UC Davis
UC Davis Electronic Theses and Dissertations

Title

Neural Mechanisms of Contingent Learning in a Hierarchically Structured World

Permalink

https://escholarship.org/uc/item/1sk628f3

Author

Zhang, Yanchang

Publication Date

2024

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Neural Mechanisms of Contingent Learning in a Hierarchically Structured World

By

YANCHANG ZHANG
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Erie D. Boorman, Chair

Charan Ranganath

Andrew P. Yonelinas

Committee in Charge
2024

© Copyright by Yanchang Zhang, 2024.

All rights reserved.

Acknowledgements

First and foremost, I would like to express my deep gratitude and appreciation to my advisor, Erie Boorman. Your immense vision and ideas, and persistent efforts in science have set standards that I continually strive to meet. You have inspired me with the essential qualities of a good scientist: curiosity about the unknown and high standards in conducting research. Your guidance and support have helped me become an independent researcher. I am honored to be your first PhD student and eternally grateful for your kindness and support during several of my personal challenges.

I would also like to thank my dissertation committee members, Charan Ranganath and Andrew Yonelinas. Thank you both for your insightful and critical input, which has helped me mature as a scientist.

Next, I extend my heartfelt thanks to Phillip Witkowski for his generous hands-on help with data analysis and his availability whenever I encountered challenges. No matter how convoluted my questions were, you never complained and always took the time to walk me through lines of code with great patience. I aspire to be as kind and knowledgeable as you are, both in science and in life.

My PhD journey would not have been the same without the unwavering support of Seongmin Alex Park. Your patience and knowledge have been invaluable, especially during moments when I struggled to see progress. I am grateful of the days you spent with me during scanning sessions, helping with tricky technical fMRI issues, teaching me how to design well with Illustrator, and encouraging me to push myself further in science. I will always cherish our deep conversations about both science and art.

Thanks to those who helped my research with their expertise, including Zhiheng Zhou who helped with functional localizers in Chapter 2, and Dr. Hamed Nili who offered his statistical knowledge on informational connectivity for future analyses building on Chapter 2.

I also wish to express my gratitude to my undergraduate advisor, Bruce Hansen. Without your guidance, I would not have pursued science. Your encouragement has been a great source of support. Thank you for taking a chance on me.

The members of LDM labs have made my PhD experience truly precious. I enjoyed our scientific discussions during lab meetings and conferences, as well as our time together outside the lab. A large portion of my cherished memories involves getting to know each of you, going on food trips, snowboarding adventures, and many more. Special thanks to the support from the friends I met in Davis including Lindsay Rondot and Xinger Yu. To my close friends outside the lab, Constance Cheng, Ivy Zheng, Jessica Chen, Zoe Zhong, Mingli Liang, and Ziwei Zhang: thank you for being by my side, supporting me, and sharing moments of sadness and joy.

My sincere appreciation extends to those who have helped my mental health, including Dr. Michelle Martin and my therapist Melody Sit. Without the professional help I received from both of you, there was no way I could have finished my PhD. Thanks to both of you, I have learned how to take care of myself in the face of adversity. Thanks to my friend and lab mate Mark Orloff, and Alex Park for encouraging me to seek help in mental health and sharing resources with me.

I would also like to thank my significant other, Austin Weber. Meeting you has been the best thing that happened during my PhD. You are the best partner I could imagine. Thank you for always being by my side and supporting me to become the person I want to be.

Last but not least, I want to thank my parents. Your unconditional love is timeless and has become my greatest source of confidence. Thank you for having supported me study abroad far from home since I was 15 until today, and for encouraging me when I decided to pursue a PhD. Your open-mindedness and love from afar have allowed me to learn and live freely and pursue my dreams.

It has been a wild journey. I will cherish what I have learned as I begin the next chapter in life.

Abstract

Humans and animals can flexibly attribute causes to past choices upon receiving outcomes and use learned information to make predictions in novel scenarios. In this doctoral dissertation, I present evidence on how the brain facilitates these processes by activating a network of regions during feedback, each supporting distinct functions in learning and decision making. Chapter 1 reviews relevant literature on key brain regions and their functions involved in causal learning and memory. Chapter 2 investigates the neural mechanisms in contingent learning, demonstrating parallel reinstatements in the frontal cortex for observed and inferred choices, with subregions in the orbitofrontal cortex representing distinct dimensions of information. It also highlights how the dopaminergic midbrain reflects individual differences in inferential ability. Chapter 3 extends beyond the hierarchical task structure used in Chapter 2 and assesses how contingent learning and abstraction are shaped by hierarchical levels in environments with multiple causal layers. Chapter 4 proposes future research directions based on these findings. Collectively, these results elucidate neural mechanisms in contingent learning, essential for building and updating cognitive maps.

Table of Contents

Acknowledgements ..... **ii**

Abstract ..... **v**

Chapter 1 Relevant Literature **1**

The hippocampus and neocortical networks in navigation and memory *1*

What is hippocampal replay? *13*

The hippocampus in configural and conjunctive reinforcement learning *21*

The orbitofrontal cortex in contingent learning *28*

The functions of dopamine in learning *37*

Chapter 2 Parallel Reinstatement of Observed and Inferred Choices in the Human Frontal Cortex during Contingent Learning **43**

Abstract *43*

Introduction *44*

Results *49*

Discussion *71*

References *99*

Supplementary Information *110*

Chapter 3. Hierarchical Causal Levels Shape and Support Learning and Abstraction **122**

Introduction *122*

Methods and Results *123*

Discussion *134*

Chapter 4. General Discussion **136**

References *143*

Chapter 1 Relevant Literature

In this chapter, I will first compare the contributions of the hippocampal network and neocortical regions in navigation and episodic memory. It is important to examine their functions in detail, since both networks appear frequently in studies related to learning and decision making. Then I will talk about the hippocampus in the contexts of replay and configural and conjunctive reinforcement learning. This chapter focuses on the works of hippocampal functions to provide a foundation for Chapter 3, in which we investigate how the hippocampus might facilitate hierarchical credit assignment in a structured environment. This Chapter could be a resourceful reference especially for the fMRI experiment proposed in Chapter 3 where we make hypotheses about the medial temporal lobe in hierarchical credit assignment. This chapter will end with a broad discussion of the functions of orbitofrontal cortex and dopamine in credit assignment.

The hippocampus and neocortical networks in navigation and memory

The hippocampus (HI) plays an important role in both spatial navigation and episodic memory. Its versatile functions range from representing spatial, temporal, and task-dependent information (Wikenheiser et al., 2016; Liman et al., 2017) to encoding episodic contexts (Ritchey et al., 2015) and enhancing and maintaining memories after encoding events (Gruber et al., 2016). However, it is not to say that the HI is the only brain region that enables animals to perform tasks related to navigation and form episodic memories. Several studies have pointed out the significance of other brain regions with similar and complementary functions as the HI in navigation and memory (Diana et al., 2008; Ritchey et al., 2015; Starwarcyzk et al., 2019; O'Mara et al., 2019). Critically, these regions work together with the HI in various tasks,

allowing efficient and holistic representations of events to be formed in navigation- and memory-related tasks. The HI functions together with these neocortical regions including the perirhinal cortex (PRC), the parahippocampal cortex (PHC) (Jenkins et al., 2010; Naya et al., 2011; Eichenbaum et al., 2012; Reagh et al., 2018), and others like the thalamic nuclei and the claustrum (O'Mara et al., 2019), many of which have strong connections with the HI. The HI and these other regions play subtly different roles depending on tasks, such as representing contexts with different specificity (Reagh et al., 2018) on different temporal scales (Jenkins et al., 2010; Naya et al., 2011), and perhaps even depending on different adaptive requirements (life-threatening vs. exploratory environments). Thus, it is reasonable to say that, functionally speaking, the HI and these neocortical and subcortical regions are highly interconnected, and these connections demonstrate a large network in forming representations in spatial navigation and episodic memory. Overall, I believe that the HI and other regions work cooperatively as a whole network and independently if we look at spatial navigation and memory encoding closely on a finer scale. In the following sections, I will first talk about the HI and other regions separately on their specific roles in episodic memory and spatial navigation, then show evidence on how the HI and other regions interact complementarily.

The role of the hippocampus

Pioneering work by O'Keefe and Nadel (1978) introduced the neural basis of Tolman's (1948) "cognitive map" in the hippocampus, revealing how it enables animals and humans to navigate and remember spatial environments. This discovery was based on identifying "place cells" in the hippocampus, neurons that fire in distinct patterns corresponding to specific locations, thereby creating an internal map of the environment. Beyond navigation, the hippocampus is

also critical for forming both long-term spatial and nonspatial declarative memories, emphasizing its broader role in memory processing (Cave & Squire, 1991; Teng & Squire, 1999).

Since then, The HI's role in spatial navigation and memory formation has become an important line of research. A common example of the HI's mapping function could be observed in rats' studies. In a spatial task such as the T-maze, hippocampal neurons have firing fields (place fields) tiling to the physical environment, encoding the animal's relative positions (Wikenheiser et al., 2016). Hippocampal pyramidal neurons (place cells and time cells) exhibited organized spatial and temporal firing patterns when rats ran along a linear track and in a T-maze water reward task (Lisman et al., 2017). If we talk about spatial navigation not only in a physical space, but also in a task-based space, the same HI neuronal assemblies were observed to encode sequential states, representing each individual state in relation to one another (Wikenheiser et al., 2016; McKenzie et al., 2013). More examples include the discovery of neurons in the hippocampus mapping to different sound frequencies (Aronov et al. 2017), numbered lap events (Sun et al., 2020), and relative values in an abstract value space (Knudsen & Wallis, 2021). In a sound-manipulation task in which it is necessary to encode various sound frequencies for reward, CA1 place cells tuned place fields to the frequency dimension (Lisman et al., 2017), allowing rats to navigate in the sound space. The HI's role in mapping sequential information is also evident from lesion studies. For example, patients with hippocampal damage showed impairments in recalling details of the sequences of ordered events and not those events per se (Lisman et al., 2017). An amnestic taxi driver with hippocampal damage failed to navigate when driving in London with a lack of detailed representation of the environment (Maguire et al., 2006). It is evident that the HI plays an important role in learning

and retrieval in spatial navigation, from physical spaces where space and time are adaptive for encoding to abstract spaces where other dimensions are crucial for task performances.

The HI also facilitates the encoding and post-encoding memory maintenance of episodic events. Rats with bilateral HI lesions indicate signs of a loss of recollection (Eichenbaum et al., 2012). Within the HI, the anterior and posterior portions represent memory on different scales over time, therefore contributing to episodic memory maintenance on different levels (Ritchey et al., 2015). In this study by Ritchey and colleagues, subjects encoded sentences with associations between items and contexts and were tested for memory on both delayed and immediate timescales. Using multi-voxel pattern similarity analysis, they showed that the left anterior HI displayed higher pattern similarity for information under the same context compared to that of different contexts. During delayed memory recollection, subjects who performed better at context associations showed a larger context similarity effect in left anterior HI. Therefore, anterior HI may be especially crucial in representing episodic context over time. On the other hand, posterior HI did not show the same effect over time, but evidence suggests that it may encode local and specific episodic details (Poppenk et al., 2013). Evensmoen et al. (2014) showed that the posterior HI had higher BOLD activations during the retrieval of specific contexts within a virtual environmental learning task. Thus, anterior HI and posterior HI may be responsible for global and local contexts in episodic memory, respectively.

The HI also exhibits post-learning dynamics that prioritize the retention of salient episodic memory. Gruber et al. (2016) used resting-state and task-based fMRI and showed that the HI supports the consolidation of memories associated with reward salience. Subjects encoded objects with different contexts that started off with different amounts of rewards.

After training a classifier that aimed to dissociate high- and low-reward contexts, they counted how often this classifier associated the resting-state activities after encoding with the high-reward contexts pattern and used this as the measurement for reactivation of high-reward context. This measurement positively correlated to an index that indicates the degree of memory advantage in a high-reward context, even after controlling for encoding and overall memory performance. In other words, episodic memory with reward was better remembered, and the HI showed such a neural indicator with its reactivation of high-reward contexts after encoding. This study supports the general idea that the HI is important in episodic memory, displaying post-encoding dynamics that adaptively prioritize memories with salience. The HI's critical role in recollecting episodic details has been demonstrated in many other studies, highlighting its ability to bind specific items with contexts with a high level of detail (Yonelinas, 2013). The importance of the HI in episodic memory is therefore evident from these studies above.

Neocortical and subcortical regions

There is abundant evidence of subcortical structures' contribution to spatial navigation. Vitally interconnected to the HI, these structures, including the anterior and rostral midline thalamic nuclei, the claustrum, and others, also contain spatially tuned cells coding for spaces, like cells within the HI. One tempting hypothesis is that different scales of spatial information are represented in different subcortical regions depending on the size of the firing field. For example, the subiculum, a part of the hippocampal formation, contains spatially-selective neurons like the HI, but with larger firing fields, and also place- and grid-like cells that are sensitive to environmental boundaries. Interestingly, the subiculum also contains neurons that

specifically code for the direction of travel of an animal. O'Mara et al. (2019) mentioned that activations in the anterior thalamus and the subiculum showed the horizontal axis of a subject when traveling in a virtual 3D environment, and the retrosplenial cortex activation reflects the vertical axis (Kim & Maguire, 2018). The claustrum, containing spatially responsive neurons, seems to track the changing positioning of objects and is sensitive to perimeters.

These subcortical regions are mostly interconnected. For example, the subiculum provides strong efferent projections to the anterior thalamic nuclei. The nucleus reuniens, also containing cells coding for the head direction, strongly project to the subiculum and CA1. The claustrum is connected to the HI formation, retrosplenial, and entorhinal cortices. These strong interconnections resonate well with the functional similarity of these subcortical structures in spatial navigation.

Neocortical regions differ from the subcortical regions in their role of spatial navigation. Specifically, neocortical structures, such as the orbitofrontal cortex (OFC), contribute to navigation in abstract task spaces compared to the HI's similar role in physical spaces. For example, in an operant decision-making task where rats must learn the task space amongst odors, actions, and outcomes, the OFC represents the state space with sequential firing to the "odor-state" and the "reward-state", potentially allowing the animal to navigate in this abstract task space (Wikenheiser et al., 2016).

Subcortical structures also play a critical role in explicit and event memory. O'Mara et al. (2019) reviewed other studies that provided evidence that the degeneration of the anterior thalamic nuclei results in diencephalic amnesia. They mentioned that patients with rostral

thalamic strokes and abnormal afferent projections to the anterior thalamic nuclei displayed changes in episodic memory performance. Item memory, which is a part of episodic memory, showed impairments from damages to the bilateral fornix, an important output tract of the HI connecting it to the other regions.

The role of other neocortical structures in episodic memory also extends to the midline default network (DN) core's functions in event representations. The DN core includes the posterior cingulate cortex (PCC) and the medial prefrontal cortex (mPFC). Stawarczyk et al. (2019) reviewed the evidence on this network and proposed that the PCC shows BOLD activations that are sensitive to event boundaries in tasks with naturalistic stimuli, such as movie-watching in scanners. At event boundaries, event models that contain multimodal representations of events integrated with sensory inputs and existing memories typically undergo transitions. Additionally, the time course of the PCC was similar across subjects when they listened to intact narratives. This between-subject similarity across time-series BOLD signals suggests the potential effect of the PCC in representing different episodes during a series of events. Representational similarity analysis (RSA) also demonstrates strong evidence for the PCC's involvement in event models. Pattern reinstatement in the PCC after encoding predicts the number of details subjects were able to recall. The evidence above shows that other neocortical regions, such as the PCC, are important in constructing, representing, and maintaining event models.

Neocortical regions such as the mPFC represent episodic contexts on a more schematic level (Stawarzyk et al., 2019), compared to subcortical structures that represent episodic memory on a spatiotemporal scale. It is suggested that in an fMRI, mPFC activations correlated

with self-specific information, whereas the PCC activations correlated with information containing locations (Martial et al., 2018).

In sum, both subcortical and neocortical regions show significance in spatial navigation and episodic memory, with neocortical regions reflecting higher-order and self-specific information and subcortical regions reflecting spatiotemporal information like the HI but with potential differences in informational scales.

Coordination between the hippocampus and other regions

The HI and neocortical regions function together in abstract spatial navigation tasks, especially for those tasks involving reward or high-level task rules. For example, the HI and prefrontal network showed coherence in theta oscillations when rats learned new rules in a Y maze (Benchenane et al., 2010). 4 Rats spatially navigated on a Y maze and learned 2 reward contingency rules: go to the right and then left arm of the maze to obtain rewards. After acquiring the rules, HI-PFC coherence increased on the theta band, and this coherence predicted the rats' performance level in the rule-dependent spatial navigation task. PFC neurons that were modulated by the HI theta after rule acquisition were significantly intercorrelated, showing changes in the dynamics of cell assemblies. This study shows that the HI and the neocortical network functionally couple during spatial navigation, which requires learning, aiding informational flow, and enhancing plasticity across regions after rule acquisition. Such a type of synchronization between cell assemblies may also be important in episodic memory formation, as theta activity is closely related to episodic memory retrieval in humans (Hsieh & Ranganath, 2014).

The HI and subcortical regions work cooperatively in episodic memory formation, with each region representing information on distinct scales, such as global or local, what or where information. Eichenbaum et al. (2012) proposed a model based on neural evidence, double dissociating the MTL network in representing “what” and “where” information. They suggested that neocortical areas provide inputs to the perirhinal cortex (PRC) and lateral entorhinal cortex (LEC), representing perceptual objects and events. In parallel, visuospatial information is sent from the parietal and retrosplenial cortex to the parahippocampal cortex (PHC) and medial entorhinal cortex. “What” and “where” streams of information then converge on the HI, as the HI binds the information together, providing a holistic representation of episodic events. Komorowski et al. (2009) conducted an interesting experiment supporting the HI’s role in binding “what” and “where” information. Rats moved between contexts with reverse reward contingency with odor stimuli. As learning progressed, more neurons became stimuli-context selective and less spatially-only selective and maintained throughout subsequent sessions. This evidence shows the binding role of the HI of object-oriented and spatial information.

Another similar model by Reagh et al. (2018) suggested that PHC and PRC play different roles in encoding episodic events. Like the “where” stream by Eichenbaum et al. (2012), PHC, along with other posterior-medial regions, may encode contextual and situational information, like scaffolds. PRC and anterior-temporal regions are thought to encode specific and local information, like the “what” stream, that is mapped onto the scaffolds with the help of the HI. PHC scales positively with memory confidence when item memory is being tested, suggesting its relevance in local information. This model is built upon the evidence suggesting PRC’s specificity to items and objects and PHC’s common representations of scene and spatial content

(Diana et al., 2007). Intrinsic functional connectivity within the PMAT network further suggests that these regions couple with each other during encoding.

Differences between PRC and PHC in episodic encoding are also shown in a study by Ritchey et al. (2015). In a delayed recollection task, PRC and PHC activations both correlated with delayed recognition over a day, and PRC not PHC was sensitive to context information. This resonates with the general idea that the PRC may work along with the HI to support item and local contextual information. However, we cannot say that PHC does not play a role in item memory. Diana et al. (2008) suggested that PHC's activation successfully classified individual object categories and is not exclusively sensitive to scenes. Therefore, PHC is not purely selective for spatial information in episodic memory.

The evidence above shows that the hippocampus, neocortical, and subcortical regions are all important in spatial navigation and episodic memory. These regions strongly interconnect structurally and functionally, which reflects their complementary and cooperative roles in representing spatiotemporal and contextual information. With a closer look, each region displays unique roles in forming and maintaining these representations, varying by different scales that are tailored to relevant dimensions of each task. Cooperation between the hippocampus and these regions enables a holistic and efficient representation over time. A potential future direction may be to elucidate how neural coupling among these regions could depend on the time available for a task (short vs. long). If a task requires fast actions, would these regions still function cooperatively, or would the hippocampus work independently for fast adaptive behaviors in navigation? Overall, interactions between the hippocampus and other regions remain as topics of continuing interest.

References

Aronov, D., Nevers, R., & Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature*, *543*(7647), 719-722.

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. Neuron, 66(6), 921-936.

Cave, C. B., & Squire, L. R. (1991). Equivalent impairment of spatial and nonspatial memory following damage to the human hippocampus. *Hippocampus*, *1*(3), 329-340.

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2008). High-resolution multi-voxel pattern analysis of category selectivity in the medial temporal lobes. Hippocampus, 18(6), 536-541.

Eichenbaum, H., Sauvage, M., Fortin, N., Komorowski, R., & Lipton, P. (2012). Towards a functional organization of episodic memory in the medial temporal lobe. Neuroscience & Biobehavioral Reviews, 36(7), 1597-1608.

Gruber, M. J., Ritchey, M., Wang, S. F., Doss, M. K., & Ranganath, C. (2016). Post-learning hippocampal dynamics promote preferential retention of rewarding events. Neuron, 89(5), 1110-1120.

Hsieh, L. T., & Ranganath, C. (2014). Frontal midline theta oscillations during working memory maintenance and episodic encoding and retrieval. Neuroimage, 85, 721-729.

Jenkins, L. J., & Ranganath, C. (2010). Prefrontal and medial temporal lobe activity at encoding predicts temporal context memory. Journal of Neuroscience, 30(46), 15558-15565.

Lisman, J., Buzsáki, G., Eichenbaum, H., Nadel, L., Ranganath, C., & Redish, A. D. (2017). Viewpoints: how the hippocampus contributes to memory, navigation and cognition.

Kim, M., & Maguire, E. A. (2019). Encoding of 3D head direction information in the human brain. Hippocampus, 29(7), 619-629.

Knudsen, E. B., & Wallis, J. D. (2021). Hippocampal neurons construct a map of an abstract value space. *Cell*, *184*(18), 4640-4650.

Komorowski, R. W., Manns, J. R., & Eichenbaum, H. (2009). Robust conjunctive item–place coding by hippocampal neurons parallels learning what happens where. Journal of Neuroscience, 29(31), 9918-9929.

Maguire, E. A., Woollett, K., & Spiers, H. J. (2006). London taxi drivers and bus drivers: a structural MRI and neuropsychological analysis. Hippocampus, 16(12), 1091-1101.

Martial, C., Stawarczyk, D., & D'Argembeau, A. (2018). Neural correlates of context-independent and context-dependent self-knowledge. Brain and cognition, 125, 23-31.

McKenzie, S., Robinson, N. T., Herrera, L., Churchill, J. C., & Eichenbaum, H. (2013). Learning causes reorganization of neuronal firing patterns to represent related experiences within a hippocampal schema. Journal of Neuroscience, 33(25), 10243-10256.

Naya, Y., & Suzuki, W. A. (2011). Integrating what and when across the primate medial temporal lobe. Science, 333(6043), 773-776.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain research*.

