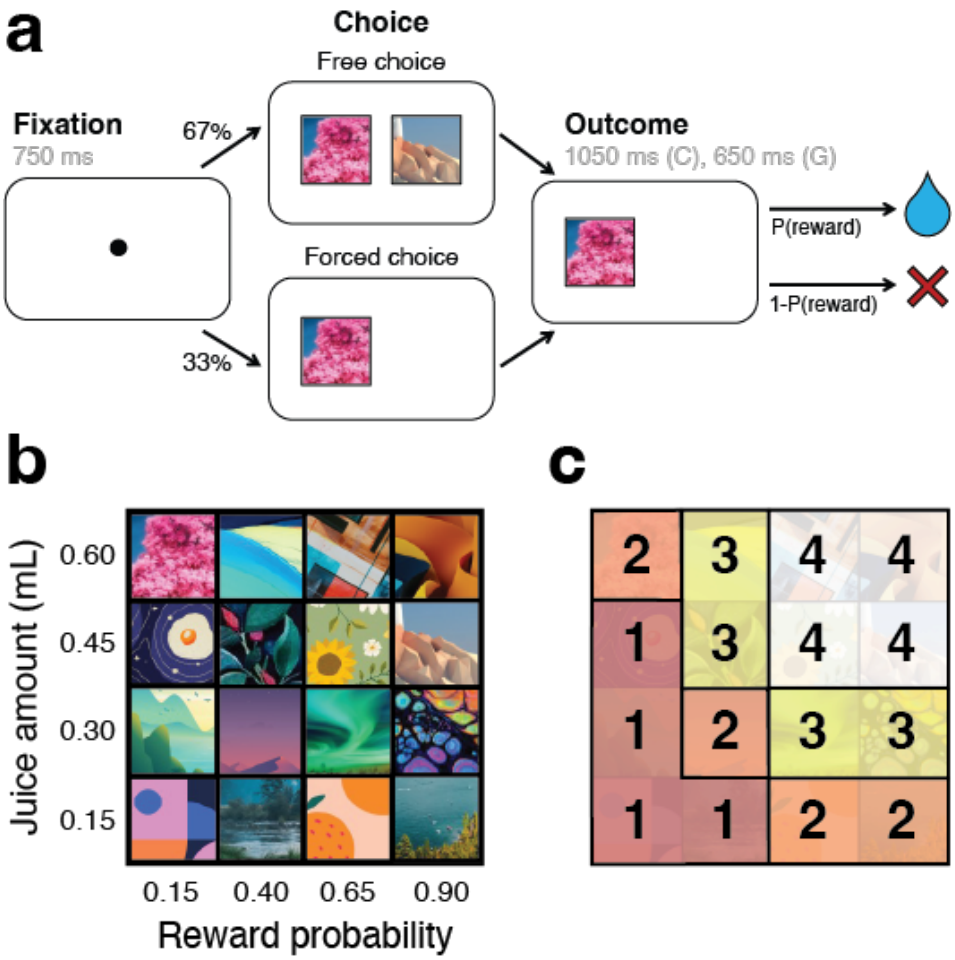

Figure 4.1: Task design. (a) Subjects fixated on a central point to start each trial. On 67% of trials, they were presented with two pictures to choose between (free choice), and on the remaining 33% of trials, they only saw one picture (forced choice). In either condition, subjects indicated their choice with a lever (left or right) and received a juice reward with a probability corresponding to the picture. (b) Subjects learned 16 pictures associated with different juice amounts and reward probabilities. (c) We split the pictures into four equal-sized value bins, with 1 indicating the lowest expected value and 4 indicating the highest expected value. Groupings were the same for both subjects.

Subject	Session	OFC neurons	LFP channels
C	rec07	72	53
	rec09	101	38
	rec17	14	14
	rec20	71	41
	rec33	33	27
	rec47	48	40
`	rec54	38	29
	rec60	42	36
	rec61	138	91
	rec62	79	55
	rec65	81	67
	rec66	116	83
G	rec15	292	181
	rec18	215	163
	rec19	210	166
	rec20	191	145
	rec21	177	132
	rec28	206	165
	rec29	208	155
	rec31	207	157
	rec32	168	137

Table 4.1: Neural data. Number of neurons recorded per session in OFC for subjects C and G. Note that subject G had a bilateral chamber implant, and subject C had a unilateral chamber implant.

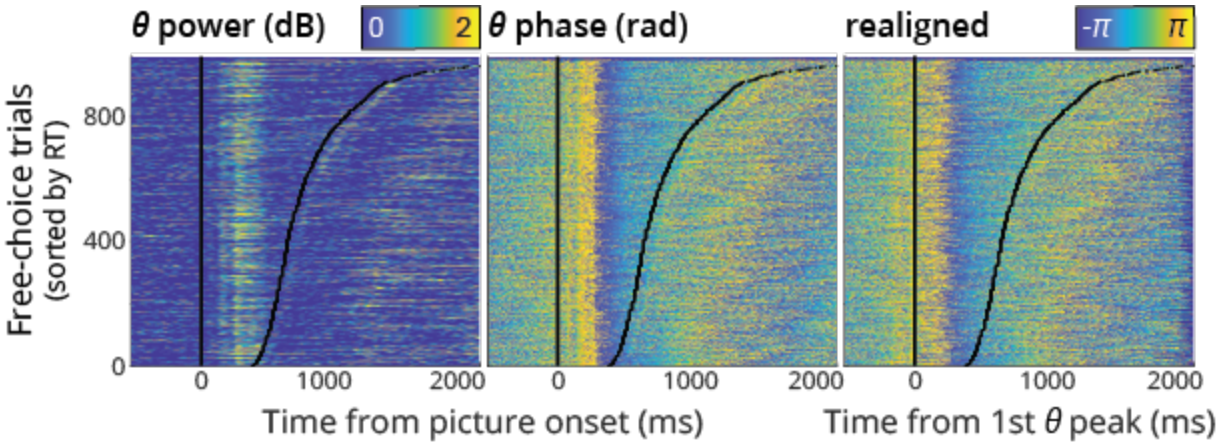

Figure 4.2: Example theta LFP signal. All panels show Hilbert-transformed LFP signals from the theta frequency band (4-8 Hz) recorded from a single electrode channel during one recording session (monkey C). Analytic amplitude is on the left, phase is in the middle, and realigned phase (where values from each trial were shifted so that 0 = time of the first theta peak following picture onset) is on the right. Dark blue patches on the far right of the plot illustrate missing data, after shifting trials. Black vertical lines indicate the time of picture onset (left and middle) or the first theta peak after picture onset (right). All panels show free-choice trials (rows) sorted by response times, with response time marked by a black dot for each trial.