O'keefe, J., & Nadel, L. The Hippocampus as a Cognitive Map (Oxford University Press).

O'Mara, S. M., & Aggleton, J. (2019). Space and Memory (Far) Beyond the Hippocampus: Many Subcortical Structures Also Support Cognitive Mapping and Mnemonic Processing. Frontiers in neural circuits, 13, 52.

Poppenk, J., Evensmoen, H. R., Moscovitch, M., & Nadel, L. (2013). Long-axis specialization of the human hippocampus. Trends in cognitive sciences, 17(5), 230-240.

Reagh, Z. M., & Ranganath, C. (2018). What does the functional organization of cortico-hippocampal networks tell us about the functional organization of memory?. Neuroscience letters, 680, 69-76.

Ritchey, M., Libby, L. A., & Ranganath, C. (2015). Cortico-hippocampal systems involved in memory and cognition: the PMAT framework. In Progress in brain research (Vol. 219, pp. 45-64). Elsevier.

Ritchey, M., Montchal, M. E., Yonelinas, A. P., & Ranganath, C. (2015). Delay-dependent contributions of medial temporal lobe regions to episodic memory retrieval. Elife, 4, e05025.

Stawarczyk, D., Bezdek, M. A., & Zacks, J. M. (2019). Event Representations and Predictive Processing: The Role of the Midline Default Network Core. Topics in cognitive science.

Sun, C., Yang, W., Martin, J., & Tonegawa, S. (2020). Hippocampal neurons represent events as transferable units of experience. *Nature neuroscience*, *23*(5), 651-663.

Teng, E., & Squire, L. R. (1999). Memory for places learned long ago is intact after hippocampal damage. *Nature*, *400*(6745), 675-677.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological review*, *55*(4), 189.

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.

Yonelinas, A. P. (2013). The hippocampus supports high-resolution binding in the service of perception, working memory and long-term memory. Behavioural brain research, 254, 34-44.

What is hippocampal replay?

During non-exploratory states such as sleeping and resting when animals are awake, the hippocampal network displays brief neuronal sequences named “replays”, which are thought to code for the spatiotemporal paths that animals experienced. These brief neuronal events take the form of ripples, or oscillatory patterns, characterized by the high-frequency (150-300 Hz) and briefness (50-100ms) (Pfeiffer, 2018). These sharp-wave ripples are found in both the hippocampus (HI) and other inter-connected regions, including the medial entorhinal cortex (O’Neil et al., 2017). Ripples are temporally compressed events and are thought to represent behaviorally relevant information (Pfeiffer, 2018). Replay could take place in simple navigational tasks for rats, such as running on a linear track or an open field. Using decoding methods, it could be shown that ripples show the replay of trajectories that rats just went through (Pfeiffer & Foster, 2015).

Replays can take various forms, ranging from forward-replay, reverse-ordered “replays” predicting animals’ future actions, or simulations of mental trajectories that have not been experienced (Pfeiffer, 2018). Therefore, replays do not only represent prior paths that animals have taken but could also reflect information that is currently being processed for planning. In the following section, I will provide evidence in rodent studies on replays during sleep, awake states, and “replays” for future planning behaviors.

Replay evidence in rodents

Replay is not a newly discovered hippocampal phenomenon. Early studies already indicated the first evidence of replay during animals’ sleep. In the review article, Pfeiffer (2018) mentioned

that an early study found that highly active hippocampal neurons during spatial exploration showed higher activities during the following sleep session compared to control hippocampal neurons (Pavlides & Winson, 1989). Ripples may be neuronal markers for reactivation, as an early study showed high correlations between hippocampal neurons during exploration and subsequent ripples compared to other non-ripple events during sleep (Wilson & McNaughton, 1994). Ripples during sleep preserved the temporal order of spatial sequences rats experienced (Pfeiffer, 2018).

Replays could also occur when rodents are awake and experiencing brief rests in an exploratory environment. These replays could be both forward and reverse in the temporal direction that rodents initially experienced in prior activities (Pfeiffer, 2018). As Pfeiffer (2018) mentioned, reverse replays could occur when an animal travels on a track only once, without repetitions. Interestingly, the rate of reverse replays but not forward replays increase with higher rewards (Ambrose et al., 2016), suggesting a potential relationship between reverse replays and credit assignment. Awake replays last long after rodents' original spatial exploratory experience in W-shaped environments in a high-walled rest box, and this type of remote awake replay is found to reflect information on coherent remote trajectories (Karlsson & Frank, 2009). This suggests that awake replays could take place in delayed exploratory pauses during rodents' navigation, with the initial spatial information well preserved over time.

Replay, especially forward replay, contains learned and planned information on the spatial representations of an environment rather than a simple remapping of the experience. Thus, the word "replay" does not necessarily suggest that the information encoded has already been experienced. In a study by Gupta et al. (2010), rats experienced two different behavioral

sequences within a maze, neither forward nor backward replays encoded highly experienced sequences. Surprisingly, they found that a significant number of replays encoded novel paths that the rats had not experienced. These novel paths represented efficient trajectories that could help the rats navigate the maze. Furthermore, replays of a previous sequence occurred while rats were currently going through the other sequence even more than 10 minutes, reflecting the learned information of the previous sequence. Therefore, replays could reflect learned information, potentially as a mechanism for memory consolidation. Also, replays could encode information related to the construction of cognitive maps, helping rodents to plan in the environment. A potential replay candidate for planning is the forward replay, which often occurs before an animal's movement (Pfeiffer, 2018). Forward replay may encode content predicting an animal's future actions, as Singer et al. (2013) found the information of forward replay prior to a spatial decision in a two-choice maze task is more likely to encode the right path following an animal's learning. However, forward replay does not deterministically predict goal-directed future behaviors (Pfeiffer, 2018).

In sum, existing rodent literature provides evidence of replay during sleep and awake states. Replay could encode temporally compressed information on trajectories that rodents experienced or learned, as well as potential future trajectories, which could be part of animals' cognitive maps in planning.

Replay and memory consolidation

As the rodent replay literature suggests, replay serves as a neural mechanism in consolidating already learned spatial trajectories and disruption of hippocampal ripples could lead to losses in

memory (van de Ven et al., 2016; Pfeiffer, 2018). Some studies found evidence in favor of reverse replay in memory consolidation (Pfeiffer, 2018). The critical role of replay in memory consolidation is also evident in humans, according to a study by Schapiro et al. (2018). Interestingly, their results demonstrate that replay in the human HI prioritizes weakly learned information and is predictable of subsequent memory performance and memory improvement only after sleep. Similar results of replay prioritization for less often encountered information are also found in a rodent study (Gupta et al., 2010). Therefore, replay of previously experienced events may consolidate memories but with different weights depending on how well the information is learned, or even perhaps how necessary the information needs to be consolidated for future performance.

To elaborate on the study by Schapiro et al. (2018), subjects learned the features of 15 objects in 3 categories, with each category's stimuli sharing some common features. Neural responses generated by each object were measured in the scanner. They also used pattern analysis to assess if the objects were replayed by in the HI during a resting period in the scanner. The effect of sleep was controlled by having a group of subjects who slept between the learning and testing phases and another group awake. Replay events were defined as the significant correlations between resting period activity and the multivoxel templates for each object in the HI during object encoding. A negative correlation was shown between memory and subsequent replay, meaning replay was the strongest for weak memory. A positive correlation was shown between the amount of replay and subsequent memory performance only for those subjects who slept between learning and the subsequent memory test. To ensure the validity of the results, they scrambled the labels of objects during multivoxel template

generation of the objects, therefore resulting in meaningless templates. As a result, significantly less replay was detected. This suggests that replay activities that were originally measured reflect the reinstated representations of specific objects.

I think a potential logical caveat of this study may be perhaps additional replay during sleep was what caused the memory to consolidate instead of the original replay events before sleep. Because they did not report replay events during sleep, and replay during sleep is common, it is hard to dissociate awake replay from replay during sleep. It is also possible that memory consolidation results from both replay during sleep and the effect of sleep after awake replay. Future studies could dissociate these two effects and their separate contributions to memory consolidation, as well as the contributions of reverse and forward replay in humans.

Motivational salience and replay

Besides memory strengths during encoding, other motivational or salient dimensions, such as reward or fear, could also change replay dynamics. For example, the rate of ripples in the rats' CA3 was significantly higher at rewarding sites versus non-rewarding sites (Singer & Frank, 2009). One possibility is that through hippocampal replay, new information on reward is bound to existent knowledge (e.g. spatial layout) to generate adaptive goal-directed future behaviors. This hypothesis also resonates with the HI's function in high-resolution binding (Yonelinas, 2013).

Replay quantity seems to favor trajectories leading to rewards in navigational tasks. In goal-directed navigation within an open arena, rats had to learn the daily changing locations of the goal along with the starting position. After rats experienced the recently learned goal

location, replay prioritized spatial paths from the rat's current location to the goal location (Pfeiffer & Foster, 2013). Thus, in a reward-seeking situation, replay enables the binding of existing knowledge with newly learned salient information for adaptive behavior. This study also shows that the replayed events significantly correlated with the rat's subsequent behaviors, supporting the idea that replays reflect mental simulations of future goal-directed actions.

Given the evidence above, replay is very likely to be captured in the model-free reinforcement learning models, such as the temporal difference model which emphasizes that the reinforcement of rewarded information. In the temporal difference model, spiny neurons in the striatum learn the values, and the basal ganglia are thought to be initiating actions based on the striatal value signals. Replay events in the HI might be responsible for updating the values of different trajectories/states through coupling with the striatum. The HI replay events may also functionally connect to the orbitofrontal cortex in updating the stimulus-specific value representation. As for model-based reinforcement learning, replays that reflect the future trajectories in planning instead of repeating occurred sequences might play a crucial role. This type of replay, or "preplay" if the term "replay" seems misleading here, would facilitate the generation of cognitive maps in the hippocampus . In addition, transition structures and reward states are learned in model-based learning, which involves the prefrontal cortex and the striatum. It would be interesting to find out how different types of replay (forward/reverse) might couple differentially with the prefrontal regions in learning the transition structures in an environment with changing statistics. Another interesting direction would be to test if

hippocampal replay acts as an intermediate hub between the prefrontal cortex and sensory cortices, binding incoming sensory information with the existing knowledge of task structures.

References

Ambrose, R. E., Pfeiffer, B. E., & Foster, D. J. (2016). Reverse replay of hippocampal place cells is uniquely modulated by changing reward. *Neuron*, *91*(5), 1124-1136.

Karlsson, M. P., & Frank, L. M. (2009). Awake replay of remote experiences in the hippocampus. Nature neuroscience, 12(7), 913-918.

O'Neill, J., Boccara, C. N., Stella, F., Schönenberger, P., & Csicsvari, J. (2017). Superficial layers of the medial entorhinal cortex replay independently of the hippocampus. Science, 355(6321), 184-188.

Pavlides, C., & Winson, J. (1989). Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. Journal of Neuroscience, 9(8), 2907-2918.

Pfeiffer, B. E., & Foster, D. J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. Nature, 497(7447), 74.

Pfeiffer, B. E., & Foster, D. J. (2015). Autoassociative dynamics in the generation of sequences of hippocampal place cells. Science, 349(6244), 180-183.

Pfeiffer, B. E. (2018). The content of hippocampal "replay". Hippocampus.

Schapiro, A. C., McDevitt, E. A., Rogers, T. T., Mednick, S. C., & Norman, K. A. (2018). Human hippocampal replay during rest prioritizes weakly learned information and predicts memory performance. Nature communications, 9(1), 3920.

Singer, A. C., & Frank, L. M. (2009). Rewarded outcomes enhance reactivation of experience in the hippocampus. Neuron, 64(6), 910-921.

Singer, A. C., Carr, M. F., Karlsson, M. P., & Frank, L. M. (2013). Hippocampal SWR activity predicts correct decisions during the initial learning of an alternation task. Neuron, 77(6), 1163–1173.

van de Ven, G. M., Trouche, S., McNamara, C. G., Allen, K., & Dupret, D. (2016). Hippocampal offline reactivation consolidates recently formed cell assembly patterns during sharp wave-ripples. Neuron, 92(5), 968-974.

Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. Science, 265(5172), 676-679.

Yonelinas, A. P. (2013). The hippocampus supports high-resolution binding in the service of perception, working memory and long-term memory. Behavioural brain research, 254, 34-44.

The hippocampus in configural and conjunctive reinforcement learning

Because of its critical functions in pattern separation and vital connections to the striatum, the hippocampus (HI) enables configural reinforcement learning (RL). The HI seems to encode configural stimuli distinctively from the elemental stimuli with common features, enabling a clear representation of stimuli with complex forms. This representation may then be sent to the striatal regions for RL value updating. Furthermore, the HI also seems to distinctively represent other task-relevant dimensions, such as values, preserving the uniqueness of stimuli of a similar dimension of interest, thereby aiding RL in subsequent processes of these stimuli.

Humans can process causal stimuli in both elemental and configural forms (Duncan et al., 2018). Configurations of different cues sometimes combine as a single unit that leads to outcomes. Therefore, in a statistical environment where causal events have complexity in forms, it is important for humans to learn configural cues. In this study, Duncal et al. (2018) found that the HI showed predictive BOLD activity for configural cues, and this activity is correlated to the configural learning performance. Specifically, subjects learned to predict weather outcomes based on elemental and configural visual cues with predetermined outcome statistics. They performed this task under 2 different conditions: “inseparable” and “separable” conditions. In the inseparable condition, weather outcomes could only be predicted by the configural elements, whereas in the separable condition, both elemental and configural cues were informative of weather outcomes. Using a reinforcement learning model, they quantified the relative weights of configural and elemental learning in contribution to every choice the subjects made and defined the degree of configural learning as the “configural choice index”. The anterior HI had more within-subject activation when choices had a higher “configural

choice index", and this activation also positively predicts the level of between-subject configural learning. This result is in line with the idea that the anterior HI supports global context encoding (Ritchey et al., 2015), and indicates that the hippocampal activity is a neural mechanism for configural learning. The HI responses at choice showed higher unidirectional coupling with the nucleus accumbens during outcomes in the inseparable task compared to the separable task, and this activation is higher for subjects who relied more on higher configural learning. Such a functional coupling between the HI and the striatum suggests that these regions cooperate to support configural RL, possibly with the HI sending inputs to the striatum to update the value of configural stimuli. The HI is likely to represent configurations differentially from elemental stimuli, and supports distinctive representations of stimuli taking various forms, namely, pattern separation. Clear-cut representations in the HI reduce ambiguity among perceptually similar stimuli and allow subsequent stimulus-outcome associations to occur with better precision, therefore facilitates configural learning.

Another study by Ballard et al. (2019) also found similar evidence supporting the dual contribution of the hippocampal pattern separation and connections with the striatum to configural reinforcement learning. Subjects did a non-spatial, probabilistic stimulus-response learning task including elemental and conjunctive stimuli with overlapping features: AB+, B-, AC-, AC+, +/- meaning reward or no reward. Behavioral modeling showed the spread of value among stimuli sharing the same features. Specifically, for any given value update to a conjunctive stimulus (AB), about 44% of that update was also applied to a stimulus (B) within that conjunctive stimulus. If pattern separation takes place in the HI, then representations for stimuli that shared features (AB vs. B) should be more separated in HI and, thus, less correlated

with each other. Pattern similarity analysis shows that the HI is the only region with low similarity in representing stimuli with overlapping features. All the other tested regions, including the parahippocampal and perirhinal cortex, the inferior frontal sulcus, and the medial orbitofrontal cortex, had similar representations of stimuli with overlapping features. This strongly supports HI's role in disambiguating a configural stimulus as a unique unit, distinctive from the perceptual features that it is made up of. In the supplementary analyses, they analyzed each run's hippocampal activation for stimulus with overlapping features and correlated with striatal prediction error for configural learning. Interestingly, the more distinctive the HI-represented stimuli with overlapping features, the higher the striatal prediction errors for configural learning. Given that the striatal prediction error is a reliable neural correlate for learning signals, this result reflects that the hippocampal pattern separation aids configural learning. The HI separates not only patterns in perceptual features, but also in values. Pattern similarity analysis showed that stimuli with similar outcome associations were represented more differentially in HI compared to those with dissimilar outcome associations, suggesting that the HI represents stimuli with similar values distinctively. Since the subjects had to learn the values of each stimulus for better task performance, value is an important space besides the feature space that may require precise representations. Therefore, the hippocampal pattern separation may target specific task-relevant dimensions to guide RL behaviors.

The hippocampus learns both statistics and the structure of an environment

The hippocampus facilitates both statistical learning and learning of the environmental structure. In a study by Vilà-Balló et al. (2017), the gray matter of the HI was measured and

correlated with the performance of a reversal-learning task. They hypothesized that the HI may be necessary to generate anticipation of reversal occurrences. Subjects had to choose a square on the left or another square on the right of the screen, and then received feedback on whether the choice was correct or not. The probability of one side being the correct choice was constant for some trials, followed by reversals. Correct choices are defined as responses to the side with the higher probability. Behavioral results showed that subjects showed anticipatory switching of choices followed by spurious negative (made a correct choice but received negative feedback because of the predetermined probabilities) and incorrect negative (negative feedback with the current incorrect action) feedback. Over time, subjects learned to predict reversal occurrence through experience, as shown by the increasing of choice switches after a spurious and not incorrect feedback. To quantify this anticipatory switching behavior developed over trials, they did a principal component analysis, including differences in choice switches between early and late trials following spurious and incorrect negative feedback. Hippocampal sclerosis patients with less hippocampal grey matter than controls exhibited less anticipatory switching behavior. Therefore, the hippocampal grey matter seems crucial in learning the high-order statistics of the environment.

Besides its vital role in learning task statistics, the HI also rapidly learns the general structure of a task. Schapiro et al. (2015) designed a task paradigm with high-order temporal structures but no variance in transition probabilities and tested if the HI is sensitive to this form of temporal community structure. To elaborate, subjects experienced sequences of stimuli generated from a random walk with 3 communities, with each community defined by a cluster of nodes with many interconnections among themselves. Each node in this abstract structural

space was assigned a visual stimulus in the task. Stimulus-to-stimulus transition probabilities were constant for all stimuli regardless of which community a stimulus belongs to. Stimuli within a community, because of their large interconnections, tend to connect to one another and not to stimuli in other clusters. Using representational similarity analysis, they found the bilateral HI to have greater within-community than across-community similarity. Other MTL regions did not show the same effect as the HI. Using GLM, the beta weights for the regressor that coded for community boundaries were computed, and the overall level of hippocampal activity was sensitive to community boundaries. Specifically, the HI was more active during presentations of within-community nodes and was deactivated at community boundaries. In sum, HI is sensitive to higher-order structural information of a task space.

Regions beyond the hippocampus: the medial temporal lobe in detecting regularities across time.

Besides the hippocampus, the medial temporal lobe (MTL) as a whole is involved in encoding or retrieval regularities, aiding the aggregation across experiences and extracting systematic information. In a study by Schapiro et al. (2014), controls and a patient named LSJ who had bilateral hippocampal losses and broader MTL damage did a task in which a continuous sequence of shapes, syllables, scenes, or tones with temporal regularities were presented. Critically, these stimuli were divided into triplets, and items within a triplet always appeared in a consistent and consecutive order, but the order of the triplets was random. After viewing these stimuli, subjects were judged for the familiarity of a presented triplet with a foil. Compared to the controls who rated the presented triplets with a higher familiarity score, patient LSJ did not show evidence of learning the regularities. The level of task difficulty was

reduced in the subsequent experiment. Instead of triplets, simpler regularities were presented as stimulus pairs. However, patient LSJ still was not able to learn the regularities. Although the MTL as a whole is essential for the detection of regularities throughout time when there is no segmentations or trial structures, it is important to note that the study did not dissociate the HI from the other MTL regions for their separate contributions in detecting regularities.

Conclusion

The evidence above highlights the importance of the hippocampus in configural reinforcement learning and the learning of environmental statistics and temporal structures. The MTL, in general, is also important in detecting regularities across time. An interesting direction for future studies would be to investigate the functional connectivity between the MTL and other regions involved in RL, such as the prefrontal cortex. In my opinion, the distinctive signals of configural stimuli and stimuli with similar values processed by the hippocampus might project to the orbitofrontal cortex (OFC), which has been shown to encode stimulus-specific value representations. This type of high-resolution hippocampal inputs may offer highly detailed information and is thus essential for the OFC's detailed stimulus-to-value representation.

References

Ballard, I. C., Wagner, A. D., & McClure, S. M. (2019). Hippocampal pattern separation supports reinforcement learning. Nature communications, 10(1), 1073.

Duncan, K., Doll, B. B., Daw, N. D., & Shohamy, D. (2018). More than the sum of its parts: a role for the hippocampus in configural reinforcement learning. Neuron, 98(3), 645-657.

Ritchey, M., Montchal, M. E., Yonelinas, A. P., & Ranganath, C. (2015). Delay-dependent contributions of medial temporal lobe regions to episodic memory retrieval. Elife, 4, e05025.

Schapiro, A. C., Turk-Browne, N. B., Norman, K. A., & Botvinick, M. M. (2016). Statistical learning of temporal community structure in the hippocampus. Hippocampus, 26(1), 3-8.

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. Journal of cognitive neuroscience, 26(8), 1736-1747.

Vilà-Balló, A., Mas-Herrero, E., Ripollés, P., Simó, M., Miró, J., Cucurell, D., ... & Rodríguez-Fornells, A. (2017). Unraveling the role of the hippocampus in reversal learning. Journal of Neuroscience, 37(28), 6686-6697.

The orbitofrontal cortex in contingent learning

The orbitofrontal cortex encodes reward contingency

The orbitofrontal cortex (OFC) has been shown to encode contingent relationships between stimuli and outcomes in both lesion studies (Walton et al., 2010; Noonan et al., 2017) and recording studies (Stalnaker et al., 2014; Jocham et al., 2016). Animals with OFC lesions could still approximate the link between choice and reward after a change in reward contingency, but at a much slower rate with the use of alternative learning mechanisms such as integrating recent choice and reward history. Without a functioning OFC, statistical relationships in a task with changing statistics could not be updated and learned at a normal speed.

One of the studies that supports this idea had macaque monkeys with OFC lesions perform different versions of a three-armed bandit task (Walton et al., 2010). The reward statistics of the 3 stimuli slowly drifted throughout the trials, and in the middle of the experiment, the most rewarding stimulus reversed. Before a reversal, learning was intact for the OFC lesioned group, as the lesioned monkeys and controls both tracked the highest rewarded stimulus with comparable learning speeds. However, after a reversal, the likelihood of choosing the highest rewarded option decreased significantly for the OFC lesioned group compared to that of the controls. Multiple regression analysis showed that after OFC lesions, the effect of the previous choice-outcome association on predicting the current choice was significantly reduced. Credit assignment for an outcome was not attributed to the most recent contingent choice but instead to an integrated choice history over many trials. For example, if B is the most rewarded choice before a reversal, and A is the newly rewarded choice, OFC lesioned monkeys tended to choose A more often following a reward on B, because of a longer

reward history of B than A. Similarly, the lesion group was less likely to choose A after seeing B with a negative outcome. With an impaired ability to rapidly learn new contingent stimulus-outcome relationships, the OFC lesioned monkeys relied on recent choice history instead.

Besides making a current choice by integrating choice history, other non-contingent heuristic learning mechanisms also exist and operate in parallel with distinct neural circuitries compared to the contingent learning circuitry, which involves the OFC, reflecting a distinct role of the OFC in forming and updating causal relationships. Jocham et al. (2016) found several non-contingent learning mechanisms in human subjects, while the OFC's role in developing contingent associations dominated the subjects' behaviors. In the first experiment, participants were presented with trials of stimuli of different shapes, and choose whether to select a stimulus, then received contingent rewards 3s after a choice. Noncontingent rewards, distinguished in color from contingent rewards, were also presented independent of the subjects' behaviors. Lateral OFC (lOFC) showed selective responses at the time of reward to contingent but not noncontingent rewards. Furthermore, this lOFC reward signal showed a positive correlation with the contingent learning effect and a negative correlation with two noncontingent learning effects, including the spread of effect to proximal choices (recent reward biases recent but noncontingent choices) and the spread of effect to recent frequent choices (reward biases frequently selected choices in choice history). In the first experiment, they showed the natural occurrence of both contingent and noncontingent learning, but in the second experiment, they selectively controlled for different learning strategies by delivering 3 different instructions for each block. On each trial, subjects chose between 2 fractals, and in condition 1) received an outcome on the next trial, condition 2) after a known number of trials

later, condition 3) after a random unknown number of trials. Critically, all 3 runs had the same type of random reward schedule regardless of the instructions. Subjects deployed contingent learning in condition 1) and 2), but not condition 3). The OFC signals were precisely aligned with the contingencies of both conditions 1) and 2), correlating with the subtraction of the BOLD response of noncontingent sequences from contingent sequences. With the consistent results from both experiments, it is evident that lOFC facilitates the learning of choice-outcome contingencies through the use of a task model.

The OFC encodes task states

Besides its crucial role in representing reward contingency, the OFC has also been suggested as a region that supports the representation of one's current position within a task state (Wilson et al., 2014; Takahashi et al., 2011; Miller et al., 2017). As the only prefrontal region that receives sensory inputs from almost all modalities (Price, 2007), the OFC could provide an abstraction of the multisensory information relevant to the task. A task state usually refers to the underlying structure of an environment of systematic changes or a particular location on a "cognitive map" of the task. Some states may look perceptually similar but are indeed conceptually distinct. For example, to an experienced dealer, a period when the stock market is filled with low-priced stocks might be a state of true financial depression, or a state of temporary price fluctuations instead. Thus, sometimes it is hard to tell which state we are in from perceptual information alone, and the neural region that enables state distinction is therefore important. Several studies found results supporting the OFC's role in disambiguating different states and facilitating goal-directed learning.

Wilson et al. (2014) simulated several different experimental results by modeling an agent with and without the ability to distinguish among states. They were able to replicate several behavioral results of several previous experiments with OFC-lesioned animals, suggesting a causal role of the OFC in representing task states. For example, they were able to simulate the results of a reversal learning task with OFC-lesioned animals (Butter, 1969). Specifically, monkeys had to press a left or right plaque to receive food, and there was only one correct location on every block, but the location changed after the monkeys reached a 90% performance. Before a reversal, the performance of OFC-lesioned monkeys was comparable to the controls. However, after a reversal, the OFC-lesioned monkeys showed impaired learning in finding the new correct plaque compared to controls. Wilson et al. (2014) used a simple Q-learning algorithm to model the action values in 2 different states for the control group. In other words, the model assumes the control groups had two different states: state 1 choosing the “right” plaque yielding reward and the “left” plaque with no reward, and state 2, with the reversed contingencies. On the contrary, the OFC-lesioned group was modeled with a single state: with no defined reward contingencies for either plaque. This simple model under 2 different assumptions of states replicated the results in the original study by Butter (1969). The OFC-lesioned group took much longer than the control group to learn the reversal because, with a single state, it had to rely on the integration of recent choice-reward history. Using a similar approach, they also replicated the behavioral data of OFC-lesioned animals in other tasks, including delayed alteration, extinction, and devaluation, providing a strong theoretical account for the OFC’s role in representing task state.

A study by Takahashi et al. (2011) also highlighted the role of OFC in providing state-wise information to the ventral tegmental area (VTA) where value is computed. In their task, rats sought rewards in an odor-guided choice task where unexpected rewards were delivered for amount and time. Using computational modeling, they found that the best model replicating the dopamine firing patterns in the VTA was under the assumption that OFC provided inputs to the VTA on state identity. Other models, including one that assumed OFC might convey expected future reward throughout time, underperformed compared to this model. This demonstrates that the OFC lesion might not impair the value learning system per se but would result in losses of state information instead.

The OFC encodes value of specific items

Another major theory of OFC function is that OFC encodes value (Padoa-Schioppa & Assad, 2006; Padoa-Schioppa & Assad, 2008). However, other studies implied that the firing patterns of OFC neurons might not be a value signal only but reflect other variables as such uncertainty (Kepecs et al., 2008; Wilson et al., 2014). The study mentioned above by Takahashi et al. (2011) also indicated that OFC lesions did not affect value computations of dopamine prediction errors in the VTA. Therefore, it is doubtful to claim that OFC encodes all types of values. Nevertheless, according to several lines of evidence, OFC may encode a specific type of value: item-specific values (Kahnt et al., 2014; Howard et al., 2015). This theory also supports the idea that OFC represents contingency because stimulus-outcome relationship needs to be item-specific.

In the study by Kahnt et al. (2014), subjects did a simple outcome-prediction task with two sets of visual cues deterministically predicting the gain or loss of a small or large amount of

money (0.50/5). Using a searchlight decoding approach, they found that the central OFC carried significant information on the predicted value of cues. They further tested whether these signals contain not only the sign but also the magnitude of the outcome. Decoding analysis using a support vector regression (trained on values cues in set I and tested on set II) showed a high accuracy on OFC signals in encoding values for both sign and magnitude. To test if the same neural code is shared between appetitive and aversive cues (same multivoxel pattern for $0.5&$5 and -$0.5&$5), they trained a support vector machine and decoded cross-valence values. Interestingly, the central OFC was invariant to the valence of the expected outcome. Howard et al. (2015) implemented a similar paradigm with food odors as reward and found that OFC, but not the vmPFC, encoded item-specific reward. Therefore, OFC plays an important role in encoding item-specific values for both appetitive and aversive stimuli and represents values from multiple sensory modalities on a common scale.

OFC subdivisions

Several studies reported evidence of the lateral OFC's (lOFC) role in representing reward contingencies (Jocham et al., 2016; Noonan et al., 2017). Given the strong projections of the central OFC (mOFC) to the striatum (Lehman et al., 2011), mOFC plays a more vital role in value-related processes. mOFC-lesioned human subjects with intact lOFC were still able to correctly assign credits to stimuli in a probabilistic 3-choice decision-making task but were more distracted by irrelevant alternative third options, suggesting that mOFC exerts cognitive control during value comparison. mOFC has also been reported to represent stimulus value for both elemental and configural stimuli (Ballard et al., 2019). In monkeys, lOFC and mOFC also seem to play similarly dissociable roles in learning (Noonan et al., 2010). lOFC-lesioned monkeys showed

impaired credit assignment, and mOFC-lesioned monkeys were more likely to choose the second-best choice because of being distracted by a third irrelevant option (Noonan et al., 2010).

mOFC and lOFC are also dissociable for the types of saliences they each process. In a task by Weber et al. (2018), subjects rated how much they wanted and liked each visual stimulus object. mOFC activation significantly correlated with how much subjects wanted an item, and lateral/central OFC was associated with how much an item was liked. Both motivational (wanted) and hedonic (liked) signals from the OFC converged in the striatum for value computation.

Conclusion

OFC is a highly intricate prefrontal region with a wide range of cognitive functions, including representing task state, and encoding stimulus-outcome contingencies and item-specific values. Among subdivisions of OFC, lOFC damage results in impaired credit assignment, and mOFC damage results in abnormal value comparison. An interesting direction for future studies would be to further examine the relationship between OFC and the hippocampus, another highly sophisticated region that is found to sequentially represent spatiotemporal regularities and other dimensions (values) relevant to task performance. One hypothesis is that the hippocampus rapidly binds salient information to item and sends inputs to the OFC to form stimulus-outcome/stimulus-stimulus contingencies and value representations. The OFC would also project inputs regarding motivational or hedonic values of stimuli to the hippocampus for a value-based formation of a cognitive map.

References

Ballard, I. C., Wagner, A. D., & McClure, S. M. (2019). Hippocampal pattern separation supports reinforcement learning. Nature communications, 10(1), 1073.

Butter, C. M. (1969). Perseveration in extinction and in discrimination reversal tasks following selective frontal ablations in Macaca mulatta. Physiology & Behavior, 4(2), 163-171.

Jocham G, Brodersen KH, Constantinescu AO, Kahn MC, Ianni AM, Walton ME, Rushworth MF, Behrens TE. (2016). Reward-Guided Learning with and without Causal Attribution. Neuron 6;90(1):177-90.

Howard, J. D., Gottfried, J. A., Tobler, P. N., & Kahnt, T. (2015). Identity-specific coding of future rewards in the human orbitofrontal cortex. Proceedings of the National Academy of Sciences, 112(16), 5195-5200.

Lehman JF, Greenberg BD, McIntyre CC, Rasmussen SA, Haber SN. (2011). Rules ventral prefrontal cortical axons use to reach their targets: implications for diffusion tensor imaging tractography and deep brain stimulation for psychiatric illness. Journal of Neuroscience 31(28):10392-402.

Kahnt, T., Park, S. Q., Haynes, J. D., & Tobler, P. N. (2014). Disentangling neural representations of value and salience in the human brain. Proceedings of the National Academy of Sciences, 111(13), 5000-5005.

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. Nature, 455 (7210), 227.

Miller, K. J., Botvinick, M. M., & Brody, C. D. (2017). Dorsal hippocampus contributes to model-based planning. Nature neuroscience, 20(9), 1269.

Noonan, M. P., Walton, M. E., Behrens, T. E. J., Sallet, J., Buckley, M. J., & Rushworth, M. F. S. (2010). Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. Proceedings of the National Academy of Sciences, 107(47), 20547-20552.

Noonan MP, Chau BKH, Rushworth MFS, Fellows LK. (2017). Contrasting Effects of Medial and Lateral Orbitofrontal Cortex Lesions on Credit Assignment and Decision-Making in Humans. J Neuroscience 19;37(29):7023-7035.

Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. Nature, 441(7090), 223.

Padoa-Schioppa, C., & Assad, J. A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. Nature neuroscience, 11(1), 95.

Price JL (2007) Definition of the orbital cortex in relation to specific connections with limbic and visceral structures and other cortical regions. Ann N Y Acad Sci 1121:54–71

Stalnaker, T. A., Cooch, N. K., McDannald, M. A., Liu, T. L., Wied, H., & Schoenbaum, G. (2014). Orbitofrontal neurons infer the value and identity of predicted outcomes. Nature communications, 5, 3926.

Takahashi YK, Roesch MR, Wilson RC, 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):1590-7.

Walton ME, Behrens TE, Buckley MJ, Rudebeck PH, Rushworth MF. (2010) Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. Neuron 65(6):927-39.

Weber, S. C., Kahnt, T., Quednow, B. B., & Tobler, P. N. (2018). Frontostriatal pathways gate processing of behaviorally relevant reward dimensions. PLoS biology, 16(10), e2005722.

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.

The functions of dopamine in learning

Dopamine encodes belief of task state

Besides signaling the canonical model-free reward prediction error (RPE), dopamine (DA) plays a central role in belief updating, representing information related to changes regarding task states. This role of DA in facilitating task performance under an environment with multiple structures or states is in line with studies that found DA to be important for the flexibility to switch choices in reversal tasks (Cools et al., 2009; Cools et al., 2011), and the response to unexpected novel stimuli (Redgrave et al., 1999; DeYoung et al., 2013). Task state could be perceived as conceptual rather than numerical information. Therefore, updating belief of task state could be interpreted as computing a specific form of the identity prediction error. Along with the canonical reward prediction error in temporal difference learning, DA also encodes state/identity prediction errors to drive learning.

Results from a human fMRI study highlight the midbrain DA function of encoding shifts in state beliefs (Schwartenbeck et al., 2016). They defined two different types of surprising information: 1) “Bayesian surprise” which could be informative of a state change, and 2) “uninformative surprise” that occurs not because a new state has replaced an old state but instead, is due to a low probability within an unchanged state. The task consisted of 4 stimuli: 2 “good” (1 tone and 1 shape) and 2 “bad” (1 tone and 1 shape). “Good” or “bad” stimuli led to positive outcomes and negative outcomes with 90% chance, respectively. On a given trial, subjects were presented with a tone and a shape, followed by an outcome. They were then asked to rate their beliefs about the stimulus-outcome contingency, whether they believed it was the shape or the tone that led to the outcome. Critically, “good” and “bad” stimulus

identities remained the same throughout trials. However, the cue modality (audio/visual) was deterministically associated with the outcome, and this relationship switched periodically. Half of the trials were informative, in which one cue predicted win and the other cue (of another modality) predicted loss. And the other half of the trials were uninformative, with both cues being "good" or "bad". "Bayesian surprise" could only be elicited in informative trials. To elaborate, if a subject assumes the current state is the one with the visual stimulus being predictive of outcome, but after noticing that the "good" visual stimulus actually led to a negative outcome, the subject would know that the state has been changed to the other state in which audio stimulus predicts outcome. "Uninformative surprise" could be perceived from uninformative trials. For example, if both tone and shape are "good" but the outcome is negative, this means it is due to the within-state low probability of "good" stimuli leading to a negative outcome. In sum, this design clearly dissociates the two kinds of surprises of interest.

Neural results showed that the BOLD signals of the midbrain DA regions significantly correlated with "Bayesian surprise", the surprise which is indicative of a state change for cue modality. Critically, the midbrain DA region did not encode "uninformative surprise". Furthermore, subjects who performed better in estimating which stimulus determined outcome had less midbrain activation and more activation in the frontal areas that encoded "uninformative surprise". One explanation could be that surprising events within the current state could not be anticipated and might generate more pronounced effects in subjects who were better at learning real contingencies. This study provides strong evidence supporting that dopamine is sensitive to information signaling a change in task state. It also suggests that the

dopaminergic midbrain regions are not exclusively selective to value-related information since both types of surprises were associated with values.

Another study using rats (Hamid et al., 2016) also supports the theory that DA encodes state value, not value per se. In this study, rats learned the drifting reward probabilities of a left and right port for food. While tonic minute-by-minute mesolimbic DA activities correlated with reward rate, phasic DA showed a significant positive correlation with the state value that was estimated by a reinforcement learning model with discrete states. This positive correlation between phasic DA and state value is even stronger than the relationship between DA and reward prediction error. The source of the midbrain DA patterns in encoding task state might partially come from the orbitofrontal cortex, which has been shown to provide input information to the ventral tegmental area on state identity for perceptually similar but conceptually distinct states (Takahashi et al., 2011).

Dopamine strengthens associative relationships and enhances motivational behavior

Besides representing beliefs of task state, midbrain DA has also been shown to enhance motivational behavior (Hamid et al., 2016; Parker et al., 2016) and strengthen contingent relationships, with and without the presence of reward (Parker et al., 2016; Sharpe et al., 2017).

In the study by Hamid et al. (2016), phasic mesolimbic DA was optogenetically stimulated when rats chose to enter the left or right port. With the increased DA, rats were faster at engaging with port entry. Also, the chance that the same port was reentered increased significantly after DA stimulation, regardless of whether the entered port was rewarded or not. Inhibiting DA with the expression of Halorhodopsin resulted in the opposite pattern: rats were

less likely to reenter the same port on the next trial for both the rewarded and unrewarded conditions. This study provides strong causal evidence on the importance of DA in driving motivational learning.

An increase in motivational behavior could be partially attributed to a stronger bond between stimulus and outcome. In the example by Hamid et al. (2016), it could be that DA increased the contingent connection between an entered port and the outcome (not necessarily a reward outcome), therefore rats were more likely to reenter the same port. Indeed, another study by Sharpe et al. (2017) pointed out that DA transients enhance the formation of associations between two events in the presence or absence of reward. In this study, rats did a model-based sensory preconditioning task with both elemental and compound stimuli, with the idea that compound stimuli would partially block the association between the elemental stimulus and outcome. DA transients were optogenetically activated and were able to unblock the supposedly reduced learning in a compound stimulus with both positive and negative outcomes, implying that DA alone is sufficient to form model-based associations. In a second experiment, ventral tegmental area DA neurons were suppressed across the transitions between stimuli in preconditioning. Such an inhibition of DA transients prevented the formation of normal associations between 2 stimuli. This study provides direct causal evidence for DA's role in forming associative bonds between the neural representations of stimuli, with or without reward.

Conclusion

From the evidence mentioned above, midbrain dopamine transients seem to aid several behaviors in goal-directed learning: from representing beliefs of task state by encoding state prediction error to enhancing motivation and forming associative links between stimuli. Importantly, these studies all pointed out that reward is not a necessary factor for dopamine activities to exhibit high-level functions. All these dopaminergic functions, including the canonical role of dopamine for RPE in temporal difference theory, seem to converge on the idea that dopamine is a key neuromodulator to strengthen plasticity between the neural representations of external events related to model-free and model-based task performance. A possible future direction could be to directly test for the different anatomical and functional plasticity driven by dopamine in neural regions related to learning and examine the potential differences from the contributions of dopamine in forming this plasticity.

References

Cools, R., Frank, M. J., Gibbs, S. E., Miyakawa, A., Jagust, W., & D'Esposito, M. (2009). Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. Journal of Neuroscience, 29(5), 1538-1543.

Cools, R. (2011). Dopaminergic control of the striatum for high-level cognition. Current opinion in neurobiology, 21(3), 402-407.

DeYoung, C. G. (2013). The neuromodulator of exploration: A unifying theory of the role of dopamine in personality. Frontiers in human neuroscience, 7, 762.

Hamid AA, Pettibone JR, Mabrouk OS, Hetrick VL, Schmidt R, Vander Weele CM, Kennedy RT, Aragona BJ, Berke JD. (2016). Mesolimbic dopamine signals the value of work. Nature Neurosci 19(1):117-26.

Parker, N. F., Cameron, C. M., Taliaferro, J. P., Lee, J., Choi, J. Y., Davidson, T. J., ... & Witten, I. B. (2016). Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. Nature neuroscience, 19(6), 845.

Redgrave, P., Prescott, T. J., & Gurney, K. (1999). Is the short-latency dopamine response too short to signal reward error?. Trends in neurosciences, 22(4), 146-151.

Schwartenbeck P, FitzGerald THB, Dolan R. (2016). Neural signals encoding shifts in beliefs. Neuroimage 125:578-586.

Sharpe, M. J., Chang, C. Y., Liu, M. A., Batchelor, H. M., Mueller, L. E., Jones, J. L., ... & Schoenbaum, G. (2017). Dopamine transients are sufficient and necessary for acquisition of model-based associations. Nature Neuroscience, 20(5), 735.

Takahashi YK, Roesch MR, Wilson RC, Toreson K, O'Donnell P, Niv Y, Schoenbaum G. (2011). Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. Nature Neurosci 14(12):1590-7.

Chapter 2 Parallel Reinstatement of Observed and Inferred Choices in the Human Frontal Cortex during Contingent Learning

Abstract

Humans and animals can attribute the cause of a reward to multiple causal choices, including directly taken actions and untaken but inferred possible actions. While substantial progress has been made concerning the neural mechanisms of credit assignment for recently taken actions, the neural mechanisms supporting the parallel learning of taken and inferred associations remain largely unknown. We developed a structured probabilistic learning task and a Bayesian learning model, coupled with multivariate pattern-based analyses of fMRI data, to examine the neural mechanisms of credit assignment to both causal choices and inferred alternative choices during the receipt of different food rewards. We found a general representation of the causal choice taken reactivated in the lateral orbitofrontal cortex (lOFC) upon reward receipt and an identity-specific representation in the central orbitofrontal cortex (cOFC) and amygdala. Inferred alternative choices were simultaneously reinstated in the lateral frontopolar cortex (lFPC) and medial prefrontal cortex (mPFC), and the degree to which behavior reflected these inferences was associated with stronger activity in the ventral tegmental area (VTA) during updating. These findings reveal two parallel frontal networks for the learning of taken and inferred choices, demonstrating the flexibility to learn information simultaneously in an uncertain yet structured environment.

Introduction

Causal relationships between actions and outcomes can be realized at various levels within the structure of an environment, requiring organisms to learn in ways that reflect this structure. This often means that observing the outcome of one choice can provide insight into what would have happened if a similar alternative choice had been made. For example, imagine waking up with a severe headache and attributing the cause of your headache to the beer you enjoyed the night before. This observation could also give you information about what would have happened if you had made a similar choice, such as having the same amount of wine, preventing the need to make such mistakes in the future. Such inference relies on the ability to represent the relational structure of the environment, allowing a learner to assign credit to both the direct causal choice and infer that related events governed by the same latent cause (alcohol) would have led to the same outcome (a headache). Although this inference process is essential to flexible learning, the neural mechanisms that support these processes are not well understood. In the current study, we aim to identify the neural network enabling the formation and updating of causal relationships in structured task space and to parse this network to identify the neural mechanisms for learning from both experienced and inferred choice-outcome associations.

Pioneering research across species has made tremendous progress towards understanding the neural mechanisms underpinning contingent learning about choice- and stimulus-outcome associations learned from direct experience (Doll et al., 2009; Gallagher et al., 1999; Rushworth et al., 2007; Schoenbaum & Roesch, 2005; Valentin et al., 2007). These studies have pointed to complementary roles for lateral orbitofrontal cortex (lOFC) and

amygdala in learning stimulus and choice-outcome associations from direct experience (Boorman et al., 2013; Boorman et al., 2016; Jocham et al., 2016; Noonan et al., 2017; Saddoris et al., 2005; Sias et al., 2024; Walton et al., 2010). When delays between choices and outcomes intervene, as is commonplace in the real world, the neural mechanisms are less well understood. One plausible mechanism for forming associations across lengthier delays would be by reinstating the causal choice at the time of the outcome, thereby enabling synaptic plasticity between coactive neuronal ensembles for the online outcome and its causal choice, further reinforced by dopamine. Notably, neuronal ensembles re-coding recent choices have been identified in lOFC, specifically near feedback time in monkeys performing a delayed rule task, albeit in the absence of any learning (Tsujimoto et al., 2009). The coding of choice-related information at non-decision times might function as a mechanism to link choices with concurrent outcome signals for learning. This potential mechanism of choice coding is likely to operate in tandem with the coding of outcomes during cues and in anticipation of feedback (Howard et al., 2015; Kahnt et al., 2014), both working in a complementary fashion to strengthen stimulus-to-outcome contingencies.