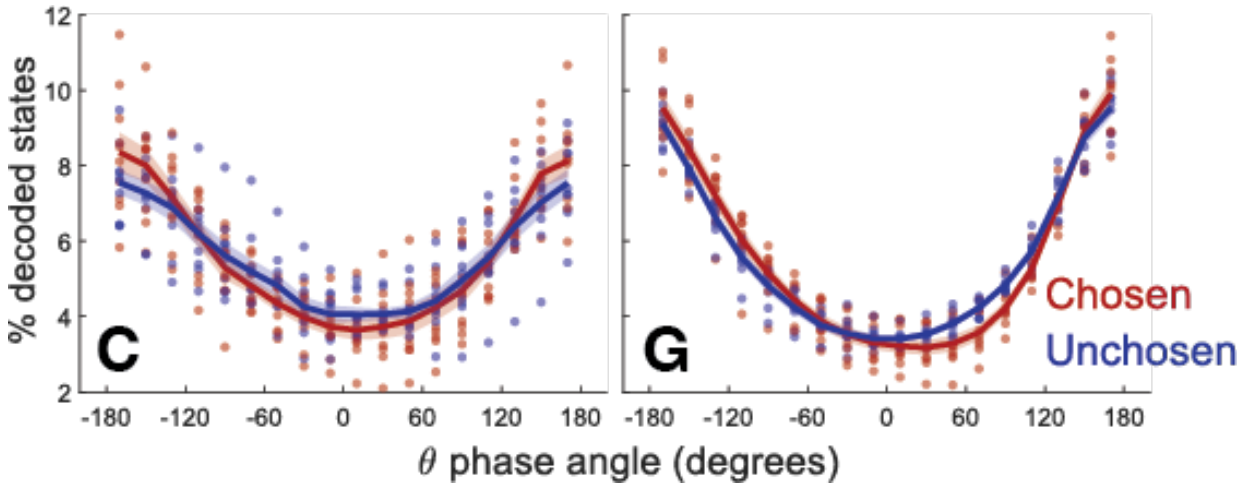

Figure 4.3: Distribution of decoded states across the theta cycle. Each point illustrates the percentage of total OFC value states decoded across the first two theta cycles for a single session. Cycles are broken into 20-degree phase angle bins. Lines represent the mean (±SEM) percentage across all sessions. Red represents chosen value states, and blue represents unchosen value states (left = subject C, right = subject G).

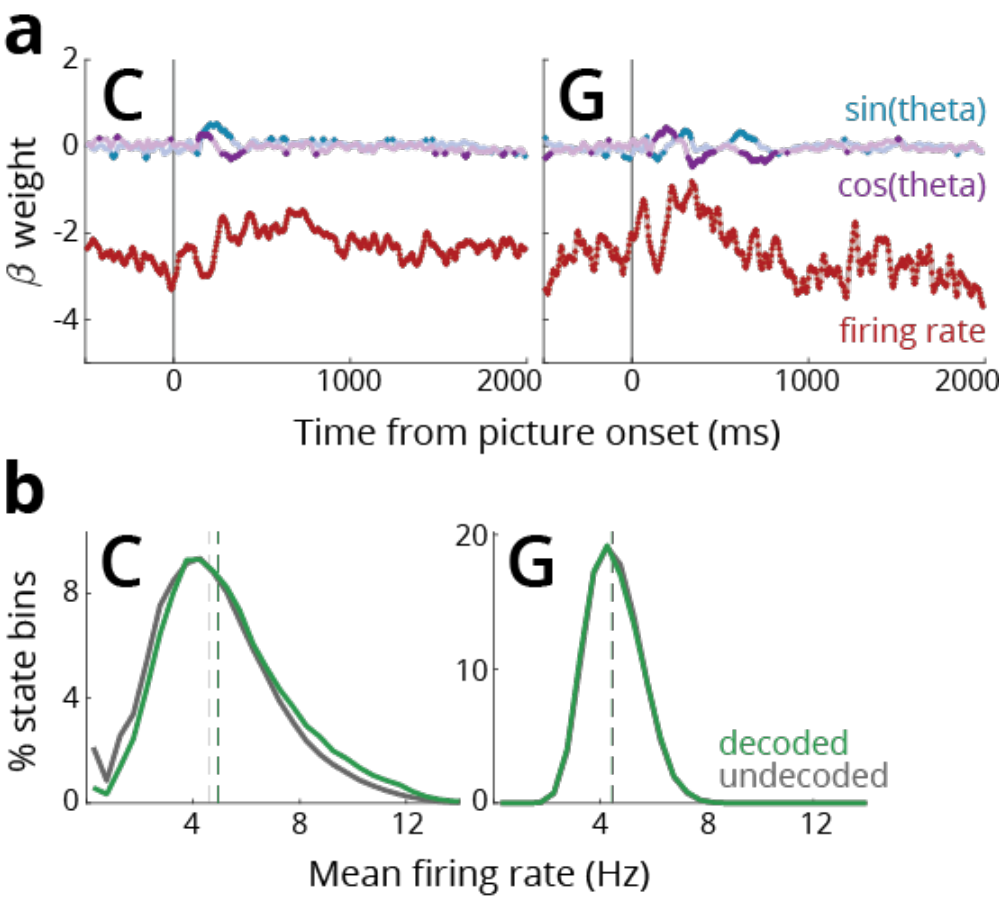

Figure 4.4: Firing rate predicts state decodability. (a) Beta weights (y axis) over time (x axis) for predictors in a logistic regression model of state decodability (red = firing rate, blue/purple = theta phase angle, separated into sine and cosine). Time stamps with significant betas ($p < 0.01$) are represented by bold dots. In both (a) and (b), subject C is on the left and subject G is on the right. (b) Percentage of time bins (y axis, combined across sessions for each subject) falling in each 0.2 Hz firing rate bin, where firing rates were averaged within each bin. Bins where no states were decoded are represented by gray lines, and bins where either the chosen or the unchosen state was decoded are represented by green lines.

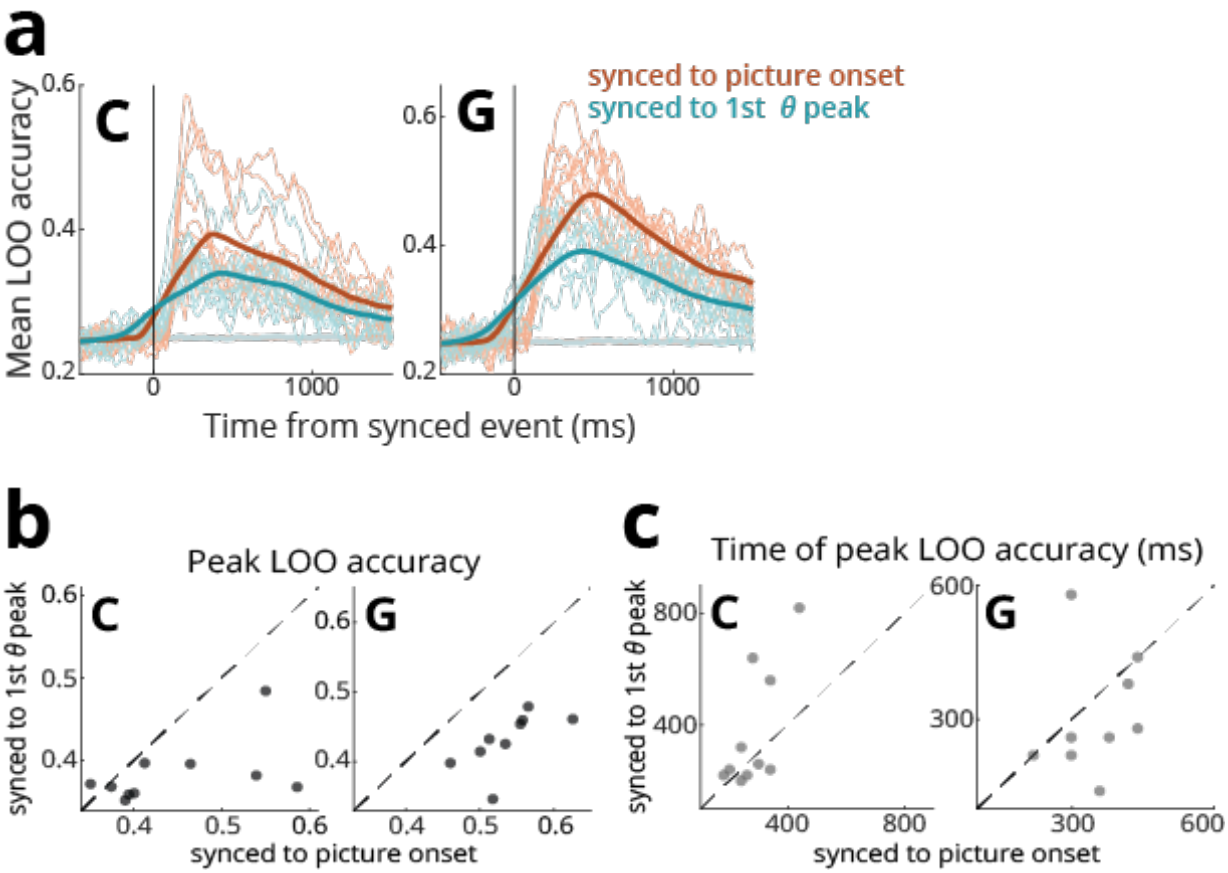

Figure 4.5: OFC value state decoder performance. (a) Mean leave-one-out (LOO) LDA accuracy for decoders trained on OFC data synced to picture onset (orange) and the first theta peak following picture onset (teal). Thin lines represent individual sessions and thick lines represent across-session means. Desaturated lines represent LOO LDA accuracy for decoders trained on shuffled data (shuffle accuracy is roughly chance, 25%). For all panels, monkey C is on the left and monkey G is on the right. (b) Peak LOO accuracy for decoders synced to picture onset (x axis) plotted against peak LOO accuracy for decoders synced to the first theta peak (y axis). Points represent individual sessions. (c) The number of milliseconds after the synced event (picture onset or theta peak) when peak LOO accuracy occurred. Conventions are the same as (b).