While there is no consensus on where and when the representation of a choice occurs outside of the decision time, some evidence suggests that a reinstatement of a choice can be decoded. However, the functional mechanisms and timing of the reinstatement of choice remain unclear. For example, sensory content about choice is found in category-specific visual and temporal sensory cortices before the decision, and the activation is predictive of the to-be-chosen stimulus (Doll et al., 2015). To address this question, we hypothesize that at the time of feedback, there exists a representation of the choice stimulus that aligns with the

representation of the outcome and dopaminergic prediction errors. This alignment would reinforce the stimulus-to-outcome association in accordance with Hebbian plasticity. The reinstatement of a stimulus during feedback could be encoded in either higher-level prefrontal regions or lower-level sensory cortices. Works in object recognition identify that within the prefrontal cortex, the orbitofrontal cortex (OFC) functions as a "top-down" region in the cortical hierarchy for object representation (Bar et al., 2006; Summerfield et al., 2006). As more information is acquired about object identity, the neural activity in the temporal lobe increases and moves forward toward the anterior regions (Bar et al., 2001). Moreover, the OFC receives abundant inputs from sensory cortices (Öngür & Price, 2000). Therefore, it could contain a representation of choice after enough exposure. Nevertheless, we could not rule out the possibility that the encoding of perceptual features of stimuli is preserved and reactivated in sensory cortices during feedback.

Within the prefrontal cortex, the lateral orbitofrontal cortex (lOFC) is the most plausible region to contain a reinstatement of choice. LOFC has been shown to exhibit a high level of sensitivity to stimuli across multiple modalities (Wilson et al., 2014), including taste and texture for reward outcomes (Rolls, 2008), salient visual cues (Bouret & Richmond, 2010), and odors (Howard & Kahnt, 2017; Li et al., 2010; Sagar et al., 2023; Tegelbeckers et al., 2023). As a "heteromodal nexus" (Rushworth et al., 2007), the OFC maintains detailed representations of the identity of expected outcome (Howard & Kahnt, 2017; Howard et al., 2018; Izquierdo et al., 2004; McDannald et al., 2011; Ostlund & Balleine, 2007; Rudebeck & Murray, 2014; Stalnaker et al., 2014; Wallis & Miller, 2003). If LOFC possesses a neuronal ensemble representing choice during feedback, it will dovetail with its representation of outcome, therefore offering a

convincing explanation of the lOFC function in precisely assigning credit for specific outcomes to past choices (Boorman et al., 2016; Izquierdo et al., 2004; Jocham et al., 2016; Noonan et al., 2017; Rushworth et al., 2011; Walton et al., 2010), and that disrupting OFC activity at feedback time would reduce the representation of schemas and structures in building a cognitive map (Lin & Zhou, 2024; Miller et al., 2022; Miller et al., 2017; Seo & Lee, 2010; Takahashi et al., 2011). It is, therefore, sensible that some studies have suggested the OFC might represent stimulus information during learning (Schoenbaum & Eichenbaum, 1995; Walton et al., 2010; Wilson et al., 2014), although these studies do not test for the type of choice-related information in this potential representation, or if the representation exists during feedback during contingent learning in humans.

Adaptive decision-making in complex environments often requires learning both directly experienced choice-outcome contingencies but also from contingencies that are inferred via latent associations. A growing body of work focuses on the encoding and tracking of positions in task space, with a particular emphasis on how the brain represents latent states or the shared structures among stimuli (Lin & Zhou, 2024; Schuck et al., 2016; Wilson et al., 2014; Witkowski et al., 2022). Essential for planning and flexible behaviors (Miller et al., 2017), such a representation of latent task structure exists in the prefrontal cortex, including the orbitofrontal cortex (OFC) (Boorman et al., 2021; Lin & Zhou, 2024; Schuck et al., 2016; Seo & Lee, 2010; Stalnaker et al., 2015; Takahashi et al., 2011; Wilson et al., 2014) and the medial prefrontal cortex (mPFC) (Witkowski et al., 2022). However, analogous to representations of choice, it remains unknown whether the brain encodes a representation of the unobserved but inferred stimulus during feedback. When navigating among different task states, agents need to

process information among alternative options before deciding. Recent evidence points to the frontopolar cortex (FPC) as playing a critical role in monitoring and comparing alternative strategies, goals, and options (Boorman et al., 2011; Boorman et al., 2009; Daw et al., 2006; Koechlin et al., 1999; Park et al., 2017). Due to its high hierarchical position in the prefrontal organization for cognition (Koechlin & Hyafil, 2007), FPC seems to process the latent information about stimuli that are critical for decisions, including counterfactual and unchosen stimuli (Boorman et al., 2011; Boorman et al., 2009), pending task sets (Koechlin et al., 1999) and credibility of alternative options (Park et al., 2017). Therefore, it is FPC carries a reinstatement the pending inferred stimuli.

In this study, we test these hypotheses using a "structured probabilistic learning task", where participants learned to track 2 probabilistic systems of stimulus-reward associations. This paradigm allows the testing for reinstatement at the feedback of both choice and the inferred stimulus through inference. Previous studies have subjects simultaneously viewing two or more options during a decision (Rudebeck & Murray, 2014; Witkowski et al., 2022), therefore posing challenges to separate neuronal responses associated with individual stimuli. To address this concern, participants were presented with options sequentially, enabling us to test for accurate representations of individual stimuli at feedback. Our design also features a separate functional localizer to test whether there exists a reactivation of stimuli in the category-specific visual temporal cortices. We found a reinstatement of choice stimulus at feedback in the lOFC and an analogous code of the inferred stimulus in the lFPC and mPFC. Moreover, we pinpoint the dopaminergic ventral tegmental area (VTA) signals at feedback that predict the individual differences in the ability to infer during learning. Taken together, we disentangle a network of

regions processing distinct learning signals and show novel results regarding representations of task-relevant stimuli at feedback.

Results

Structured probabilistic learning task

We designed a novel structured probabilistic learning task in which participants learned the probability that each of the four visual stimuli (face, tool, body part, scene) would lead to one of two food rewards (Figure 2.1). There were two types of trials: free-choice and forced-choice trials. In each free-choice trial (Figure 2.1B), participants selected one of the two sequentially presented options while considering two pieces of information: the amount of each food being offered and the probability that choosing either stimulus would lead to a specific outcome. The food rewards for each participant were chosen to be approximately equally desirable according to their previously recorded ratings (see Methods). The amount of food on offer in each trial was indicated by colored numbers at the top of the screen (e.g., a pink “7” means seven macaroons can be won). Participants knew that at the end of the experiment, one trial would be selected at random and paid out as their winnings for the experiment. This setup required participants to choose the stimulus with the highest probability of leading to the desired outcome in each trial.

Importantly, the four visual stimuli formed two sets of "linked pairs" (Figure 2.1C). Stimuli within a linked pair shared a common associative probability with each food reward, and participants were instructed that these stimuli would always lead to the same outcome. Thus, participants could update their knowledge that one stimulus in a linked pair would yield a

certain food reward based on the experienced outcomes of the other. The identities of the choices in each linked pair changed in each run, requiring participants to relearn which stimuli were in each linked pair.

Participants acquired knowledge of the two linked pairs through deterministic forced-choice trials (Figure 2.1A). In each trial, two stimuli were sequentially presented on the screen, and one of the stimuli was highlighted as the choice stimulus. Participants were required to choose the highlighted stimulus, after which the associated outcome was shown. Participants could learn which trials were in a linked pair by observing which stimuli led to the same outcome across forced-choice trials.

After learning the two linked pairs, participants completed free-choice trials in which they were free to choose from either of the two available stimuli on each trial. They completed 24 forced-choice trials and 48 free-choice trials in three separate runs.

Figure 2.1. Learning task design

A

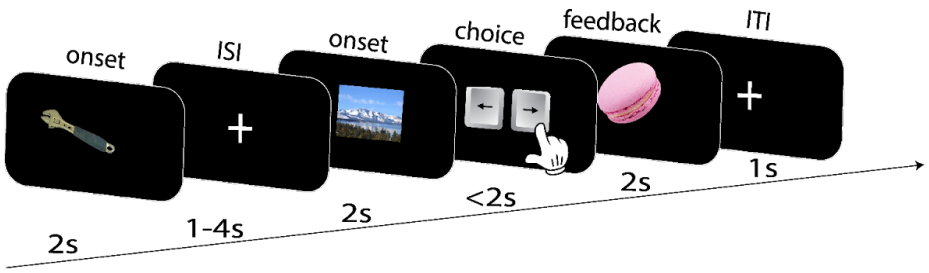


Illustration of the forced-choice trials in the fMRI task. Participants were presented with 2 stimuli, and then were instructed on which stimulus to choose. The left arrow indicated the first stimulus, and the right indicated the second.

B

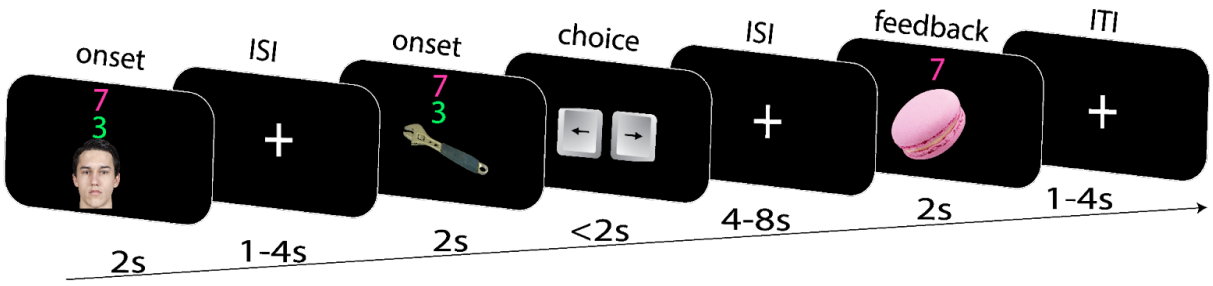
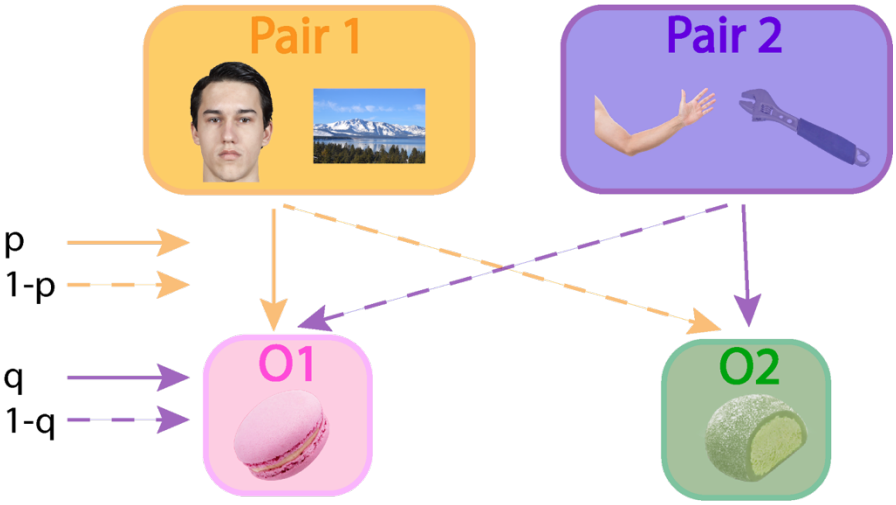


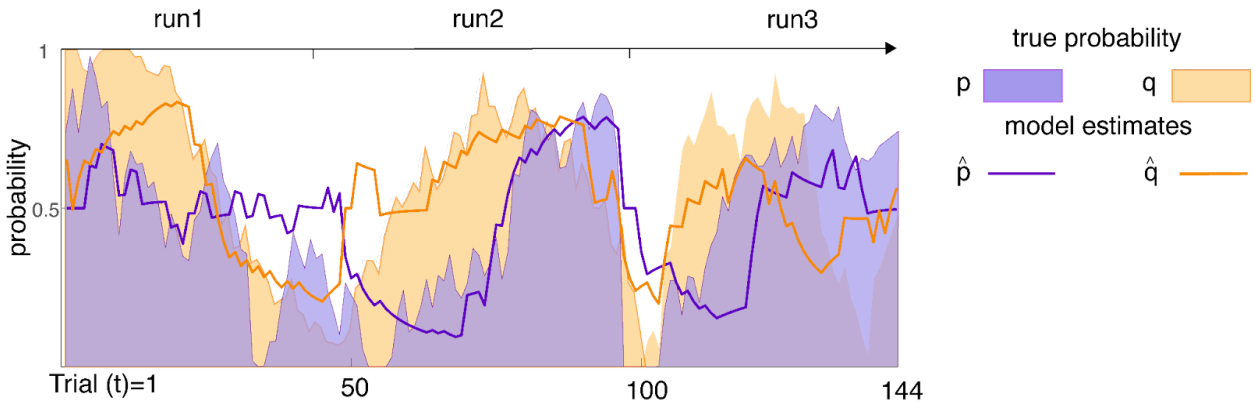
Illustration of the free-choice trials in the fMRI task. Participants freely choose between two stimuli based on two factors: their estimate of the associative probabilities (p, q) between a stimulus and each food, and the randomly generated number of food they could receive if obtained. The color of each number suggested the magnitude of the outcome on which that number of food could be received. Finally, they observed the outcome after the choice with a delay.

C



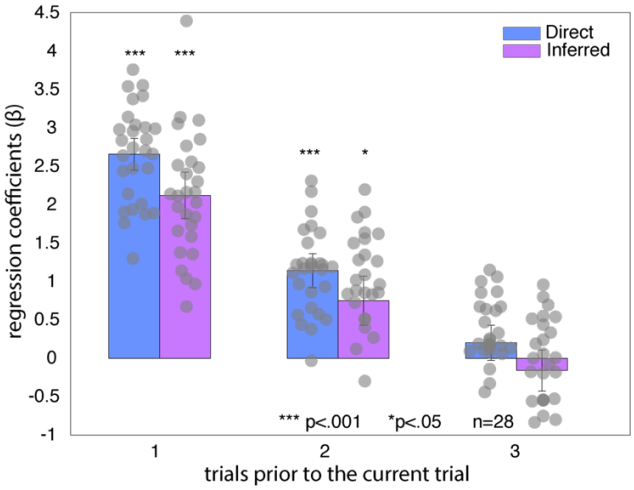
C) Four stimuli were organized into two independent systems called “linked pairs”. Each “linked pair” consists of two stimuli with shared transition probabilities leading to the two types of food outcomes. Participants could therefore infer and update their estimates of the transition probabilities of the “paired stimulus” (e.g., scene) after observing the outcome of choosing a stimulus (e.g., face) within the same linked pair.

D



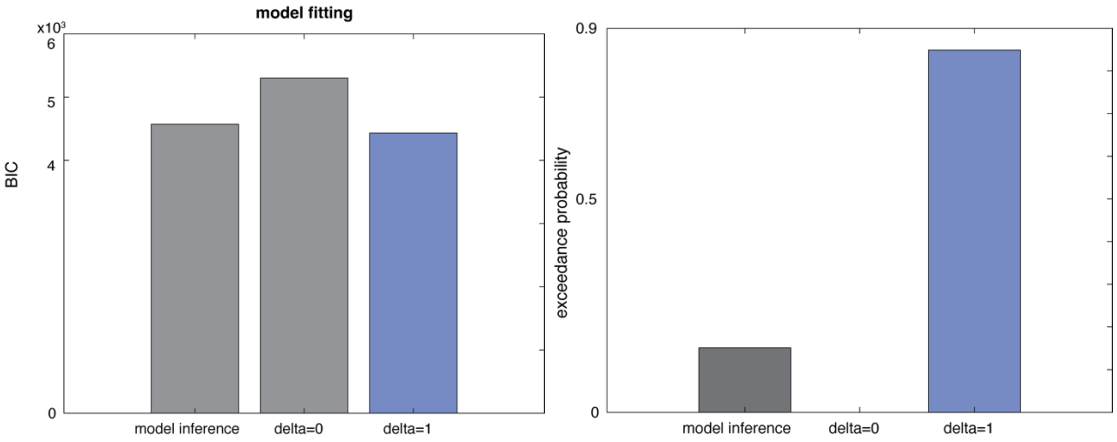
(D) Example of a participant's learning trajectory for the transition probabilities of linked pairs to outcomes. Shaded regions indicate the true associations of linked pair 1 (p, orange) and linked pair 2 (q, purple). Orange and purple lines were the estimated values of p and q based on the perfect-inference learning model (see computational models for details).

E



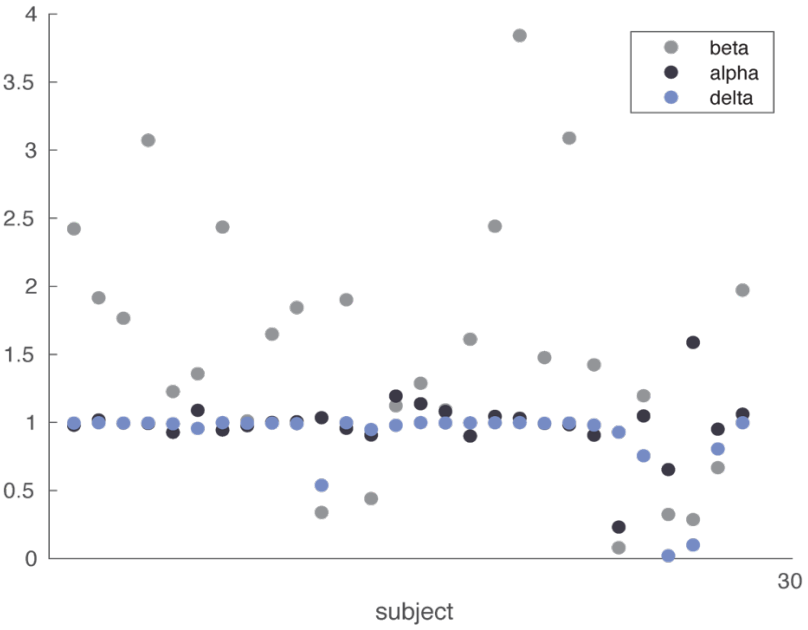
(E)The logistic regression analysis results indicate the impact of previous choices and outcomes on current decision-making. The bar heights represent the mean of regression coefficients ± SEM.

F



(F) Results of model comparisons using BIC (left) and exceedance probability (right) for weighted inference, no-inference, and perfect-inference models (See Methods for details).

G



(G) Estimated parameters for the weighted inference model include delta (inference term), alpha (relative preference for outcomes), and beta (choice stochasticity).

Direct and indirectly experienced choice outcome associations influence current choices

To maximize rewards, participants needed to track the probability that each stimulus led to each outcome over time and use that knowledge to decide which choice will lead to the outcome they currently desire. To test whether previous choice-outcome associations influenced current decisions, we employed a logistic regression model that predicted choices based on the currently preferred outcome and the choice-outcome associations from the previous three trials (Equation 11). Importantly, successful learning of this task would also involve inferred learning of the received outcome for the other stimulus within the same linked pair as the choice stimulus. Therefore, we included regressors for the previous outcomes of the complementary stimulus in the linked pair. Finally, based on reinforcement learning theory, we predicted that choice-outcome pairings that occurred further in the past would have less influence on the current choice. (Daw & Tobler, 2014; Ito & Doya, 2009; Seymour et al., 2012).

The logistic regression results indicate that participants exhibit positive learning for both the direct (choice) stimulus and the inferred (paired) stimulus. The regression coefficients for the direct choice stimulus are significant for the 1-back ($t[27] > 12.82$, $p < .001$) and 2-back ($t[27] > 5.15$, $p < .001$) conditions (Figure 2.1E). Participants also show inferred learning, as indicated by the regression coefficients for the 1-back ($t[27] > 7.03$, $p < .001$) and 2-back ($t[27] > 2.36$, $p < .05$) conditions. However, participants do not seem to be significantly influenced by either direct ($t[27] = 0.88$, $p = 0.19$) or inferred ($t[27] = -0.60$, $p = 0.28$) information from 3 trials back. The drifting probabilistic schedule poses a challenge for participants to integrate information further into the past, unlike a task that uses a reversal probabilistic schedule that shows significant 3-back regression coefficients (Witkowski et al., 2022).

Next, we tested whether these effects differed over time (1-back, 2-back, 3-back) and across information time (directly experienced or inferred) using a two-way ANOVA. Previous trials that are closer in time to the current trial exert a more pronounced influence on the current decision compared to trials from further back in time ($F(2,167) = 14.49$, $p < 0.001$). The effect of past direct outcomes is greater than those during inferred trials ($F(1,167) = 148.7$, $p < 0.001$), and there is no interaction between time and outcome type ($F(2,167) = 0.23$, $p = .80$). The ANOVA results indicate substantial learning for both direct and inferred stimuli, with a notably stronger learning effect observed for direct stimuli.

Model comparison

To estimate subjects' beliefs on stimulus-outcome associations throughout free trials, we used a Bayesian learning model (See Methods) that updates trial-by-trial beliefs about the probability a particular choice would lead to one outcome or the other based on the history of choices and outcomes made by the participant. We fit three different models, each reflecting a different possibility for how participants could integrate information from inferring the paired stimulus from the observed outcome (via the linked pairs). The first was the "no inference model" reflecting no additional information from inference, and the second was a "perfect inference" model, which counted inferred information just as meaningful as the directly experienced information. The third model included a free parameter (δ) which accounted for individual differences in the weight given to the inferred information. All models included two additional free parameters: α that quantifies the subjective value of one outcome relative to the other; β, an inverse temperature term that measures the stochasticity of participant choices. (See methods for parameter estimates).

We compared the three learning models using Bayesian information criterion (BIC) and exceedance probability metrics. The "perfect-inference" model outperforms the other 2 models with the lowest Bayesian information criterion (BIC) across participants (Figure 2.1F). This model also showed a high exceedance probability (EP=0.85) relative to other models "no inference" model (EP = 0.00) and the "weighted inference" model (EP = 0.15) (Figure 2.1F). Notably, 23 out of 28 participants have almost perfect inference terms with $\delta > 0.9$ (Figure 2.1G) according to the estimates in the "weighted inference model", which suggests that the best fits for the weighted model were nearly synonymous with the perfect inference model. In the "perfect-inference model", inference term δ and inverse temperature β both positively correlate with logistic regression coefficients ($r=0.48$, $p<0.02$; $r = 0.55$, $p <0.005$), showing our model estimates to be consistent with behavioral outcomes.

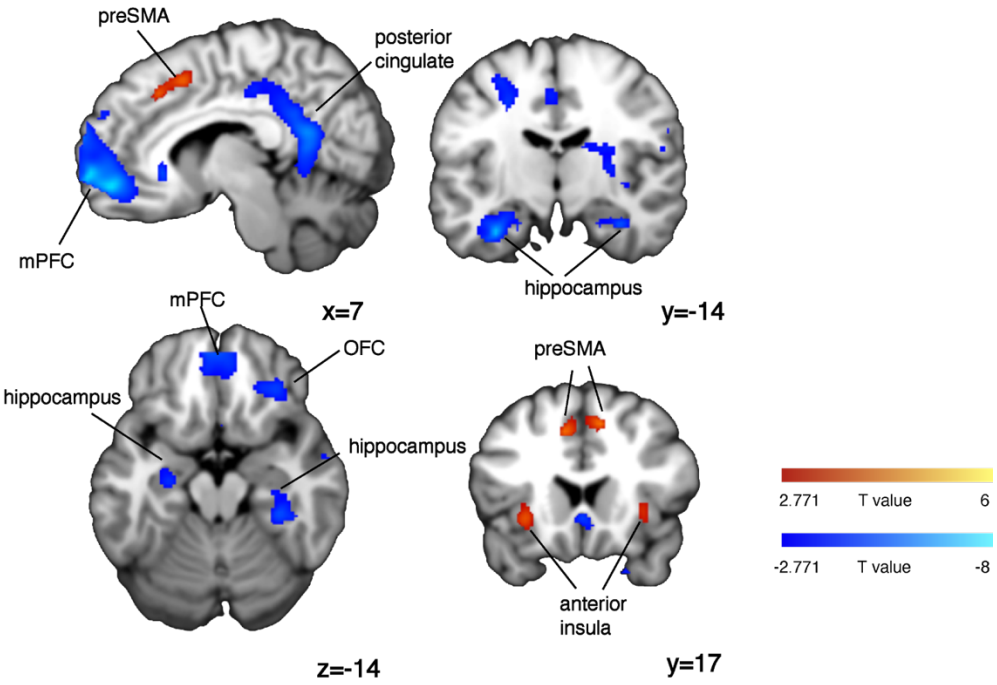
A network for belief updating and confirmation at feedback

Our next analysis sought to identify the network of regions that support updating the associations between task-relevant stimuli and outcomes during the receipt of outcomes. In this task, participants not only update their beliefs about the choice-to-outcome associations but also about that between the outcome and the complementary stimulus in the linked pair. To quantify the degree of belief updates, we used Kullback-Leibler divergence (DKL), defined as the change in information between the prior and the posterior belief distribution, also called "Bayesian surprise" (Iglesias et al., 2013; Schwartenbeck et al., 2016), as it reflects how new information has shifted our beliefs. Because beliefs about each stimulus in the linked pair were being updated in each trial, we summed the DKL for each update and input it as a parametric modulator for the blood-oxygen-level-dependent (BOLD) activity during feedback (see

Methods). Regions with positive effects include the pre-supplementary motor area (pre-SMA)/dorsal anterior cingulate cortex (dACC) (left, [x,y,z] = [-10, 20, 44], t[27] = 4.62; right, [x,y,z] = [10, 14, 50], t[27] = 4.54), and bilateral anterior insula (left, [x,y,z] = [-32, 18, -2], t[27]= 3.84; right, [x, y, z] = [32, 22, -4], t[27] = 3.19), (small volume correction within a priori defined ROIs using permutation-based threshold-free cluster enhancement [TFCE] (Smith & Nichols, 2009) at pTFCE < 0.05). These clusters together form a network for updating the belief of choice-outcome probabilistic associations. We also found clusters with negative effects for belief confirmation, reflecting greater activity when outcomes were consistent with previously held beliefs. We found these signals in the bilateral hippocampus (left, [x,y,z] = [-30, -16, -18], t[27] = 4.42; right, [x,y,z] = [26, -8 -20], t[27] = 3.79), the right lateral orbitofrontal cortex (lOFC) ([x,y,z] = [34, 34, -18], t[27] = 4.45), the ventromedial prefrontal cortex (vmPFC) ([x,y,z] = [10, 52, -6], t[27] = 7.88), and the posterior cingulate cortex ([x,y,z] = [8, -54, 14], t[27] = 5.18) (all whole-brain cluster-corrected with permutation-based threshold-free cluster enhancement [TFCE] (Smith & Nichols, 2009) (at pTFCE < 0.05). Together, these univariate results suggest 2 separable networks responsible for belief updating and belief confirmation.

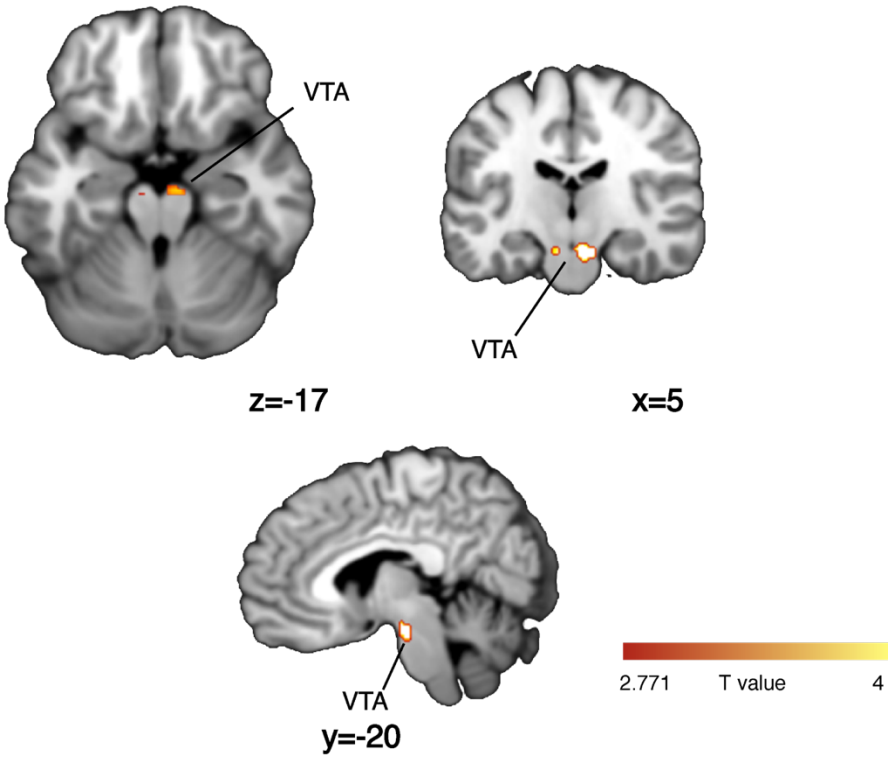
Figure 2.2. Network of regions during feedback of updating and confirming stimulus-to-outcome associations of inferred and observed information

A



(A) T statistics maps of brain regions whose activity at feedback reflected the sum of information gained from experienced and inferred information (DKL). Positive t-statistics (red and yellow) indicate the network reflecting the degree of belief update or reconsideration of estimates for transition probabilities. Negative t statistics (blue and cyan) indicate the network reflecting the degree of "stamped-in" belief or confirmation of estimates for transition probabilities. For illustration, maps display regions at a threshold of $t[27]=2.771$, $p < 0.005$, uncorrected.

B



(B) T-statistic effect within VTA of a regression correlating the significant behavioral index of learning from inferred information (mean of 1-back & 2-back logistic correlation coefficient for the inferred information) with the neural DKL effect (change of information from experienced and inferred stimuli). For illustration, maps display regions at a threshold of $t[27]=2.771$, $p < 0.005$, uncorrected.

VTA signal at feedback predicts individual differences in inferred learning

Recent studies have highlighted the importance of the dopaminergic midbrain, particularly the ventral tegmental area (VTA), in forming model-based associations (Langdon et al., 2018; Sharpe et al., 2017) and encoding action policies or task states (Gershman et al., 2019; Hamid et al., 2015) including directly observed as well as inferred outcomes (Witkowski et al., 2022).

Univariate results do not show a group-average VTA signal at feedback for changes in belief (Figure 2.2A), which might be explained by the possibility that VTA signals vary individually according to differences in inferred learning. To test this idea, we conducted a between-subject analysis, correlating individual participants' univariate signals with 2 behavioral measurements: the mean of significant (1- and 2-back) regression coefficients from the logistic regression for the direct and inferred conditions (See Equation 11). Results confirmed our hypothesis, showing individual differences in the extent to which participants integrated inferred information in their learning behaviorally predicted the magnitude of the univariate effect of DKL in VTA ([x,y,z] = [8, -18, -18], t[27] = 3.78, small volume correction within a priori defined ROI pTFCE< .05; Figure 2.2B) (Smith & Nichols, 2009).

lOFC contains a neural code for the causal choice at feedback

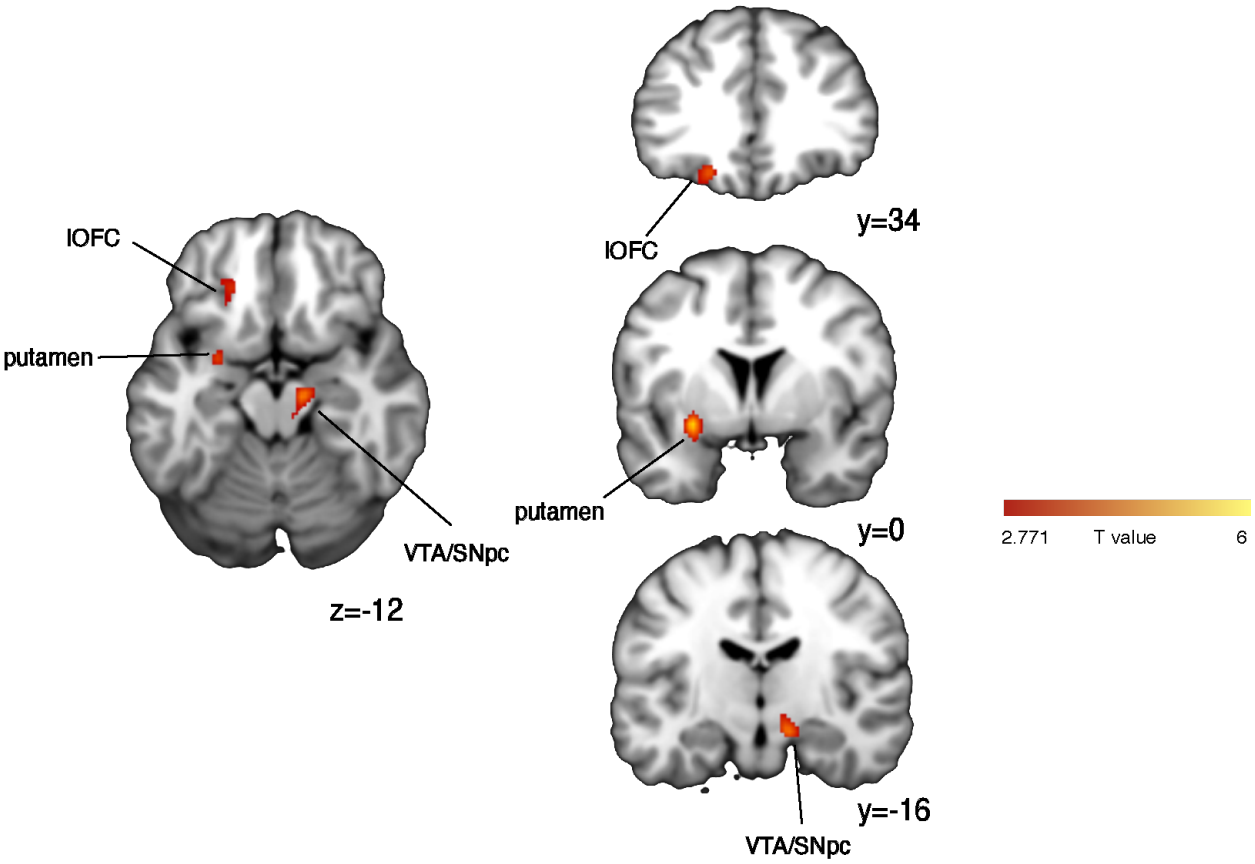
Based on our univariate findings indicating the neural correlates of belief confirmation at feedback in the lateral orbitofrontal cortex (lOFC) and the hippocampus, we proceeded to investigate whether these regions encoded specific information about the causal choice. We hypothesized that upon observing the outcome of a decision, the lOFC would reactivate a representation about the causal choice, thereby enabling credit assignment of the outcome to the relevant choice.

To test this, we employed a multivariate pattern analysis (MVPA) within predefined regions of interest (ROIs), specifically targeting the lOFC, using a searchlight approach to decode information related to the causal choice. We conducted the MVPA in a cross-validation manner to prevent overfitting. Specifically, we trained a support vector machine (SVM) to classify

patterns of activity in the lOFC at the time when participants viewed the outcome, according to which choice caused the outcome (scene, face, body part, or tool). We trained the classifier with data from two runs and tested it in the third left-out run (See Methods). Importantly, choice labels were balanced across outcome types, ensuring the accuracy of the classifier during training. To control for potential biases, we created a permuted null distribution by retraining the classifier with randomized labels and testing on the held-out set 100 times per searchlight. The mean accuracy of the permuted labels was subtracted from the accuracy of the classifier using the true training labels.

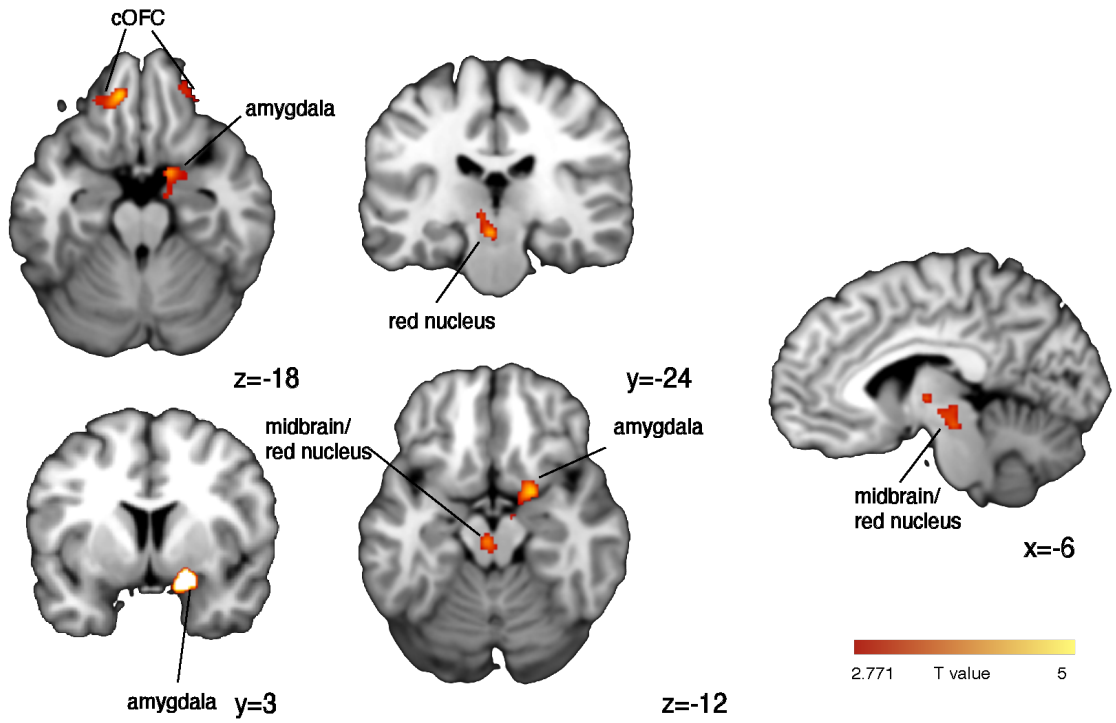
Our a priori hypothesis centered on the lOFC as the region likely to exhibit reactivation of the choice code during feedback. To test this, we employed a whole-brain searchlight approach for the MVPA analysis. We identified voxels within our a priori defined lOFC showing significantly above chance decoding accuracy, suggesting that this region contains a choice code that reactivates at the time of received outcome ([x,y,z] = [-22, 36, -16], t[27] = 3.89, pTFCE = 0.0136; Figure 2.3A). Notably, this analysis also identified significant clusters of voxels in the left putamen ([x,y,z] = [-28, 0, -6], t[27] = 5.37, pTFCE < 0.005), and the ventral tegmental area (VTA) substantia nigra pars compacta (SNpc) ([x,y,z] = [12, -18, -12], t[27] = 4.27, pTFCE < 0.005). This suggests that the lOFC, as well as the putamen and SNpc, contain a representation of the choice code, which reactivates when outcomes are received.

Figure 2.3. A code for choice reactivates in the OFC during feedback



(A) T statistical map showing effects of decoding of the general information of choice from analysis depicted in (x) in the lOFC (survived SVC within an a priori OFC ROI), putamen, and VTA/SNpc displayed using the same conventions as Figure 2.2.

B



(B) T statistic map showing effects of decoding of the identity-specific information of choice from analysis depicted in (x) in the cOFC (SVC within an a priori OFC ROI), BLA nucleus in amygdala (SVC within an amygdala ROI), midbrain/red nucleus (SVC within a midbrain ROI) displayed using the same conventions as Figure 2.2. The clusters all survived small volume correction.

An identity code for causal choice is reinstated in the OFC and the amygdala

Our previous results indicate that a choice code reactivates in the lOFC during feedback. This code represents general and more abstract information about choice beyond its perceptual identity . As such, it complements our previous analysis by disentangling the various dimensions of choice-related information and testing for an identity-specific code for the causal choice. The

design of our paradigm includes 3 sets of forced-choice trials that are separate from free-choice trials. In these forced-choice trials, the onsets of the stimuli include no other information besides the stimuli themselves (Figure 2.1A), since there is no additional visual information, such as the colored numbers in the free-choice trials. We utilize this design to perform an MVPA to test our hypothesis that the reactivation of choice code contains relatively perceptual information about the choice stimulus and thus ties to the *identity* of the stimulus.

Using a whole-brain search light, we trained an SVM using the activity patterns during the onsets of forced-choice stimuli so that the SVM learns to distinguish the identities of the 4 stimuli. Importantly, the neural data used to train the classifier consists of the pure representations of the stimuli' identities, therefore guaranteeing the classifier's validity to test for the identities of these 4 stimuli. Importantly, because of the separate forced-choice blocks, the neural data used to train the classifier does not include task-state relevant information but consists of identities of the stimuli, and therefore guarantees the classifier's validity to test for the identities of these 4 stimuli. The trained classifier is then used to test during feedback of the free-choice trials to see if the activity patterns in any region of the brain could be parsed according to stimulus identity during choice. Motivated by our a priori hypotheses about the OFC's role in representing choice identity, we examined anatomically defined ROIs within the OFC expected to contain these representations and applied TFCE to correct for multiple comparisons. We found significant decoding accuracy in bilateral OFC (left OFC [x,y,z] = [-18, 38, -16], t[27] = 4.28, pTFCE = 0.013; right OFC [x,y,z] = [12, 42, -18], t[27] = 4.63, pTFCE = 0.007) (Figure 2.3B). Additionally, this analysis identified significant decoding accuracy within clusters in the right amygdala ([x,y,z] = [18, -2, -18], t[27] = 3.13, pTFCE = 0.017), and the midbrain

([x,y,z] = [-4, -24, -10], t[27] = 4.18, pTFCE = 0.0148) (Figure 2.3B), indicating that these regions contain an identity-specific representation of the choice during feedback. Critically, early occipitotemporal areas where object recognition typically occurs do not show significant decoding accuracy. This MVPA generates results that build on our previous findings, demonstrating that the cOFC contains a more specific stimulus identity code rather than just a general choice representation.

Reactivation of a representation of inferred stimulus at feedback in FPC & mPFC

We hypothesized that the brain would reinstate not only the choice stimulus at feedback time but also the latent stimulus with the same task-relevant statistics. Behavioral results indicate strong learning for both directly experienced chosen stimulus as well as paired but inferred stimulus. In other words, with the knowledge of the task structure that the choice stimulus and the paired stimulus are statistically identical, participants successfully assigned credit to the paired stimulus when reward is received even though it was not directly chosen or observed in a trial. Given previous results showing that representations of the hierarchical task structure appear in the mPFC during inference (Witkowski et al., 2022), and that information about task-relevant alternative options is processed in the FPC (Boorman et al., 2011; Boorman et al., 2009; Koechlin et al., 1999; Park et al., 2017), we hypothesize that the mPFC or the FPC would contain a code for the paired stimulus that reinstates at the time of feedback. To find a potential code for inferred causal stimulus, we trained an SVM to differentiate activity patterns at feedback based on the paired stimulus of the choice stimulus. For example, in one case, we trained the SVM to differentiate trials “body part” was the inferred causal stimulus for both “scene” and “face,” then tested the classifier on left-out trials from a run where “body part”

was the inferred causal choice for “tool”. This procedure was then repeated across all combinations of training and test blocks. Importantly, this approach allows 3 runs to have 3 distinct pairs, therefore ensuring that the choice stimulus or the type of food would not confound our results. Given the behavioral result that participants, having learned the task structure of the linked pair, flexibly attributed credit to the paired stimulus when seeing the food type received, we could determine the specific brain regions in which the paired stimulus is reactivated, as indicated by decoding accuracy. When we used the trained SVM to test the identity of the paired stimulus at feedback, we identified a cluster in the right frontopolar cortex with significant decoding accuracy ([x,y,z] = [24, 56, -2], t[27]=4.16; pTFCE = 0.0232 ROI corrected; Figure 2.4A), suggesting the location of the neural signature for inferred learning in the form of a code for paired stimulus that reactivates at feedback. We also found a cluster in the left medial prefrontal cortex ([x,y,z] = [0, 52, -12], t[27]=3.02; pTFCE <0.031 ROI corrected; Figure 2.4A). This finding reveals the neural basis of indirect model-based credit assignment by confirming that these prefrontal regions reinstate a representation of the inferred stimulus at feedback time for adaptive decision-making in subsequent trials.