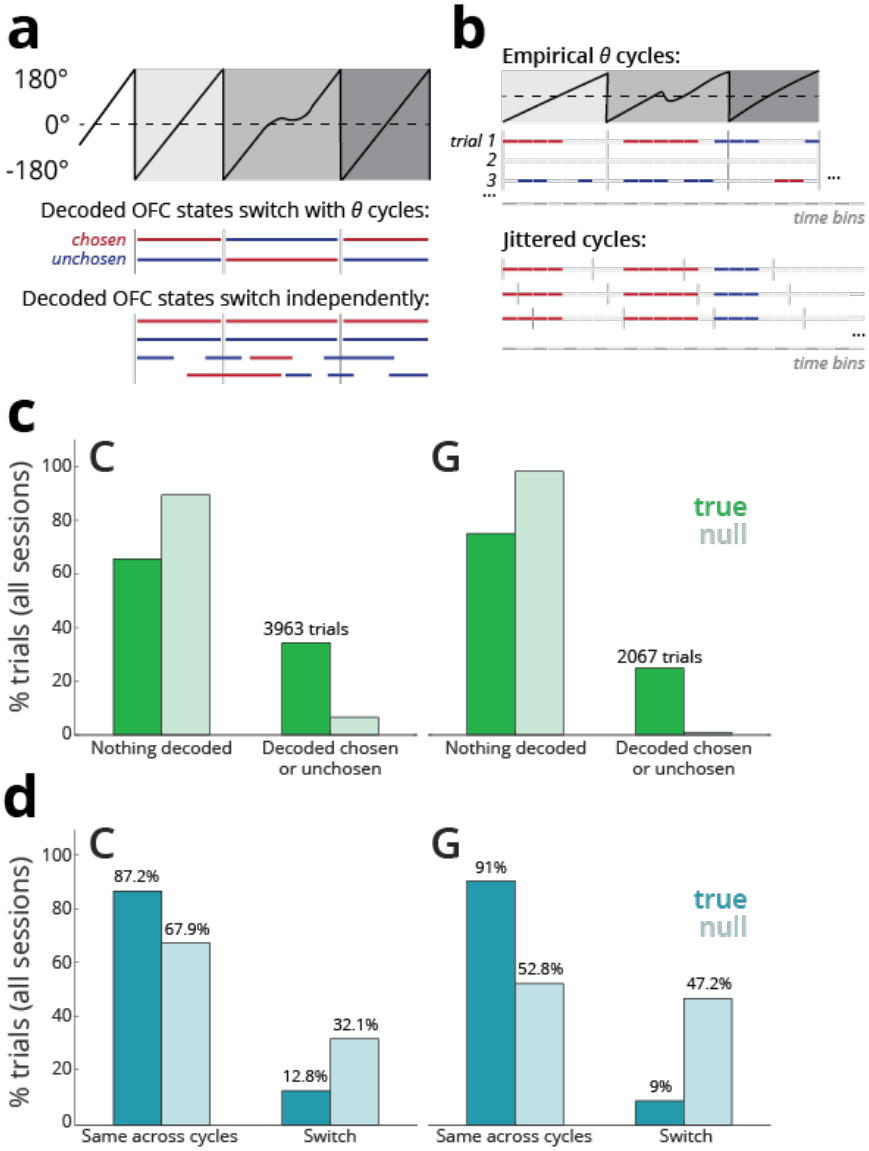

Figure 4.6: State decoding dynamics across theta cycles. (a) Cartoon illustrating theta cycle boundaries within a given trial (top row, with edges depicted as gray vertical lines). One hypothesis (middle row) is that decoded states (red and blue horizontal lines, representing chosen value and unchosen value, respectively) are bound by theta cycles, alternating across cycles. Alternatively (bottom row), decoded states could switch independently of theta cycle boundaries. (b) Illustration of method for labeling each theta cycle. Within each cycle (top row), each time bin was labeled as containing a decoded state, chosen (red) or unchosen (blue), or not (light gray). A cycle was labeled "chosen," "unchosen," or "not decoded" according to the label assigned to the most time bins within that cycle. For comparison, we followed the same procedure to label jittered cycles (bottom row). (c) The percentage of trials, pooled across sessions, where at least one theta cycle was labeled as containing a decoded state between picture onset and choice, given empirical (dark green) and jittered (light green) cycle boundaries. For (c) and (d), subject C is on the left and subject G is on the right. (d) Of those trials where at least one cycle contained a decoded state, we compared the number of trials where the decoded state was the same across theta cycles versus where the decoded state switched across cycles. Counts given empirical cycle boundaries are represented by dark blue bars, and jittered cycle boundaries are represented by light blue bars.

Chapter 5

Conclusion

In this thesis, we aimed to understand how the orbitofrontal cortex (OFC) and the hippocampus (HPC) jointly contribute to value-guided decision making. State inference, or the process of guessing the current state of the world given a set of observations, is critical to successful decision making, especially in a dynamic environment. We hypothesized that neural representations of task state in both OFC and HPC would predict changes in behavior during reversal learning, as animals adapted choice preferences to switches in reward contingencies (Chapter 3). Given recent findings from our lab demonstrating that OFC represents value states sequentially during deliberation (Balewski et al., 2022; Rich & Wallis, 2016), and that HPC-driven theta oscillations in OFC are important for value-guided learning (Knudsen & Wallis, 2020), we also predicted that theta oscillations shape OFC value dynamics (Chapter 4). We used acute electrophysiology to simultaneously record from OFC and HPC during reversal learning (Chapter 2), and observed that both regions represent task state, but theta oscillations do not determine the timing of representational switches in OFC. This chapter will summarize the findings from our experiments, address lingering questions, and propose future directions for research.

Task state is represented by the orbitofrontal cortex and the hippocampus

In both OFC and HPC, we found that the current task state was encoded stably within each trial, independent of stimulus presentation and reward outcome. As representations of task state strengthened in the trials following each reversal, behavioral performance improved. We also observed anticipatory neural representations of upcoming task states in trials immediately preceding upcoming reversals, while neurons sensitive to value expectations continued to reflect current conditions. Together, these results suggest that subjects used their understanding of task structure to anticipate and adapt to changes in reward contingencies, and that OFC and HPC are both involved in this process.

This is consistent with recent ideas arguing that the role of OFC is not in economic choice or valuation narrowly, but in representing task states (Schuck et al., 2018; Stalnaker et al., 2015; Wikenheiser & Schoenbaum, 2016). Our findings are also consistent with the argument that HPC also represents task states and the relationships between task states, or the “cognitive map” (Behrens et al., 2018; Knudsen & Wallis, 2022; Sanders et al., 2020; Whittington et al., 2020). This raises the question of what the distinctions are between each region in task state representation, since it seems unlikely that OFC and

HPC serve redundant roles. One hypothesis, which can be tested in future studies, is that OFC represents the current goal, and uses a map of state transitions stored in the HPC to compute state values, given the current goal (Knudsen & Wallis, 2022).

Orbitofrontal value dynamics

Given the functional connection between OFC and HPC (Barbas & Blatt, 1995; Carmichael & Price, 1995; Young & Shapiro, 2011), and the recent finding that orbitofrontal theta oscillations are critically important for value-guided learning (Knudsen & Wallis, 2020), we tested the hypothesis that theta oscillations shape OFC value dynamics (Balewski et al., 2022; Rich & Wallis, 2016). We related the timing of decoded value states to theta filtered LFP signals and observed that while decodability varied as a function of theta phase angle, states switched independent of theta cycles. Additional analyses revealed that synchronizing training data to a theta phase reset, rather than to an external trial event, decreases decoder accuracy, further emphasizing that factors other than neural oscillations are responsible for guiding value state alternations in OFC.

These findings present a potential difference between OFC and HPC: hippocampal neurons have been shown to flicker between representational states in pace with theta oscillations (Jezek et al., 2011; Kay et al., 2020), while OFC states appear to flicker to the beat of their own drum. It is possible that cross-species differences, both in task design and in anatomy, contribute to this discrepancy. Both OFC and HPC represent value and task state in non-spatial domains (Knudsen & Wallis, 2021, 2022), so future studies may be able to address this by comparing the relationship between theta oscillations and value state switches across the two regions, within a single task and animal model.

Orbitofrontal-hippocampal contributions and interactions

While some recent studies have explored the parallel contributions of OFC And HPC to state inference during value-based decision-making (Wikenheiser & Schoenbaum, 2016; Zhou, Gardner, et al., 2019; Zhou, Montesinos-Cartagena, et al., 2019), to our knowledge, our results are the first to extend this framework to nonhuman primates. Given the dramatic differences in size and complexity of the anterior HPC (Insausti, 1993) between rodents and primates, and differences in anatomical and functional organization of the OFC (Wallis, 2012), neurophysiology in nonhuman primates is an important bridge between basic science and potential therapeutic applications in humans. Moving forward, applying techniques like closed-loop microstimulation (Knudsen & Wallis, 2020) or pharmacogenetic tools like DREADDs (Grayson et al., 2016) to temporarily deactivate OFC or HPC during state inference will help disentangle the relative contributions of each region.

While our results did not support a role for neural oscillations in OFC value dynamics, it is still likely that LFP phase-locking between the prefrontal cortex and the HPC facilitates their communication during learning and decision making (Brincat & Miller, 2015; Knudsen & Wallis, 2020; Morici et al., 2022; Riceberg et al., 2022; Siapas et al., 2005; Tang et al., 2021; van Wingerden et al., 2010). Examining LFP coherence between OFC and HPC will expand our understanding of how the brain builds a cognitive map of task structure, implements it during value-guided learning, and updates it in response to change.

Bibliography

Backus, A. R., Schoffelen, J.-M., Szebényi, S., Hanslmayr, S., & Doeller, C. F. (2016). Hippocampal-Prefrontal Theta Oscillations Support Memory Integration. *Current Biology: CB*, *26*(4), 450–457.

Balewski, Z. Z., Knudsen, E. B., & Wallis, J. D. (2022). Fast and slow contributions to decision-making in corticostriatal circuits. *Neuron*, *110*(13), 2170–2182.e4.

Barbas, H., & Blatt, G. J. (1995). Topographically specific hippocampal projections target functionally distinct prefrontal areas in the rhesus monkey. *Hippocampus*, *5*(6), 511–533.

Bartolo, R., & Averbeck, B. B. (2020). Prefrontal Cortex Predicts State Switches during Reversal Learning. *Neuron*. https://doi.org/10.1016/j.neuron.2020.03.024

Baxter, M. G., Parker, A., Lindner, C. C., Izquierdo, A. D., & Murray, E. A. (2000). Control of response selection by reinforcer value requires interaction of amygdala and orbital prefrontal cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *20*(11), 4311–4319.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*(5304), 1293–1295.

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron*, *100*(2), 490–509.

Benchenane, K., Peyrache, A., Khamassi, M., Tierney, P. L., Gioanni, Y., Battaglia, F. P., & Wiener, S. I. (2010). Coherent Theta Oscillations and Reorganization of Spike Timing in the Hippocampal- Prefrontal Network upon Learning. In *Neuron* (Vol. 66, Issue 6, pp. 921–936). https://doi.org/10.1016/j.neuron.2010.05.013

Blanchard, T. C., Hayden, B. Y., & Bromberg-Martin, E. S. (2015). Orbitofrontal cortex uses distinct codes for different choice attributes in decisions motivated by curiosity. *Neuron*, *85*(3), 602–614.

Bouret, S., & Richmond, B. J. (2010). Ventromedial and orbital prefrontal neurons differentially encode internally and externally driven motivational values in monkeys. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(25), 8591–8601.

Bradfield, L. A., Dezfouli, A., van Holstein, M., Chieng, B., & Balleine, B. W. (2015). Medial Orbitofrontal Cortex Mediates Outcome Retrieval in Partially Observable Task Situations. *Neuron*, *88*(6), 1268–1280.

Brincat, S. L., & Miller, E. K. (2015). Frequency-specific hippocampal-prefrontal interactions during associative learning. *Nature Neuroscience*, *18*(4), 576–581.

Buckley, M. J., Mansouri, F. A., Hoda, H., Mahboubi, M., Browning, P. G. F., Kwok, S. C., Phillips, A., & Tanaka, K. (2009). Dissociable Components of Rule-Guided Behavior Depend on Distinct Medial and Prefrontal Regions. In *Science* (Vol. 325, Issue 5936, pp. 52–58). https://doi.org/10.1126/science.1172377

Buzsáki, G. (2002). Theta oscillations in the hippocampus. *Neuron*, *33*(3), 325–340.