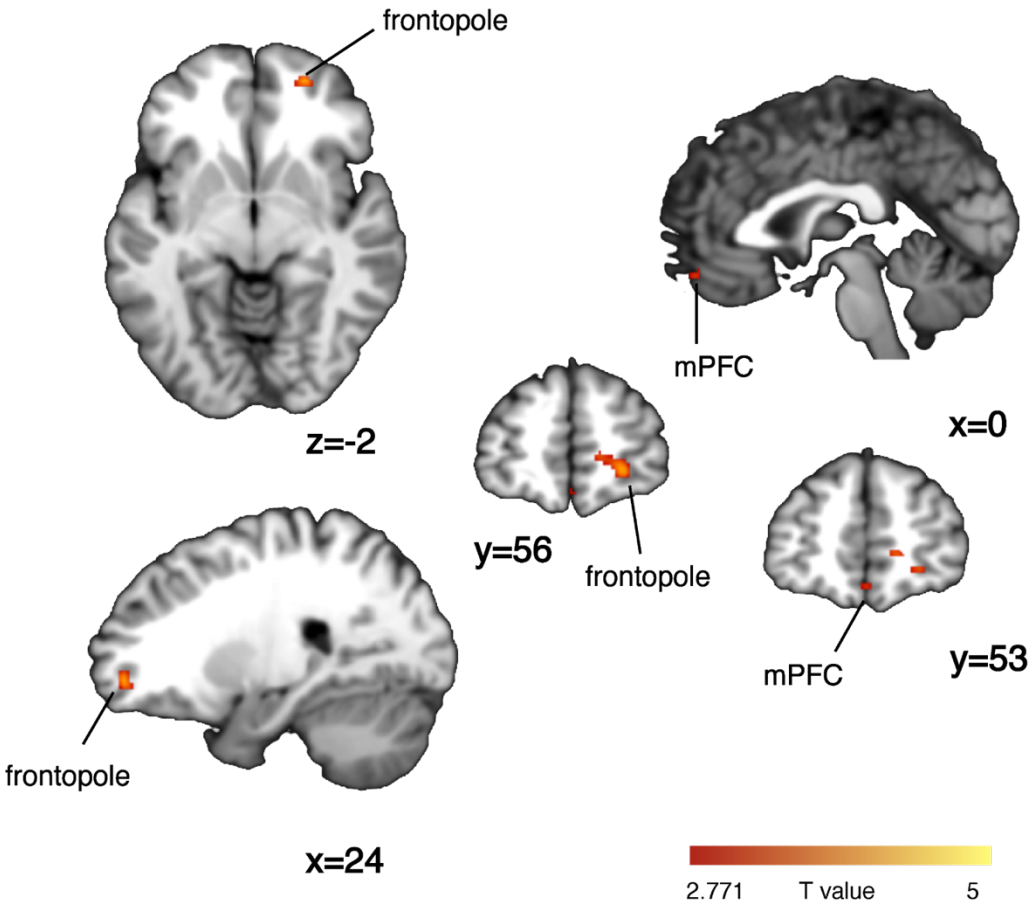
Identity code for inferred stimulus in left vlOFC

We then ask the question if, like the choice stimulus, there exists an analogous reinstatement of the identity-specific code of the paired stimulus when participants receive a reward. An identity-specific code of paired stimulus during feedback would be adaptive to the formation of the inferred stimulus-to-outcome association in a model-based task, which demands more than the learning of choice stimulus-to-outcome contingency. We trained the classifier to recognize the patterns of the identity of the paired stimulus during stimuli onsets in forced-choice trials

and used it to decode the paired stimulus at feedback in free-choice trials. Results show significant decoding accuracy in the ventral lateral OFC ([x,y,z] = [-44, 40, -20], t[27] = 3.72; Figure 2.4B). Remarkably, the feedback period contains a decodable identity-specific code for the paired stimulus even though it has not been shown during a trial. Collectively, these findings demonstrate that the brain reinstates not only information about the choice during feedback but also task-relevant inferred stimuli. This reactivation could aid in subsequent decisions by facilitating? the update of the contingencies between latent stimuli and outcomes.

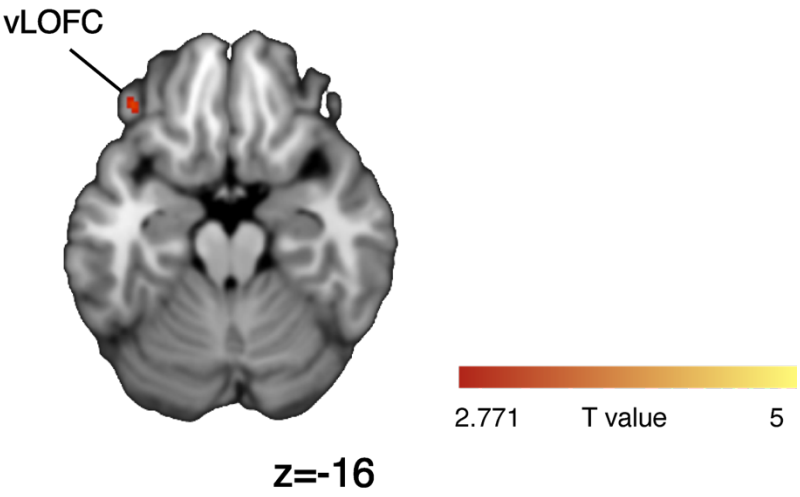
Figure 2.4

A



(A) T statistic map showing effects of decoding of the general information of inferred stimuli (the stimulus within the same linked pair as the choice) from analysis depicted in (x) in the FPC (SVC within an a priori FPC ROI), mPFC (SVC within an a priori mPFC ROI) displayed using the same conventions as Figure 2.2. The clusters all survived small volume correction.

B



(B) T statistic map showing effects of decoding the identity-specific information of inferred stimuli in vlOFC.

No evidence for reactivation of choice or inferred information on category-selective sensory cortices during feedback

To test the possibility that choice stimuli might also reinstate on lower-level category-selective sensory regions during feedback, we included a functional localizer after the main task to locate the regions responsive to different visual categories on an individual basis. We identified each participant for their commonly recognized category-selective ROIs: fusiform face area (FFA) for face (Kanwisher & Yovel, 2006), the parahippocampal place area (PPA) for scene (Epstein et al., 1999), the extrastriate body area (EBA) for body part (Downing et al., 2001), and the lateral occipital cortex (LC) for tool (Bracci et al., 2012; Chao et al., 1999). We test whether the univariate response at feedback time in the relevant category-selective ROI (e.g. PPA) is greater for the selected (e.g. scene) in comparison to the other 3 stimuli. If so, it would suggest that choice stimulus reactivates at feedback on a sensory level.

We utilized the localizer output to create 4 category-selective ROIs for each participant (See Methods). Then for every trial during feedback, we extracted the values of univariate pattern activities using the 4 category-selective ROIs as masks. We examined if the average values extracted from a category-selective ROI during feedback are higher for trials where the choice stimulus belongs to that category than other trials where the other 3 stimuli were chosen. Category-selective ROIs do not exhibit significantly higher univariate activities on trials where the choice stimulus matched the corresponding category, shown by 1-way ANOVA (left PPA $F(3, 96)=0.13$, $p=0.94$; right PPA $F(3, 96)=0.26$, $p=0.85$; left FFA $F(3, 104)=0.12$, $p=0.95$; right FFA $F(3, 96)=0.14$, $p=0.94$; left OFA $F(3, 88)=0.29$, $p=0.83$; right OFA $F(3, 88)=0.08$, $p=0.97$; left eba $F(3, 104)=0.22$, $p=0.88$; right eba $F(3, 104)=0.16$, $p=0.92$; left pfus $F(3, 104)=0.12$, $p=0.95$; right pfus $F(3, 104)=0.05$, $p=0.99$).

It remains plausible that the paired stimulus is reinstated at feedback time. We therefore conducted the same analysis with the paired stimulus but did not find evidence for reactivation at feedback on the category-selective ROIs (See Supplementary Figure 2). Taken together, we do not find evidence for category-selective reactivation of the visual category of choice stimulus or paired stimulus during feedback.

Discussion

We have shown how the human brain functions to enable flexible credit assignment to both observable choices and latent task-relevant stimuli in an uncertain environment. First, we identified a network of regions responsible for updating beliefs about stimulus-to-outcome contingencies during feedback. We highlighted that VTA signals during this period predict individual differences in the ability to assign credit to the unobservable stimuli that are critical

for task performance. We found that within the OFC, lOFC contains a reinstatement of choice at the receipt of the reward. This coincides with the cOFC and the amygdala, which encode the specific identity of the choice stimulus. In parallel, we demonstrated the existence of a reinstatement of the unobservable stimulus in the lFPC, to which credit is being assigned through inference. During this time, the vlOFC reinstates a code containing information about the identity of the unobservable stimulus.

To create an environment with changing dynamics where participants could learn from latent causal stimuli, we designed a structured probabilistic learning task. Many studies investigating flexible behaviors used reversal-learning tasks (Boorman et al., 2016; Hampton et al., 2008; O'Doherty et al., 2003; Tsuchida et al., 2010-12-15; Witkowski et al., 2022), yet this type of task only demands a specific type of learning, in which habitual stimulus-to-outcome contingencies are formed, then all associations are simultaneously inverted during a reversal (McDannald et al., 2014). Our novel task requires the learning of gradually changing stimulus-to-reward probabilities rather than sudden, concurrent changes to all factors. As a result, participants needed to continuously use the task structure to update their beliefs about contingencies, thereby avoiding reliance on habit-like strategies often employed in reversal-learning tasks. We found the best-fit model to be a “perfect inference” model, which assumes that participants updated their beliefs about the likelihood of the unobserved stimulus leading to an outcome equally well as they did for the choice stimulus. This indicates that participants were able to utilize the learned task structure (the linked pairs) in forced-choice trials to make inferences flexibly in the free-choice trials. It has been shown that primal reinforcers tend to elicit higher intensities in certain reward circuitries (Sescousse et al., 2013). Therefore, we

chose food as the rewarding outcome instead of monetary rewards. Our findings extend beyond previous studies by showing the ability to use the task structure to learn about the inferred stimuli, whereas an earlier study demonstrated the ability to learn through inferred outcomes (Witkowski et al., 2022).

We advanced the understanding of a general network of regions involved in updating choice-outcome contingencies by demonstrating that the regions also support the learning of inferred stimuli using a model of the task structure. These regions can be divided into two subnetworks with complementary functions according to signed DKL upon receiving an outcome, aligning with the direction of belief updates: one towards the confirmation (positive) of complete learning of estimates regarding stimuli-outcome associations (DKL), and another (negative) that reevaluates and updates those estimates. The former subnetwork includes the OFC/mPFC, hippocampus, and posterior cingulate, and the latter includes the pre-SMA/dACC and anterior insula. Our results align with previous studies that examined directly experienced stimuli/outcomes and inferred outcomes (Boorman et al., 2016; Iglesias et al., 2013; O'Reilly et al., 2013; Schwartenbeck et al., 2016; Witkowski et al., 2022).

The human dopaminergic midbrain is widely known to signal prediction errors at the receipt of outcomes (D'Ardenne et al., 2008; Eshel et al., 2016; Schultz et al., 1997; Waelti et al., 2001). More recent studies using model-based paradigms have identified a broader function of the midbrain in encoding information about task state and supporting the formation of task-relevant associative relationships (Hamid et al., 2015; Schwartenbeck et al., 2016; Sharpe et al., 2017; Wise, 2004; Witkowski et al., 2022), both of which are critical for forming a cognitive map. We showed that VTA activity levels predict the degree to which individuals use task

structure to update unobservable stimuli. Our findings align with another study (Schwartenbeck et al., 2016) that demonstrated a similar between-subject effect in the VTA and learning. Participants who better tracked contingencies during a model-based task exhibited smaller effect sizes in VTA activation, suggesting a strong correlation between VTA activity and task performance (Schwartenbeck et al., 2016). Notably, in our task, successful inference and updating of paired stimuli demanded more model-based learning compared to the analogous learning between choice and outcome. We observed significant correlations between VTA activity and the integration of choice history through inference but not direct experience. This finding supports the theory that dopamine contributes to model-based learning (de Wit et al., 2011; Wunderlich et al., 2012/08/08), which in part relies on the afferent projections from the prefrontal regions (Dolan & Dayan, 2013), and efferent dopaminergic projections to ventral striatum which determines the degree of goal-directed and model-based learning (Deserno et al., 2015; Yin et al., 2006). The mPFC and OFC directly project to VTA dopamine neurons (Beier et al., 2015), and using TMS to perturb the OFC modulates VTA signaling of identity prediction error, thereby disrupting reward identity learning in a reversal task (Liu et al., 2024). Similarly, OFC lesions in rodents weaken error signaling and reward expectancy signals computed in VTA dopamine neurons (Takahashi et al., 2011). Together, these findings suggest that frontal cortical regions exert top-down control over midbrain neurons, which signal diverse types of learning information, including reward identity, prediction error, and motivational value during learning (Bromberg-Martin et al., 2010). Our results showing that mPFC and OFC function to update beliefs during feedback is consistent with this claim. Along with the frontal cortical regions, the dopaminergic midbrain promotes cognitive flexibility (Floresco et al., 2006; van

Holstein et al., 2011) by specifically signaling information of priorities to task performance, thus plays a critical role in forming cognitive maps (Sharpe et al., 2017; Witkowski et al., 2022).

Previous research has identified the essential role of the OFC, particularly the lOFC, in flexibly forming contingent relationships between outcomes and specific prior stimuli and using these learned associations to guide decision-making (Boorman et al., 2016; Izquierdo et al., 2004; Jocham et al., 2016; Jones et al., 2012; Noonan et al., 2017; Rushworth et al., 2011; Seo & Lee, 2010; Stalnaker et al., 2015; Takahashi et al., 2011; Walton et al., 2010; Wilson et al., 2014), but it remains unclear what precise information the OFC processes during learning to enable this function. During feedback time, we found that the lOFC reinstates a code that instantiates general information about the choice, coinciding with a more specific code in the cOFC and amygdala, where the identity of the choice is decodable. Prior works have shown that the OFC encodes value and identity-specific information about expected outcomes at the time of cue (Howard et al., 2015; Kahnt et al., 2014). These results together suggest that the OFC processes stimulus-related information during feedback and feedback-related information during the stimulus, generating complementary teaching signals to properly assign stimuli to outcomes. Our finding could also explain how perturbation to OFC during feedback results in impaired updating of an internal model (Miller et al., 2022; Miller et al., 2017). Beyond affecting the coding of outcome value (Miller et al., 2022), silencing the OFC during feedback could also impair the reinstatement of choice, therefore disrupting learning signals. Our findings also shed light on the mechanism by which the OFC tracks state-wise information in a cognitive map to enable efficient navigation in structured environments (Stalnaker et al., 2015; Takahashi et al., 2011; Wilson et al., 2014; Witkowski et al., 2022). Specifically, we identified that during

feedback, the lOFC within the OFC contains general information (value, volatility, state, etc.) about a relevant past choice beyond just the identity of the choice stimulus.

BLA and lOFC are anatomically connected (Carmichael & Price, 1995; Lichtenberg et al., 2017; Mcdonald, 1991; Price, 2007) and functionally interdependent in encoding and updating the identities of expected outcomes (Holland & Gallagher, 2004; Pickens et al., 2003; Schoenbaum et al., 1999; Schoenbaum et al., 1998; Schoenbaum et al., 2003; Stalnaker et al., 2007). Several studies have emphasized the significance of BLA-OFC interactions during decision-making prior to outcome delivery. For instance, disruptions to the BLA impair the OFC's capacity to encode outcomes during cue sampling (Lichtenberg et al., 2017; Schoenbaum et al., 2003). The BLA encodes the motivational salience of cues signaling outcome identities (Gore et al., 2015; Pickens et al., 2003; Sias et al., 2024), while the OFC maintains and utilizes this information from BLA to signal behavioral strategies (Pickens et al., 2003; Schoenbaum et al., 1999; Schoenbaum et al., 2003). However, it is unclear how the BLA-OFC circuit supports learning during feedback when beliefs are updated. Our study addressed this gap by demonstrating that the identity of chosen cues can be decoded during feedback in both the BLA and medial OFC. Nuclei in the amygdala contain not only neurons responsive to appetitive or aversive outcomes but also neurons that process conditioned stimuli or cues of all sensory modalities (Gore et al., 2015; Herry et al., 2008; Livneh & Paz, 2012; Schoenbaum et al., 1998). Evidence from a rodent study identified a crucial circuit in the BLA for behavioral responses, consisting of neurons responsive to conditioned stimuli connected with neurons responsive to unconditioned stimuli (Gore et al., 2015). In line with these studies, our findings confirm the BLA's role in processing sensory information about cues and provide novel evidence of its role

during feedback of linking cue information to outcome representation through Hebbian plasticity. Our classifier learned the identity of stimuli, which encompasses information about stimuli's salience, perceptual qualities, and semantics. The BLA's coding of stimulus identity during feedback may include one or more of these aspects and should therefore be interpreted with caution. In future research, it would be valuable to investigate the interactions between the amygdala and OFC during feedback and how their functions may differ in processing stimulus identity.

Another novel finding of our study is identifying the network where task-relevant but unobservable options are reinstantiated during feedback. Our results align with previous research indicating that the FPC is responsive to alternative options besides the primary choice and processes information critical for subsequent task performance (Boorman et al., 2011; Boorman et al., 2009; Daw et al., 2006; Doll et al., 2015; Koechlin et al., 1999; Koechlin & Hyafil, 2007; Lee et al., 2014; Park et al., 2017; Rushworth et al., 2011). In our task, the paired stimulus shares the same statistical properties as the choice option. It is, therefore, to one's advantage for decisions in future trials to update the stimulus-outcome contingency for the paired stimulus. It is worth stressing the characteristics of a paired stimulus: not directly presented in each trial, critical for future performance, and its tracking and updating it requires inference and understanding of the task structure. Evidently, learning the paired stimulus demands more cognitive effort than the choice stimulus. We indicate that the reactivation of the choice option and the inferred option recruit distinct neural circuits during feedback. To our knowledge, our study is the first to demonstrate that the FPC contains a representation during feedback of alternative options that are statistically identical but perceptually different from the primary

choice. This finding highlights the role of the FPC in inference and extends its known role beyond processing counterfactual options (Boorman et al., 2011; Boorman et al., 2009). Additionally, it confirms the FPC's involvement in processing hidden states (Christoff et al., 2003; Yoshida & Ishii, 2006) and reinforces its position high in the hierarchy of prefrontal organization for managing complex information that requires greater cognitive control beyond simple associations (Koechlin & Hyafil, 2007; Lee et al., 2014). We also observe mPFC in the reactivation of inferred stimuli during feedback. Interestingly, this region overlaps with a previous study indicating that mPFC carries a compact representation of the latent cause when assigning credit to inferred outcomes (Witkowski et al., 2022). It would be interesting for future work to examine how the inferred stimuli and the abstract cause are different in neural representations.

In conclusion, we identified distinct learning signals manifested as neural ensembles coding for the reactivation of both choice and latent stimuli. During feedback, the lOFC reinstates general information about the choice, while identity-specific information is processed in the amygdala and cOFC. The FPC and mPFC carry information about latent stimuli requiring inference. Additionally, VTA signals correlate with individual differences in the ability to infer unobservable stimulus-outcome contingencies. These findings offer insights that could inform clinical research on impaired associative inference in neurological and psychiatric conditions, such as schizophrenia.

Methods

Subjects

Twenty-eight participants (18 females; average age = 22.5) were enlisted from the general population surrounding the University of California, Davis. None of the participants disclosed a history of neurological or psychiatric disorders. Participants received two forms of compensation for participation: 1) course credit or monetary payment ($15/hour) for their participation in the experiment; 2) A type of dessert from Trader Joe's. The number of desserts they received was determined by their performance in the task. All procedures were granted approval by the University of California, Davis Institutional Review Board (IRB), and participants provided written consent before participating in the experiment.

Task instruction

Participants completed a "structured probabilistic learning task" in which they tracked associations between 4 stimuli (choices) and food identities (outcomes) to maximize the chance of larger food rewards at the end of the experiment (Figure 2.1). The task has 3 runs in total, with each run composed of 2 sections: 24 forced-choice trials and 48 free-choice trials.

Probabilistic Learning Task

On each forced-choice trial, participants were presented with 2 of 4 stimuli, one after another, separated by a second interstimulus interval (ISI). Then, they were presented with an image of a left and a right arrow key, representing the firstly or secondly presented stimuli, respectively. A red box encircled one of the two arrow keys, indicating which stimulus participants had to select between the two. For example, if Tool and Scene were presented on a forced-choice trial, and the left arrow key is encircled by the red arrow key, participants would press the left arrow key (corresponding to the firstly presented stimulus), knowing that doing so means they have

chosen Tool (Figure 2.1A). Then, a food image was presented as the outcome of the forced-choice. Participants were instructed to learn the association between the chosen stimulus and food type. Here, they would learn that Tool is associated with Macaroon. Over forced-choice trials, participants learn what food type each of the 4 stimuli leads to. They are instructed that 2 stimuli that lead to the same food make up a "Linked Pair" (LINKED PAIR). For example, if Face and Tool both lead to the same food, then these stimuli are an LINKED PAIR. Note that the food type the 2 LINKED PAIRs are associated with would change in the middle of forced-choice trials. That is, Face and Tool lead to Macaroon for the first half of the trials, and then this LINKED PAIR starts leading to Green Tea Mochi. Nevertheless, Face and Tool always lead to the same food type, thus staying as a linked pair by definition.

After the forced-choice trials, participants were tested for their knowledge of the 2 LINKED PAIRs in a "matching task", in which they had to correctly identify the 2 linked pairs from the 4 stimuli. If they mistakenly match any pair, they would be prompted to do the matching task again until they successfully identify the 2 correct pairs. They were instructed to remember the 2 linked pairs within a Run, with the knowledge that the 2 stimuli belonging to the same linked pair share the same drifting probability of leading to either dessert throughout a Run. Participants could therefore infer the probabilistic stimulus-to-outcome association of one stimulus when observing the outcome of choosing the other stimulus in the linked pair.

After the matching task and solid learning of the 2 linked pairs, participants started free-choice trials. Importantly, they were told that a random free-choice trial would be drawn at the end of the task, and their actual food compensation would be equivalent to the outcome on

that trial. Thus, their goal should be to make decisions that would maximize the number of their preferred food rewards received on every trial.

On free-choice trials, one of the four stimuli is presented with 2 numbers of different colors. The color of a number represents the type of food (e.g. green colored number is the number of green tea mochi, pink colored number is the number of pink macaroons), and the number itself represents how many food participants would receive if they get the food on that trial. After an ISI, a second stimulus was presented. Similarly, the colored numbers reappear on top of the stimulus to remind participants of the possible outcomes. Then, participants chose between the two stimuli by pressing the left (the 1st stimulus presented) or the right (the 2nd stimulus) button. After an ISI (4-8 secs), they were presented with a food outcome with the corresponding magnitude.

Critically, each of the stimuli in the same linked pair have a probability p leading to a food outcome, and a probability of 1-p leading to the other food outcome. Similarly, the two stimuli in a different linked pair have a probability of q leading to food outcome1 and 1-q to food outcome2. The probabilities p and q are independent of one another; therefore, participants could not infer the reward probability for one linked pair using another. This ensures two independent systems of learning the stimulus-to-outcome associations. The stimulus-to-outcome probabilities p and q, refer to the type of food outcomes, not to the number/magnitude of food outcomes. In the instructions, experimenters made it clear that the number of food outcomes was random on every free-choice trial, thus were uncorrelated with reward probabilities. The purpose of the food numbers was to motivate the participants to aim for the food type with the higher number on every trial. The colored numbers were located

vertically, one on top of another. The location of the 2 numbers was random on each trial, thus did not indicate stimulus-to-outcome probabilities in any way. In the instructions, participants were told that p and q are drifting probabilities, meaning that they vary in time. They were also informed that the drifting rates of p and q changed across the task. 5) On a trial, the 2 stimuli presented always come from 2 different linked pairs, and never from the same linked pair. Therefore, participants always had to compare the stimulus-to-outcome probabilities of 2 different systems. Depending on which food had a larger number, participants chose one of the two stimuli presented on the trial with a higher estimated probability leading to that food type during a time frame.

On each new run, the stimuli that comprised each linked pair changed, so participants had to learn which stimuli were linked. For example, in run 1, Face and Tool formed an linked pair, Body and Scene formed the other linked pair; in Run 2, Face and Scene formed an linked pair, Body and Tool formed the other. Thus, participants began with no prior information about the contingencies in each run. Taken together, the 3 runs are independent of each other in terms of linked pair and stimulus-to-outcome associations, two critical components of participants' learning and task performance.

Stimuli

Four images representing distinct categories (Face, Body part, Scene, Tool) were designated as stimuli for all participants to choose from. The pairings of images linked pair 1 and those for linked pair 2 were randomized across three Runs, ensuring coverage of all possible combinations. For instance, in Run 1, linked pair 1 = Face and Tool, linked pair 2 = Scene and

Body; in Run 2, linked pair 1 = Face and Body, linked pair 2 = Tool and Scene; in Run 3, linked pair 1 = Face and Scene, while linked pair 2 = Body and Tool.

Two types of reward identities were selected from 10 different Trader Joe's desserts. Before the task, participants rated their preferences for these desserts on a scale from 1 (least desired) to 5 (most desired). The desserts included: 1 "brownie crisp coffee ice cream sandwiches", "a dozen sweet bites-chocolate/coffee opera cake", "A dozen macarons variés","ice cream bon bons with chocolate cookie crust", "mochi ice cream (green tea)", "gluten-free Joe-Joe's Chocolate Vanilla Creme Cookies", "cocoa-filled pinwheel cookies", "coffee mochi", "mango & cream bars" and "strawberry mochi".

Two reward identities were selected for each participant based on their preference ratings before the fMRI experiment. Specifically, two desserts with the smallest difference among the four most highly preferred options were chosen. These desserts were designated as outcome1 (O1) and outcome2 (O2), respectively, in the fMRI experiment. This procedure was employed to minimize potential biases stemming from individual dessert preferences during the probabilistic learning task while also considering the desirability associated with each outcome.

Task-schedule and procedure

Two distinct schedules were generated, with half of the participants (N=14) assigned to perform the task on schedule 1, while the remaining half followed schedule 2. The schedules determined what choice options (images) would be presented on every trial and what

outcomes (food) would be associated with each choice. Note that on every trial, the 2 choice options were always from different linked pairs. We designed the schedules in a way that an optimal Bayesian learner (perfect inference model; See Computational models) would choose each image and receive each outcome approximately equally, given an equal preference between the food identities. This promotes a nearly equal sample size for choices and outcomes across participants and minimizes sampling bias. Importantly, each run is independent, where participants learned the task structure and probabilistic choice-to-outcome associations without any information from previous runs. The schedule for each run was, therefore, independent from one another. Each schedule had predetermined probabilistic choice-to-outcome associations (p, q) that were independent of one another. In other words, participants were not able to learn the choice-outcome association of one linked pair from that of another.

Participants completed three runs of 48 trials (144 trials total). Before the fMRI experiment, participants were asked to fast for 2 hours, which helped motivate their desire to obtain food. They were instructed that at the completion of the task, one trial would be randomly chosen to count as a "real" trial and would be used to determine the food reward they received. As a result, participants need to make optimal decisions on each trial to maximize their rewards. At the end of the experiment, we randomly selected one trial and rewarded participants with food of the magnitude of that trial.

Behavioral training

To acquaint participants with the task, all subjects underwent a behavioral training session preceding the fMRI experiment. Following behavioral training, participants performed the fMRI experiment on a separate day within 2 weeks. The task in the behavioral training was identical to that in the fMRI experiment, with the only difference being the schedule used. If a participant did the task using schedule1, then schedule2 was used for their fMRI experiment, and vice versa.

After participants went through the instructions, they answered questions about the rules of the task structure. The experimenter ensured that participants had a solid understanding of the task by confirming and correcting their answers as needed. Before the behavioral task consisting of 3 runs, participants did a separate practice run made of forced-choice trials and free-choice trials of a different schedule than the main task. The goal of the practice run was to familiarize participants with the task. After the practice run, the experimenter checked with participants to answer any questions they had about the task. Then, participants completed the behavioral training runs independently. To verify that participants learned the 2 linked pairs after the forced-choice trials, they performed a "matching task" which required the correct grouping of the 4 stimuli into 2 pairs. Feedback was provided informing participants if they grouped the 2 linked pairs correctly. If incorrect, participants were presented with the matching task again until they performed it correctly. Participants demonstrating a good understanding of the task and achieving satisfactory performance were invited to participate in the fMRI experiments. Out of the initial 37 participants enrolled, 28 participants in the fMRI experiment.

Computational models

Perfect-inference learning model

We employed a Bayesian computational model to predict participants' choices on each trial (t) by considering the history of outcomes for each previous choice, the currently available choice options, and the magnitude of the rewards being offered. While a concise overview of the model is provided here, a comprehensive description can be referenced in (Behrens et al., 2007). For a related model, refer to (Arulampalam et al., 2002).

As the actual probability of the associative contingencies remains unobservable, the model employed a Markovian approach to estimate the subjective belief that selecting a particular choice (C) would result in outcome1 (O1), and lead to outcome2 (O2) with the inverse probability. This is denoted as follows:

$$p(S \rightarrow O1) = q_s$$
$$p(S \rightarrow O2) = 1 - q_s$$

(equation 1)

where qs denotes the probability of receiving outcome O1 given choice S. On every trial, the model estimates a prior belief in of qst based on observations from previous outcomes. The choice-to-outcome contingencies range from 0 to 1, and their likelihood is modeled as a beta distribution over possible value.

$$\beta(q_{st} \mid V)$$

(equation 2)

The beta distribution has a mean of qst and a variance of V=exp(v). v decides how likely qst the next trial would change compared to the current trial. A large v suggests that qst the following trial will be more likely to change, and a small v suggests that qst would be relatively like that of the current trial. v is therefore named as the "volatility" parameter due to its control of the extent to which observed outcomes should update the choice-to-outcome contingencies on every trial.

This volatility parameter of choice-to-outcome contingencies V changes on a trial according to the following equation:

$$p(v_t \mid v_{t-1}, k) = N(v_{t-1} \mid k)$$

(equation 3)

K denotes the rate of change of volatility v. In other words, k is the control parameter on the beliefs about how likely it is that the change rate will itself change. The distribution of v is modeled as Gaussian in a Markovian fashion.

After every observation, the model updates the choice-to-outcome contingencies in a Markovian fashion, meaning that the posterior probability is only dependent on the belief about the latest state of the contingency qs, t-1, and the volatility vt-1. The prior qs, t-1, captures the most likely state given it integrates the full history of previous choices and outcomes without the need to store the history of the environment. Therefore, the estimate of the parameters could be updated following the Bayes rule:

$$p(q_{s,t},\ v_t,\ k \mid y_{1:t}) \ = \ p(y_t \mid q_{s,t}) \int \left[\int p(q_{s,t-1}, v_{t-1}|y_{1:t-1})p(v_t|v_{t-1})dv_{t-1}\right]\ p(q_{s,t}|q_{s,t-1}, vt)\ dq_{s,t-1}$$

(equation 4)

This gives us a 3-dimensional grid of the joint probability distribution that is stored between trials. At the beginning of each run, the joint probability distribution is set to be uniform, because participants had no information about the associations or rates of change. On every trial, the Bayesian learner needs to use the estimated choice-to-outcome contingency to compute the expected value, which is found by marginalizing over v and k.

$$p(q_{s,t}) \ = \ \int\int p(q_{s,t},\ v_t,\ k)\ dv_t\, dk$$

(equation 5)

And then taking the mean of the distribution:

$$\overline{q_{s,t}} \ = \ \int q_{s,t}\, p(q_{s,t}) dq_{s,t}$$

(equation 6)

For every participant, we began the model with a uniform prior encompassing the complete parameter space at the start of the task. Numerical grid integration was employed for all integral calculations. Subsequently, we utilized beliefs in associative contingencies to determine the expected value of each stimulus during each trial according to the following equation:

$$EV_{s,t} = \overline{q_{s,t}} m_{o1,t}\alpha + (1 - q_{s,t}) m_{o2,t}a^{-1}$$

(equation 7)

Where a represented a free parameter signifying a subject's preference for O1 over O2 (with 0 < alpha < 2), and mo1,t mo2,t indicated the reward magnitudes of the 2 different outcomes available in the current trial t. Subsequently, we assessed the likelihood of each participant's choice on each trial using a SoftMax function.

$$p(choice = s) = \frac{e^{bEV_{s,t}}}{e^{bEV_{s,t}} + e^{bEV_{c,t}}}$$

(equation 8)

S and C are the two stimuli on a trial, and b is the inverse temperature parameter that indicates the level of sensitivity of choices to expected values. Free parameters were fitted using Markov Chain Monte Carlo methods (See below).

Parameter estimates

The learning model involves two free parameters, a and b. We independently estimated these parameters for each subject using custom Markov Chain Monte Carlo (MCMC) code in MATLAB R2018a. The parameter constraints were set as follows: 0<a<2 and 0<b<7. Initial values were set at = 1 (a=1 implies no preference for either choice) and b=3.5. The model was then fitted to maximize the likelihood of a subject's choices, considering the model's estimates of the expected value for each choice on every trial (Eq. 8).

Alternative models

We tested the perfect-inference learning model against two additional models: the weighted-inference learning model and the no-inference learning model. In the weighted-inference learning model, an inference learning parameter (0<gamma<inifity) was included to capture the degree of belief update for the inferred choice-to-outcome contingencies in the same Linked Pair. When an outcome is observed, participants not only update their beliefs of the contingency of the directly chosen stimulus but also that of the inferred paired stimulus. In the perfect-inference learning model, was set to 1, meaning participants updated the paired stimulus-to-outcome associations equally well as the choice-to-outcome associations. The no-inference learning model assumes no inferred learning within a linked pair, therefore was set to 1. According to the Bayes rule, the posterior belief qs is computed as the following:

$$p(q_{c,t}|\, y_{1:t}) = \gamma\, p(q_{s,t}|y_{1:t}) + (1 - \gamma)p(q_{c,t-1}|\, y_{1:t-1}) \quad \text{(equation 9)}$$

Note that C and S are two stimuli within a linked pair, and that S is the choice stimulus on trial t, and C is not shown on trial t, therefore needs to be learned through S.

Model comparisons

We compared the goodness of fit for each model type by calculating the sum of the Bayesian Information Criterion (BIC) across subjects. This provided an overall measure of how well the data were fitted by each model at the group level while penalizing models that introduced additional free parameters.

The BIC is calculated using the formula:

$$BIC = k \bullet ln(n) - 2ln\widehat{L}$$

(equation 10)

where k is the number of parameters in the learning model, n is the number of choices (i.e., trials) made by the subject, and lnL represents the log-likelihood of each model.

Model-free analysis of choice history on decisions

To assess whether participants showed behavioral learning effects from their choices, we employed logistic regression models to estimate the impact of past choice-outcome observations on the current trial t. The model considered both the experienced choice-outcome association from three trials back and the inferred paired stimulus-outcome relationships from three trials back. The logistic regression equation is expressed as follows:

$$ln(\frac{p(chosen)}{p(unchosen)}) = \beta 0 + \beta 1 w_t O_{tE-1} + \beta 2 w_t O_{tE-2} + \beta 3 w_t O_{tE-3} + \beta 4 w_t O_{tI-1} + \beta 5 w_t O_{tI-2} + \beta 6 w_t O_{tI-3}$$

(equation 11)

Here, n represents the nth previous trial when the image was chosen, with consideration given to up to three previous trials. E denotes the effect of an experienced choice-outcome pairing, and I denotes the effect of an inferred choice-outcome pair. For instance, tE-1 refers to the outcome directly experienced the last time the current image was chosen. The same notation applies to the outcome of the previously paired stimulus. Since participants were presented with two choice options among four images in each trial, the chosen option in the current trial might not have been available in the previous trial. Thus, tE-1 or tI-1 was the last trial when the current choice or the paired stimulus was chosen

respectively. Separate regression models were fitted for the choices of each of the four images for each subject.

$$O_{tE-n} = \{1 \ if \ S_{tE-n} led \ to \ the \ desired \ outcome; \ -1 \ if \ S_{tE-n} led \ to \ the \ undesired \ outcome$$

We included contextual, counterfactual details about the alternative option into the experienced regressors. For example, if a participant chose option C over option A, and obtained the desired outcome, this might discourage them from selecting image A in a subsequent trial when both A and C are available. This inclusion was made for comprehensive coverage of all experienced information that could impact the choice of a shape in any given trial.

For the effect of the paired stimulus, the regressor had a value of 1 if the paired stimulus led to the desired outcome n trials back, such that

$$O_{tI-n} = \{1 \ if S_{tI-n} led \ to \ the \ desired \ outcome; \ -1 \ if \ S_{tI-n} led \ to \ the \ undesired \ outcome$$

To account for the effect of the relative desirability of the rewards on the current trial, the difference in reward magnitude (wt = |MtO1-MtO2|) was assigned as a weight on each regressor.

After fitting individual regression models for each run, we averaged the regression coefficients across conditions, representing the subject-specific impact of prior decisions on the current choice.

MRI data acquisition

We obtained T2-weighted functional images using a Siemens Skyra 3 Tesla scanner. The gradient-echo-planar imaging (EPI) pulse sequence was employed with a slice angle of 30 degrees relative to the anterior-posterior commissure line, minimizing signal loss in the orbitofrontal cortex region (Weiskopf et al., 2006). The acquisition involved 46 slices, each 3mm thick, with the following parameters: repetition time (TR) = 1300 ms, echo time (TE) = 24 ms, flip angle = 67 degrees, field of view (FoV) = 192mm, and voxel size = 3 x 3 x 3 mm^3. Contiguous slices were acquired in an interleaved order. For deformation correction, a field map with dual echo-time images covering the entire brain was also acquired, featuring parameters: TR = 630 ms, TE1 = 10 ms, TE2 = 12.46 ms, flip angle = 40 degrees, FoV = 192mm, and voxel size = 3 x 3 x 3 mm^3. Additionally, a T1-weighted structural image was acquired using a magnetization-prepared rapid gradient echo sequence (MPRAGE) with the following parameters: TR = 1800 ms, TE = 2.96 ms, flip angle = 7 degrees, FoV = 256mm, and voxel size = 1 x 1 x 1 mm^3.

Preprocessing

The data preprocessing was performed in SPM12 (Wellcome Trust Centre for Neuroimaging) within MATLAB (2018b, MathWorks). The default options in SPM were employed for preprocessing. The images underwent slice-time correction and were realigned to the initial volume of each sequence to correct for motion using a six-parameter rigid body transformation. Correction for field inhomogeneities was performed using the phase of non-EPI gradient echo images at two echo times, which were co-registered with structural maps. Subsequently, the images were spatially normalized by warping subject-specific images to the reference brain in the MNI (Montreal Neurological Institute) coordinate system with 2mm

isotropic voxels. For univariate analysis, the images were finally spatially smoothed using a Gaussian kernel with a full width at half maximum of 8mm.

Univariate fMRI analysis

To model BOLD activity in every voxel, we used a GLM with 8 different regressors: the first stimulus in forced-choice trials (a 2s boxcar), the second stimulus in forced-choice trials (a 2s boxcar), and the reward feedback phase in forced-choice trials (a 2s boxcar). For the free-choice trials, we included the onset of the first choice stimulus (a boxcar, from the stimulus onset including the duration of 2s), an ISI (a box car with a jittered duration between 1-4s), the second stimulus (a box car, from the stimulus onset including the duration of 2s), the reward expectation period (a boxcar including jittered ISI 4-8s), the reward feedback phase (a 2s boxcar). In the GLM, we included the model-based belief updates to the choice-outcome associations after the outcome was observed and used this as a parametric regressor at the feedback phase. This belief update was quantified as the Kullback-Leibler divergence (DKL) between the prior and posterior belief in qs,t for the chosen image S.

$$D_{KL}(t) = \int ln(\frac{p(q_{s,t}|\, y_{1:t})}{p(q_{s,t-1}|\, y_{1:t-1})})p(q_{s,t}|\, y_{1:t})dq_{s,t}$$

The DKL reflected changes in the model estimated ‘‘beliefs’’ about which choice led to which outcome (food type) as participants progressed through learning. The DKL for each image was

determined through numerical integration, and subsequently, the DKL values for images within the same Linked Pair were aggregated by summation to produce DKLSum. Six motion regressors were incorporated into the model as nuisance regressors to control for translation and rotation in head position throughout the experiment. In the first-level analysis, contrast images of parameter estimates derived from the regressors of DKLSum were calculated for each participant and then utilized as input for the one-sample t-test in the second-level analysis.

Group-level statistical inference

Group-level testing was conducted through a one-sample t-test (df = 27) on the cumulative functional maps generated by the first-level analysis. All first-level maps were smoothed before being aggregated and examined at the group level. To address multiple comparisons, Threshold-Free Cluster Enhancement (TFCE) was used with permutation testing to consider both the height and extent of the cluster (Smith and Nichols, 2009). Default parameters (H=2, E=0.5) were applied, and 5000 permutations were used. For ROI analyses, we initially extracted voxels from each ROI in each subject's first-level activation map, averaged the maps, and then applied small-volume TFCE correction. We report effects that exceeded a pTFCE < .05 threshold in each ROI.

Multivariate Analyses

Choice representation at feedback

The MVPA analysis aimed to identify regions in the brain that contain a choice representation at the reward feedback phase. We began by estimating BOLD activity during the reward feedback

period of each trial independently, by creating a GLM with a regressor for the onset of each trial's feedback period. The feedback period was modeled as a boxcar function with a consistent duration of 2 seconds from the start of the feedback on each trial. There were no additional parametric modulators included.

Every trial was labeled according to which image was chosen, St. We hypothesize that during the reward feedback phase, there would be a representation somewhere in the brain of the choice made on that trial. If such a choice representation exists, then we could expect to decode the choice label (Face, Body, Scene, Tool) at the time of the reward feedback phase, regardless of what food outcome they receive. For example, we expect to be able to decode representations for Face at feedback on trials where Face was chosen, despite what type of food was received at feedback. In other words, free-choice trials on all 3 runs were grouped according to the choice label. Then, using a cross-validation approach, we trained the support vector machine (SVM, linear classifier with a cost of 100) on 2 runs during feedback and tested on the left-out run during feedback.

Our search light procedure employed a 5x5x5 voxel cube centered around a centroid voxel. The centroid voxel needed to have values in at least 10 of the neighboring voxels to be considered for further processing. The activity within each searchlight was standardized by z-scoring the beta values across voxels. The data were then split into training and test sets by block. Linear classifiers were fitted using LIBSVM (Chang & Lin, 2011) with the training data, and these classifiers were applied to classify data points from the test set. This process repeated until each block was used as a test set, and the mean decoding accuracy (average proportion of correct classifications) was then averaged across iterations. The mean decoding accuracy for

each voxel was compared to a voxel-specific null distribution, estimated by repeating the procedure while randomly assigning labels for 100 permutations at each searchlight. The mean classification accuracy of this null distribution was subtracted from the classification accuracy of each searchlight, providing a measure of how reliably information about choice representation could be decoded above chance. The resulting maps were then spatially smoothed using a Gaussian kernel with a full width at half maximum of 8mm.

Identity code of causal choice

To test for the reinstatement of the identity of casual choice stimulus, we trained an SVM with the same parameters as the previous analysis (linear classifier with a cost of 100, using 100 permutations to verify there was no bias and test for significance of decoding accuracy). In this analysis, we trained the SVM using the BOLD signals with the unsmoothed preprocessed images during the stimuli onset in forced-choice trials. Training data when participants were viewing the stimuli which was just the image itself, without any additional information. This was modeled as a boxcar with 2 seconds from the onset of the stimulus presentation. No parametric modulators were added. Each forced-choice trial contains 2 image stimuli, and each stimulus onset was labeled for the image identity and used subsequently in the SVM training. Testing was performed during the feedback period, modeled as a boxcar for 2 seconds, starting when participants saw the food. Training and testing were conducted within a run, and decoding accuracy was averaged across 3 runs after we removed the decoding accuracy from the permutations with shuffled labels.

Inferred stimulus representation

To test if there exists a representation of the paired stimulus when an outcome is received, we trained an SVM with the same parameters as previous analyses (linear classifier with a cost of 100)). The main difference with the analysis on choice representation at feedback is that the training and testing labels are not based on choice but paired stimulus of the choice stimulus. For example, a trial would be labeled as "Tool" if participants had chosen Face on a free-choice trial and that Face and Tool were a Linked Pair on that run. This analysis was done with a cross-validation approach, meaning that training was done using 2 runs, and the classifier was tested on the left-out run. We trained and tested the SVM using the BOLD activity patterns with the unsmoothed preprocessed images during feedback, which was modeled as a boxcar for 2 seconds. Critically, this analysis eliminates the choice image label, resulting in only 3 labels for training and testing. This guarantees that the choice image label does not act as a potential confound with the inferred/paired stimulus or the variable of interest.

Identity code of paired stimulus

This analysis was conducted in a similar process to the analysis of the identity code of the causal choice. Instead of testing the reinstatement of choice stimulus identity, we also expect to decode the identity of the paired stimulus at feedback. On a free-choice trial, the SVM used to test feedback was trained at forced-choice trial onsets for the paired stimulus identity. For example, on a free-choice trial where Face was chosen and Tool was the paired stimulus to Face on that run, we would use the SVM trained at forced-choice to test for the correct label "Tool" at the free-choice feedback. The choice label was eliminated in this analysis to avoid potential confounds with the paired stimulus. Therefore 3 labels were used for training and testing.