Buzsáki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents--EEG, ECoG, LFP and spikes. *Nature Reviews. Neuroscience*, *13*(6), 407–420.

Buzsáki, G., Logothetis, N., & Singer, W. (2013). Scaling brain size, keeping timing: evolutionary preservation of brain rhythms. *Neuron*, *80*(3), 751–764.

Canolty, R. T., Ganguly, K., Kennerley, S. W., Cadieu, C. F., Koepsell, K., Wallis, J. D., & Carmena, J. M. (2010). Oscillatory phase coupling coordinates anatomically dispersed functional cell assemblies. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(40), 17356–17361.

Carmichael, S. T., & Price, J. L. (1995). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *The Journal of Comparative Neurology*, *363*(4), 615–641.

Chadwick, M. J., Hassabis, D., Weiskopf, N., & Maguire, E. A. (2010). Decoding individual episodic memory traces in the human hippocampus. *Current Biology: CB*, *20*(6), 544–547.

Chan, S. C. Y., Niv, Y., & Norman, K. A. (2016). A Probability Distribution over Latent Causes, in the Orbitofrontal Cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *36*(30), 7817–7828.

Collins, A. G. E., & Cockburn, J. (2020). Beyond dichotomies in reinforcement learning. *Nature Reviews. Neuroscience*, *21*(10), 576–586.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215.

Dias, R., Robbins, T. W., & Roberts, A. C. (1996). Primate analogue of the Wisconsin Card Sorting Test: effects of excitotoxic lesions of the prefrontal cortex in the marmoset. *Behavioral Neuroscience*, *110*(5), 872–886.

Eichenbaum, H. (2017). The role of the hippocampus in navigation is memory. *Journal of Neurophysiology*, *117*(4), 1785–1796.

Fanselow, M. S., & Dong, H.-W. (2010). Are the dorsal and ventral hippocampus functionally distinct structures? *Neuron*, *65*(1), 7–19.

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J. V., Pieper, S., & Kikinis, R. (2012). 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic Resonance Imaging*, *30*(9), 1323–1341.

Fellows, L. K., & Farah, M. J. (2003). Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm. *Brain: A Journal of Neurology*, *126*(Pt 8), 1830–1837.

Fernando, A. B. P., & Robbins, T. W. (2011). Animal models of neuropsychiatric disorders. *Annual Review of Clinical Psychology*, *7*, 39–61.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical

representation of visual stimuli in the primate prefrontal cortex. *Science*, *291*(5502), 312–316.

Gallagher, M., McMahan, R. W., & Schoenbaum, G. (1999). Orbitofrontal cortex and representation of incentive value in associative learning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *19*(15), 6610–6614.

Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R., & Fried, I. (2008). Internally generated reactivation of single neurons in human hippocampus during free recall. *Science*, *322*(5898), 96–101.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*(4), 585–595.

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science*, *301*(5636), 1104–1107.

Grayson, D. S., Bliss-Moreau, E., Machado, C. J., Bennett, J., Shen, K., Grant, K. A., Fair, D. A., & Amaral, D. G. (2016). The Rhesus Monkey Connectome Predicts Disrupted Functional Networks Resulting from Pharmacogenetic Inactivation of the Amygdala. *Neuron*, *91*(2), 453–466.

Hare, T. A., Schultz, W., Camerer, C. F., O'Doherty, J. P., & Rangel, A. (2011). Transformation of stimulus value signals into motor commands during simple choice. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(44), 18120–18125.

Hartley, T., Lever, C., Burgess, N., & O'Keefe, J. (2014). Space in the brain: how the hippocampal formation supports spatial cognition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1635), 20120510.

Hayden, B. Y., & Moreno-Bote, R. (2018). A neuronal theory of sequential economic choice. *Brain and Neuroscience Advances*, *2*, 2398212818766675.

Herreras, O. (2016). Local Field Potentials: Myths and Misunderstandings. *Frontiers in Neural Circuits*, *10*, 101.

Hunt, L. T., Malalasekera, W. M. N., de Berker, A. O., Miranda, B., Farmer, S. F., Behrens, T. E. J., & Kennerley, S. W. (2018). Triple dissociation of attention and decision computations across prefrontal cortex. *Nature Neuroscience*, *21*(10), 1471–1481.

Insausti, R. (1993). Comparative anatomy of the entorhinal cortex and hippocampus in mammals. *Hippocampus*, *3 Spec No*, 19–26.

Iversen, S. D., & Mishkin, M. (1970). Perseverative interference in monkeys following selective lesions of the inferior prefrontal convexity. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *11*(4), 376–386.

Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H., & Holmes, A. (2017). The neural basis of reversal learning: An updated perspective. *Neuroscience*, *345*, 12–26.

Izquierdo, A., & Jentsch, J. D. (2012). Reversal learning as a measure of impulsive and compulsive behavior in addictions. *Psychopharmacology*, *219*(2), 607–620.
Jezek, K., Henriksen, E. J., Treves, A., Moser, E. I., & Moser, M.-B. (2011). Theta-paced flickering between place-cell maps in the hippocampus. *Nature*, *478*(7368), 246–249.
Jones, B., & Mishkin, M. (1972). Limbic lesions and the problem of stimulus--reinforcement associations. *Experimental Neurology*, *36*(2), 362–377.
Karakaş, S. (2020). A review of theta oscillation and its functional correlates. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *157*, 82–99.
Kay, K., Chung, J. E., Sosa, M., Schor, J. S., Karlsson, M. P., Larkin, M. C., Liu, D. F., & Frank, L. M. (2020). Constant Sub-second Cycling between Representations of Possible Futures in the Hippocampus. *Cell*, *180*(3), 552–567.e25.
Kjelstrup, K. B., Solstad, T., Brun, V. H., Hafting, T., Leutgeb, S., Witter, M. P., Moser, E. I., & Moser, M.-B. (2008). Finite scale of spatial representation in the hippocampus. *Science*, *321*(5885), 140–143.
Knudsen, E. B., Balewski, Z. Z., & Wallis, J. D. (2019). A model-based approach for targeted neurophysiology in the behaving non-human primate. *International IEEE/EMBS Conference on Neural Engineering: [proceedings]. International IEEE EMBS Conference on Neural Engineering*, *2019*, 195–198.
Knudsen, E. B., & Wallis, J. D. (2020). Closed-Loop Theta Stimulation in the Orbitofrontal Cortex Prevents Reward-Based Learning. *Neuron*, *106*(3), 537–547.e4.
Knudsen, E. B., & Wallis, J. D. (2021). Hippocampal neurons construct a map of an abstract value space. *Cell*, *184*(18), 4640–4650.e10.
Knudsen, E. B., & Wallis, J. D. (2022). Taking stock of value in the orbitofrontal cortex. *Nature Reviews. Neuroscience*, *23*(7), 428–438.
Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298.
Langdon, A. J., Song, M., & Niv, Y. (2019). Uncovering the "state": Tracing the hidden state representations that structure learning and decision-making. *Behavioural Processes*, *167*, 103891.
Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017). Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. *Neuron*, *93*(2), 451–463.
Lim, S.-L., O'Doherty, J. P., & Rangel, A. (2011). The decision value computations in the vmPFC and striatum use a relative value code that is guided by visual attention. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(37), 13214–13223.
Mark, S., Romani, S., Jezek, K., & Tsodyks, M. (2017). Theta-paced flickering between place-cell maps in the hippocampus: A model based on short-term synaptic

plasticity. *Hippocampus*, *27*(9), 959–970.
McGinty, V. B., Rangel, A., & Newsome, W. T. (2016). Orbitofrontal Cortex Value Signals Depend on Fixation Location during Free Viewing. *Neuron*, *90*(6), 1299–1311.
Miller, E. K., Freedman, D. J., & Wallis, J. D. (2002). The prefrontal cortex: categories, concepts and cognition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *357*(1424), 1123–1136.
Miller, E. K., Nieder, A., Freedman, D. J., & Wallis, J. D. (2003). Neural correlates of categories and concepts. *Current Opinion in Neurobiology*, *13*(2), 198–203.
Morici, J. F., Weisstaub, N. V., & Zold, C. L. (2022). Hippocampal-medial prefrontal cortex network dynamics predict performance during retrieval in a context-guided object memory task. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(20), e2203024119.
Moser, M. B., & Moser, E. I. (1998). Functional differentiation in the hippocampus. *Hippocampus*, *8*(6), 608–619.
Niv, Y. (2019). Learning task-state representations. *Nature Neuroscience*, *22*(10), 1544–1553.
Noonan, M. P., Mars, R. B., & Rushworth, M. F. S. (2011). Distinct roles of three frontal cortical areas in reward-guided behavior. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(40), 14399–14412.
O'keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon Press.
Padoa-Schioppa, C. (2011). Neurobiology of economic choice: a good-based model. *Annual Review of Neuroscience*, *34*, 333–359.
Padoa-Schioppa, C. (2013). Neuronal origins of choice variability in economic decisions. *Neuron*, *80*(5), 1322–1336.
Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, *441*(7090), 223–226.
Padoa-Schioppa, C., & Conen, K. E. (2017). Orbitofrontal Cortex: A Neural Circuit for Economic Decisions. In *Neuron* (Vol. 96, Issue 4, pp. 736–754). https://doi.org/10.1016/j.neuron.2017.09.031
Paxinos, G., Huang, X.-F., & Toga, A. W. (2000). *The Rhesus Monkey Brain in Stereotaxic Coordinates*. https://ro.uow.edu.au/hbspapers/3613/
Place, R., Farovik, A., Brockmann, M., & Eichenbaum, H. (2016). Bidirectional prefrontal-hippocampal interactions support context-guided memory. *Nature Neuroscience*, *19*(8), 992–994.
Rainer, G., Asaad, W. F., & Miller, E. K. (1998). Memory fields of neurons in the primate prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(25), 15008–15013.
Riceberg, J. S., Srinivasan, A., Guise, K. G., & Shapiro, M. L. (2022). Hippocampal signals modify orbitofrontal representations to learn new paths. *Current Biology:*

CB, *32*(15), 3407–3413.e6.
Rich, E. L., & Wallis, J. D. (2014). Medial-lateral organization of the orbitofrontal cortex. *Journal of Cognitive Neuroscience*, *26*(7), 1347–1362.
Rich, E. L., & Wallis, J. D. (2016). Decoding subjective decisions from orbitofrontal cortex. *Nature Neuroscience*, *19*(7), 973–980.
Royer, S., Sirota, A., Patel, J., & Buzsáki, G. (2010). Distinct representations and theta dynamics in dorsal and ventral hippocampus. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(5), 1777–1787.
Rudebeck, P. H., Mitz, A. R., Chacko, R. V., & Murray, E. A. (2013). Effects of amygdala lesions on reward-value coding in orbital and medial prefrontal cortex. *Neuron*, *80*(6), 1519–1531.
Sadacca, B. F., Wied, H. M., Lopatina, N., Saini, G. K., Nemirovsky, D., & Schoenbaum, G. (2018). Orbitofrontal neurons signal sensory associations underlying model-based inference in a sensory preconditioning task. *eLife*, *7*. https://doi.org/10.7554/eLife.30373
Samuel J Gershman, Kenneth A Norman, Yael Niv. (2015). Discovering latent causes in reinforcement learning. In *Current Opinion in Behavioral Sciences*. https://doi.org/10.1016/j.cobeha.2015.07.007
Sanders, H., Wilson, M. A., & Gershman, S. J. (2020). Hippocampal remapping as hidden state inference. *eLife*, *9*. https://doi.org/10.7554/eLife.51140
Schoenbaum, G., Setlow, B., Nugent, S. L., Saddoris, M. P., & Gallagher, M. (2003). Lesions of orbitofrontal cortex and basolateral amygdala complex disrupt acquisition of odor-guided discriminations and reversals. *Learning & Memory* , *10*(2), 129–140.
Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron*, *91*(6), 1402–1412.
Schuck, N. W., Wilson, R., & Niv, Y. (2018). Chapter 12 - A State Representation for Reinforcement Learning and Decision-Making in the Orbitofrontal Cortex. In R. Morris, A. Bornstein, & A. Shenhav (Eds.), *Goal-Directed Decision Making* (pp. 259–278). Academic Press.
Siapas, A. G., Lubenov, E. V., & Wilson, M. A. (2005). Prefrontal phase locking to hippocampal theta oscillations. *Neuron*, *46*(1), 141–151.
Smith, D. V., Hayden, B. Y., Truong, T.-K., Song, A. W., Platt, M. L., & Huettel, S. A. (2010). Distinct value signals in anterior and posterior ventromedial prefrontal cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(7), 2490–2495.
Stalnaker, T. A., Cooch, N. K., & Schoenbaum, G. (2015). What the orbitofrontal cortex does not do. *Nature Neuroscience*, *18*(5), 620–627.
Strait, C. E., Blanchard, T. C., & Hayden, B. Y. (2014). Reward value comparison via mutual inhibition in ventromedial prefrontal cortex. *Neuron*, *82*(6), 1357–1366.
Strait, C. E., Sleezer, B. J., Blanchard, T. C., Azab, H., Castagno, M. D., & Hayden, B. Y.

(2016). Neuronal selectivity for spatial positions of offers and choices in five reward regions. *Journal of Neurophysiology*, *115*(3), 1098–1111.

Strange, B. A., Witter, M. P., Lein, E. S., & Moser, E. I. (2014). Functional organization of the hippocampal longitudinal axis. *Nature Reviews. Neuroscience*, *15*(10), 655–669.

Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: an introduction MIT Press. *Cambridge, MA*.

Takahashi, Y. K., Roesch, M. R., Wilson, R. C., Toreson, K., O'Donnell, P., Niv, Y., & Schoenbaum, G. (2011). Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. *Nature Neuroscience*, *14*(12), 1590–1597.

Tang, W., Shin, J. D., & Jadhav, S. P. (2021). Multiple time-scales of decision-making in the hippocampus and prefrontal cortex. *eLife*, *10*. https://doi.org/10.7554/eLife.66227

Tolman, E. C. (1948). Cognitive maps in rats and men. In *Psychological Review* (Vol. 55, Issue 4, pp. 189–208). https://doi.org/10.1037/h0061626

Valentin, V. V., Dickinson, A., & O'Doherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human brain. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *27*(15), 4019–4026.

van Wingerden, M., Vinck, M., Lankelma, J., & Pennartz, C. M. A. (2010). Theta-band phase locking of orbitofrontal neurons during reward expectancy. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(20), 7078–7087.

Wallis, J. D. (2012). Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nature Neuroscience*, *15*(1), 13–19.

Wallis, J. D. (2018). Decoding Cognitive Processes from Neural Ensembles. *Trends in Cognitive Sciences*, *22*(12), 1091–1102.

Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature*, *411*(6840), 953–956.

Wallis, J. D., Dias, R., Robbins, T. W., & Roberts, A. C. (2001). Dissociable contributions of the orbitofrontal and lateral prefrontal cortex of the marmoset to performance on a detour reaching task. *The European Journal of Neuroscience*, *13*(9), 1797–1808.

Wallis, J. D., & Rich, E. L. (2011). Challenges of Interpreting Frontal Neurons during Value-Based Decision-Making. *Frontiers in Neuroscience*, *5*, 124.

Whittington, J. C. R., McCaffary, D., Bakermans, J. J. W., & Behrens, T. E. J. (2022). How to build a cognitive map. *Nature Neuroscience*, *25*(10), 1257–1272.

Whittington, J. C. R., Muller, T. H., Mark, S., Chen, G., Barry, C., Burgess, N., & Behrens, T. E. J. (2020). The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. *Cell*, *183*(5), 1249–1263.e23.

Wikenheiser, A. M., & Schoenbaum, G. (2016). Over the river, through the woods:

cognitive maps in the hippocampus and orbitofrontal cortex. *Nature Reviews. Neuroscience*, *17*(8), 513–523.

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, *81*(2), 267–279.

Xie, Y., Nie, C., & Yang, T. (2018). Covert shift of attention modulates the value encoding in the orbitofrontal cortex. *eLife*, *7*. https://doi.org/10.7554/eLife.31507

Young, J. J., & Shapiro, M. L. (2011). Dynamic coding of goal-directed paths by orbital prefrontal cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(16), 5989–6000.

Yun, M., Hwang, J. Y., & Jung, M. W. (2023). Septotemporal variations in hippocampal value and outcome processing. *Cell Reports*, *42*(2), 112094.

Zhou, J., Gardner, M. P. H., Stalnaker, T. A., Ramus, S. J., Wikenheiser, A. M., Niv, Y., & Schoenbaum, G. (2019). Rat Orbitofrontal Ensemble Activity Contains Multiplexed but Dissociable Representations of Value and Task Structure in an Odor Sequence Task. *Current Biology: CB*, *29*(6), 897–907.e3.

Zhou, J., Montesinos-Cartagena, M., Wikenheiser, A. M., Gardner, M. P. H., Niv, Y., & Schoenbaum, G. (2019). Complementary Task Structure Representations in Hippocampus and Orbitofrontal Cortex during an Odor Sequence Task. *Current Biology: CB*, *29*(20), 3402–3409.e3.

Zhu, L., Mathewson, K. E., & Hsu, M. (2012). Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(5), 1419–1424.