Localizer model and ROI selection

The localizer design aimed to identify neural responses linked to four specific visual categories. Following the main experiment, participants performed the localizer task with 12 blocks. Each block comprised stimuli from one of four categories: faces, body parts, tools, and scenes, with each category having three blocks. Within each block, 20 stimuli of the same category were presented for 300ms, separated by a 400ms Inter-Stimulus Interval (ISI). Participants were instructed to attend to the stimuli and respond to the presence of a red dot on a stimulus (10% of images, randomly interleaved) by a button press.

We conducted a General Linear Model (GLM) analysis on the time series from the 12 localizer runs. For each of the four stimulus categories, the model included separate 14-second boxcar regressors for every block. The GLM also incorporated six motion regressors to account for translation and rotation in the head position. To identify Regions of Interest (ROIs) most sensitive to each stimulus category, we contrasted each category with the other three in the temporal and occipital lobes. Each ROI was obtained in the subject's individual native brain space (See Supplementary Figure 2.3 for a group-level representation of these ROIs).

References

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M.,…Halgren, E. (2006). Top-down facilitation of visual recognition. Proceedings of the National Academy of Sciences, 103(2). https://doi.org/10.1073/pnas.0507062103

Bar, M., Tootell, R. B., Schacter, D. L., Greve, D. N., Fischl, B., Mendola, J. D.,…Dale, A. M. (2001). Cortical mechanisms specific to explicit visual object recognition. Neuron, 29(2). https://doi.org/10.1016/s0896-6273(01)00224-0

Beier, K. T., Steinberg, E. E., DeLoach, K. E., Xie, S., Miyamichi, K., Schwarz, L.,…Luo, L. (2015). Circuit Architecture of VTA Dopamine Neurons Revealed by Systematic Input-Output Mapping. Cell, 162(3). https://doi.org/10.1016/j.cell.2015.07.015

Boorman, E. D., Behrens, T. E., & Rushworth, M. F. (2011). Counterfactual Choice and Learning in a Neural Network Centered on Human Lateral Frontopolar Cortex. PLOS Biology, 9(6). https://doi.org/10.1371/journal.pbio.1001093

Boorman, Erie D., Behrens, T. E. J., Woolrich, M. W., & Rushworth, M. F. S. (2009). How Green Is the Grass on the Other Side? Frontopolar Cortex and the Evidence in Favor of Alternative Courses of Action. Neuron, 62(5). https://doi.org/10.1016/j.neuron.2009.05.014

Boorman, Erie D., O'Doherty, John P., Adolphs, R., & Rangel, A. (2013). The Behavioral and Neural Mechanisms Underlying the Tracking of Expertise. Neuron, 80(6). https://doi.org/10.1016/j.neuron.2013.10.024

Boorman, E. D., Rajendran, Vani G., O'Reilly, Jill X., & Behrens, T. E. (2016). Two Anatomically and Computationally Distinct Learning Signals Predict Changes to Stimulus-Outcome Associations in Hippocampus. Neuron, 89(6). https://doi.org/10.1016/j.neuron.2016.02.014

Boorman, E. D., Witkowski, P. P., Zhang, Y., & Park, S. A. (2021). The orbital frontal cortex, task structure, and inference. Behavioral Neuroscience, 135(2). https://doi.org/10.1037/bne0000465

Bouret, S., & Richmond, B. J. (2010). Ventromedial and Orbital Prefrontal Neurons Differentially Encode Internally and Externally Driven Motivational Values in Monkeys. Journal of Neuroscience, 30(25). https://doi.org/10.1523/JNEUROSCI.0049-10.2010

Bracci, S., Cavina-Pratesi, C., Ietswaart, M., Caramazza, A., & Peelen, M. V. (2012). Closely overlapping responses to tools and hands in left lateral occipitotemporal cortex. Journal of Neurophysiology, 107(5). https://doi.org/10.1152/jn.00619.2011

Bromberg-Martin, E. S., Matsumoto, M., & Hikosaka, O. (2010). Dopamine in Motivational Control: Rewarding, Aversive, and Alerting. Neuron, 68(5). https://doi.org/10.1016/j.neuron.2010.11.022

Carmichael, S. T., & Price, J. L. (1995). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. Journal of Comparative Neurology, 363(4). https://doi.org/10.1002/cne.903630408

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST), 2(3), 1-27. https://doi.org/10.1145/1961189.1961199

Chao, L. L., Haxby, J. V., Martin, A., Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. Nature Neuroscience, 2(10). https://doi.org/10.1038/13217

Christoff, K., Ream, J. M., Geddes, L. P. T., & Gabrieli, J. D. E. (2003). Evaluating self-generated information: anterior prefrontal contributions to human cognition – PubMed. Behavioral neuroscience, 117(6). https://doi.org/10.1037/0735-7044.117.6.1161

D'Ardenne, K., McClure, S. M., Nystrom, L. E., & Cohen, J. D. (2008). BOLD Responses Reflecting Dopaminergic Signals in the Human Ventral Tegmental Area. Science, 319(5867). https://doi.org/10.1126/science.1150605

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., Dolan, R. J., Daw, N. D.,…Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. Nature 2006 441:7095, 441(7095). https://doi.org/10.1038/nature04766

Daw, N. D., & Tobler, P. N. (2014). Value Learning through Reinforcement: The Basics of Dopamine and Reinforcement Learning. Neuroeconomics. https://doi.org/10.1016/B978-0-12-416008-8.00015-2

de Wit, S., Standing, H. R., DeVito, E. E., Robinson, O. J., Ridderinkhof, K. R., Robbins, T. W.,…Sahakian, B. J. (2011). Reliance on habits at the expense of goal-directed control following dopamine precursor depletion. Psychopharmacology 2011 219:2, 219(2). https://doi.org/10.1007/s00213-011-2563-2

Deserno, L., Huys, Q. J. M., Boehme, R., Buchert, R., Heinze, H.-J., Grace, A. A.,…Schlagenhauf, F. (2015). Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. Proceedings of the National Academy of Sciences, 112(5). https://doi.org/10.1073/pnas.1417219112

Dolan, Ray J., & Dayan, P. (2013). Goals and Habits in the Brain. Neuron, 80(2). https://doi.org/10.1016/j.neuron.2013.09.007

Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D., & Daw, N. D. (2015). Model-based choices involve prospective neural activity. Nature Neuroscience 2015 18:5, 18(5). https://doi.org/10.1038/nn.3981

Doll, B. B., Jacobs, W. J., Sanfey, A. G., & Frank, M. J. (2009). Instructional control of reinforcement learning: A behavioral and neurocomputational investigation. Brain Research, 1299. https://doi.org/10.1016/j.brainres.2009.07.007

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A Cortical Area Selective for Visual Processing of the Human Body. Science, 293(5539). https://doi.org/10.1126/science.1063414

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The Parahippocampal Place Area. Neuron, 23(1). https://doi.org/10.1016/S0896-6273(00)80758-8

Eshel, N., Tian, J., Bukwich, M., & Uchida, N. (2016). Dopamine neurons share common response function for reward prediction error. Nature Neuroscience 2016 19:3, 19(3). https://doi.org/10.1038/nn.4239

Floresco, S. B., Magyar, O., Floresco, S. B., & Magyar, O. (2006). Mesocortical dopamine modulation of executive functions: beyond working memory. Psychopharmacology 2006 188:4, 188(4). https://doi.org/10.1007/s00213-006-0404-5

Gallagher, M., McMahan, R. W., & Schoenbaum, G. (1999). Orbitofrontal Cortex and Representation of Incentive Value in Associative Learning. Journal of Neuroscience, 19(15). https://doi.org/10.1523/JNEUROSCI.19-15-06610.1999

Gershman, S. J., Uchida, N., Gershman, S. J., & Uchida, N. (2019). Believing in dopamine. Nature Reviews Neuroscience 2019 20:11, 20(11). https://doi.org/10.1038/s41583-019-0220-7

Gore, F., Schwartz, E. C., Brangers, B. C., Aladi, S., Stujenske, J. M., Likhtik, E.,…Axel , R. (2015). Neural Representations of Unconditioned Stimuli in Basolateral Amygdala Mediate Innate and Learned Responses. Cell, 162(1). https://doi.org/10.1016/j.cell.2015.06.027

Hamid, A. A., Pettibone, J. R., Mabrouk, O. S., Hetrick, V. L., Schmidt, R., Vander Weele, C. M.,…Berke, J. D. (2015). Mesolimbic dopamine signals the value of work. Nature Neuroscience 2016 19:1, 19(1). https://doi.org/10.1038/nn.4173

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. Proceedings of the National Academy of Sciences, 105(18). https://doi.org/10.1073/pnas.0711099105

Herry, C., Ciocchi, S., Senn, V., Demmou, L., Müller, C., Lüthi, A.,…Lüthi, A. (2008). Switching on and off fear by distinct neuronal circuits. Nature, 454(7204). https://doi.org/10.1038/nature07166

Holland, P. C., & Gallagher, M. (2004). Amygdala-frontal interactions and reward expectancy – PubMed. Current opinion in neurobiology, 14(2). https://doi.org/10.1016/j.conb.2004.03.007

Howard, J. D., A., G. J., Tobler, P. N., & Kahnt, T. (2015). Identity-specific coding of future rewards in the human orbitofrontal cortex. Proceedings of the National Academy of Sciences, 112(16). https://doi.org/10.1073/pnas.1503550112

Howard, J. D., & Kahnt, T. (2017). Identity-Specific Reward Representations in Orbitofrontal Cortex Are Modulated by Selective Devaluation. Journal of Neuroscience, 37(10). https://doi.org/10.1523/JNEUROSCI.3473-16.2017

Howard, J. D., Kahnt, T., Howard, J. D., & Kahnt, T. (2018). Identity prediction errors in the human midbrain update reward-identity expectations in the orbitofrontal cortex. Nature Communications 2018 9:1, 9(1). https://doi.org/10.1038/s41467-018-04055-5

Iglesias, S., Mathys, C., Brodersen, K. H., Kasper, L., Piccirelli, M., den Ouden, H. E., & Stephan, K. E. (2013). Hierarchical Prediction Errors in Midbrain and Basal Forebrain during Sensory Learning. Neuron, 80(2). https://doi.org/10.1016/j.neuron.2013.09.009

Ito, M., & Doya, K. (2009). Validation of Decision-Making Models and Analysis of Decision Variables in the Rat Basal Ganglia. Journal of Neuroscience, 29(31). https://doi.org/10.1523/JNEUROSCI.6157-08.2009

Izquierdo, A., Suda, R. K., & Murray, E. A. (2004). Bilateral Orbital Prefrontal Cortex Lesions in Rhesus Monkeys Disrupt Choices Guided by Both Reward Value and Reward Contingency. Journal of Neuroscience, 24(34). https://doi.org/10.1523/JNEUROSCI.1921-04.2004

Jocham, G., Brodersen, Kay H., Constantinescu, Alexandra O., Kahn, Martin C., Ianni, A. M., Walton, Mark E.,…Behrens, T. E. (2016). Reward-Guided Learning with and without Causal Attribution. Neuron, 90(1). https://doi.org/10.1016/j.neuron.2016.02.018

Jones, J. L., Esber, G. R., McDannald, M. A., Gruber, A. J., Hernandez, A., Mirenzi, A., & Schoenbaum, G. (2012). Orbitofrontal Cortex Supports Behavior and Learning Using Inferred But Not Cached Values. Science, 338(6109). https://doi.org/10.1126/science.1227489

Kahnt, T., Park, S. Q., Haynes, J.-D., & Tobler, P. N. (2014). Disentangling neural representations of value and salience in the human brain. Proceedings of the National Academy of Sciences, 111(13). https://doi.org/10.1073/pnas.1320189111

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. Philosophical Transactions of the Royal Society B: Biological Sciences, 361(1476). https://doi.org/10.1098/rstb.2006.1934

Koechlin, E., Basso, G., Pietrini, P., Panzer, S., Grafman, J., Koechlin, E.,…Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition. Nature 1999 399:6732, 399(6732). https://doi.org/10.1038/20178

Koechlin, E., & Hyafil, A. (2007). Anterior Prefrontal Function and the Limits of Human Decision-Making. Science, 318(5850). https://doi.org/10.1126/science.1142995

Langdon, A. J., Sharpe, M. J., Schoenbaum, G., & Niv, Y. (2018). Model-based predictions for dopamine. Current Opinion in Neurobiology, 49. https://doi.org/10.1016/j.conb.2017.10.006

Lee, S. W., Shimojo, S. S., & P., O. D. J. (2014). Neural Computations Underlying Arbitration between Model-Based and Model-free Learning. Neuron, 81(3). https://doi.org/10.1016/j.neuron.2013.11.028

Li, W., Lopez, L., Osher, J., Howard, J. D., Parrish, T. B., Gottfried, J. A., & Wen Li, L. L., Jason Osher, James D. Howard, Todd B. Parrish, Jay A. Gottfried. (2010). Right Orbitofrontal Cortex Mediates Conscious Olfactory Perception. Psychological Science, 21(10). https://doi.org/10.1177/0956797610382121

Lichtenberg, N. T., Pennington, Z. T., Holley, S. M., Greenfield, V. Y., Cepeda, C., Levine, M. S., & Wassum, K. M. (2017). Basolateral Amygdala to Orbitofrontal Cortex Projections Enable Cue-Triggered Reward Expectations. Journal of Neuroscience, 37(35). https://doi.org/10.1523/JNEUROSCI.0486-17.2017

Lin, H., & Zhou, J. (2024). Hippocampal and orbitofrontal neurons contribute to complementary aspects of associative structure. Nature Communications 2024 15:1, 15(1). https://doi.org/10.1038/s41467-024-49652-9

Liu, Q., Zhao, Y., Attanti, S., Voss, J. L., Schoenbaum, G., Kahnt, T.,…Kahnt, T. (2024). Midbrain signaling of identity prediction errors depends on orbitofrontal cortex networks. Nature Communications 2024 15:1, 15(1). https://doi.org/10.1038/s41467-024-45880-1

Livneh, U., & Paz, R. (2012). Aversive-Bias and Stage-Selectivity in Neurons of the Primate Amygdala during Acquisition, Extinction, and Overnight Retention. Journal of Neuroscience, 32(25). https://doi.org/10.1523/JNEUROSCI.0323-12.2012

McDannald, M. A., Jones, J. L., Takahashi, Y. K., & Schoenbaum, G. (2014). Learning theory: A driving force in understanding orbitofrontal function. Neurobiology of Learning and Memory, 108. https://doi.org/10.1016/j.nlm.2013.06.003

McDannald, M. A., Lucantonio, F., Burke, K. A., Niv, Y., & Schoenbaum, G. (2011). Ventral Striatum and Orbitofrontal Cortex Are Both Required for Model-Based, But Not Model-Free, Reinforcement Learning. Journal of Neuroscience, 31(7). https://doi.org/10.1523/JNEUROSCI.5499-10.2011

Mcdonald, A. J. (1991). Organization of amygdaloid projections to the prefrontal cortex and associated striatum in the rat. Neuroscience, 44(1). https://doi.org/10.1016/0306-4522(91)90247-L

Miller, K. J., Botvinick, M. M., & Brody, C. D. (2022). Value representations in the rodent orbitofrontal cortex drive learning, not choice. eLife, 11. https://doi.org/10.7554/eLife.64575

Miller, K. J., Botvinick, M. M., Brody, C. D., Miller, K. J., Botvinick, M. M., & Brody, C. D. (2017). Dorsal hippocampus contributes to model-based planning. Nature Neuroscience 2017 20:9, 20(9). https://doi.org/10.1038/nn.4613

Noonan, M. P., Chau, B. K. H., Rushworth, M. F., & Fellows, L. K. (2017). Contrasting Effects of Medial and Lateral Orbitofrontal Cortex Lesions on Credit Assignment and Decision-Making in Humans. Journal of Neuroscience, 37(29). https://doi.org/10.1523/JNEUROSCI.0692-17.2017

O'Doherty, J., Critchley, H., Deichmann, R., & Dolan, R. J. (2003). Dissociating Valence of Outcome from Behavioral Control in Human Orbital and Ventral Prefrontal Cortices. Journal of Neuroscience, 23(21). https://doi.org/10.1523/JNEUROSCI.23-21-07931.2003

Ostlund, S. B., & Balleine, B. W. (2007). Orbitofrontal Cortex Mediates Outcome Encoding in Pavlovian But Not Instrumental Conditioning. Journal of Neuroscience, 27(18). https://doi.org/10.1523/JNEUROSCI.5443-06.2007

O'Reilly, Jill X., Schüffelgen, U., Cuell, S. F., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. Proceedings of the National Academy of Sciences, 110(38). https://doi.org/10.1073/pnas.1305373110

Park, S. A., Goïame, S., O'Connor, D. A., & Dreher, J.-C. (2017). Integration of individual and social information for decision-making in groups of different sizes. PLOS Biology, 15(6). https://doi.org/10.1371/journal.pbio.2001958

Pickens, C. L., Saddoris, M. P., Setlow, B., Gallagher, M., Holland, P. C., & Schoenbaum, G. (2003). Different Roles for Orbitofrontal Cortex and Basolateral Amygdala in a Reinforcer Devaluation Task. Journal of Neuroscience, 23(35). https://doi.org/10.1523/JNEUROSCI.23-35-11078.2003

Price, J. L. (2007). Definition of the orbital cortex in relation to specific connections with limbic and visceral structures and other cortical regions. Annals of the New York Academy of Sciences, 1121(1). https://doi.org/10.1196/annals.1401.008

Rolls, E. (2008). Functions of the orbitofrontal and pregenual cingulate cortex in taste, olfaction, appetite and emotion. Acta physiologica Hungarica, 95(2). https://doi.org/10.1556/Aphysiol.95.2008.2.1

Rudebeck, P. H., & Murray, E. A. (2014). The Orbitofrontal Oracle: Cortical Mechanisms for the Prediction and Evaluation of Specific Behavioral Outcomes. Neuron, 84(6). https://doi.org/10.1016/j.neuron.2014.10.049

Rushworth, M. F., Noonan, M. P., Boorman, E. D., Walton, M. E., & Behrens, T. E. (2011). Frontal Cortex and Reward-Guided Learning and Decision-Making. Neuron, 70(6). https://doi.org/10.1016/j.neuron.2011.05.014

Rushworth, M. F. S., Behrens, T.E.J. , Rudebeck, P. H., & Walton, M. E. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. Trends in Cognitive Sciences, 11(4). https://doi.org/10.1016/j.tics.2007.01.004

Saddoris, M. P., Gallagher, M., & Schoenbaum, G. (2005). Rapid Associative Encoding in Basolateral Amygdala Depends on Connections with Orbitofrontal Cortex. Neuron, 46(2). https://doi.org/10.1016/j.neuron.2005.02.018

Sagar, V., Shanahan, L. K., Zelano, C. M., Gottfried, J. A., Kahnt, T., Sagar, V.,…Kahnt, T. (2023). High-precision mapping reveals the structure of odor coding in the human brain. Nature Neuroscience 2023 26:9, 26(9). https://doi.org/10.1038/s41593-023-01414-4

Schoenbaum, G., Chiba, A. A., & Gallagher, M. (1999). Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning – PubMed. The Journal of neuroscience : the official journal of the Society for Neuroscience, 19(5). https://doi.org/10.1523/JNEUROSCI.19-05-01876.1999

Schoenbaum, G., Chiba, A. A., Gallagher, M., Schoenbaum, G., Chiba, A. A., & Gallagher, M. (1998). Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. Nature Neuroscience 1998 1:2, 1(2). https://doi.org/10.1038/407

Schoenbaum, G., & Eichenbaum, H. (1995). Information coding in the rodent prefrontal cortex. I. Single-neuron activity in orbitofrontal cortex compared with that in pyriform cortex. Journal of Neurophysiology, 74(2). https://doi.org/10.1152/jn.1995.74.2.733

Schoenbaum, G., & Roesch, M. (2005). Orbitofrontal Cortex, Associative Learning, and Expectancies. Neuron, 47(5). https://doi.org/10.1016/j.neuron.2005.07.018

Schoenbaum, G., Setlow, B., Saddoris, M. P., & Gallagher, M. (2003). Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala – PubMed. Neuron, 39(5). https://doi.org/10.1016/s0896-6273(03)00474-4

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). https://doi.org/10.1016/j.neuron.2016.08.019

Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. Science, 275(5306). https://doi.org/10.1126/science.275.5306.1593

Schwartenbeck, P., FitzGerald, T. H., & Dolan, R. (2016). Neural signals encoding shifts in beliefs. NeuroImage, 125. https://doi.org/10.1016/j.neuroimage.2015.10.067

Seo, H., & Lee, D. (2010). Orbitofrontal cortex assigns credit wisely. Neuron, 65(6). https://doi.org/10.1016/j.neuron.2010.03.016

Sescousse, G., Caldú, X., Segura, B., & Dreher, J. C. (2013). Processing of primary and secondary rewards: A quantitative meta-analysis and review of human functional neuroimaging studies. Neuroscience & Biobehavioral Reviews, 37(4). https://doi.org/10.1016/j.neubiorev.2013.02.002

Seymour, B., Daw, N. D., Roiser, J. P., Dayan, P., & Dolan, R. (2012). Serotonin Selectively Modulates Reward Value in Human Decision-Making. Journal of Neuroscience, 32(17). https://doi.org/10.1523/JNEUROSCI.0053-12.2012

Sharpe, M. J., Chang, C. Y., Liu, M. A., Batchelor, H. M., Mueller, L. E., Jones, J. L.,…Schoenbaum, G. (2017). Dopamine transients are sufficient and necessary for acquisition of model-based associations. Nature Neuroscience 2017 20:5, 20(5). https://doi.org/10.1038/nn.4538

Sias, A. C., Jafar, Y., Goodpaster, C. M., Ramírez-Armenta, K., Wrenn, T. M., Griffin, N. K.,…Wassum, K. M. (2024). Dopamine projections to the basolateral amygdala drive the encoding of identity-specific reward memories. Nature Neuroscience 2024 27:4, 27(4). https://doi.org/10.1038/s41593-024-01586-7

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and 107ocalization in cluster inference. NeuroImage, 44(1). https://doi.org/10.1016/j.neuroimage.2008.03.061

Stalnaker, T. A., Cooch, N. K., McDannald, M. A., Liu, T.-L., Wied, H., Schoenbaum, G.,…Schoenbaum, G. (2014). Orbitofrontal neurons infer the value and identity of predicted outcomes. Nature Communications 2014 5:1, 5(1). https://doi.org/10.1038/ncomms4926

Stalnaker, T. A., Cooch, N. K., Schoenbaum, G., Stalnaker, T. A., Cooch, N. K., & Schoenbaum, G. (2015). What the orbitofrontal cortex does not do. Nature Neuroscience 2015 18:5, 18(5). https://doi.org/10.1038/nn.3982

Stalnaker, T. A., Franz, T. M., Singh, T., & Schoenbaum, G. (2007). Basolateral Amygdala Lesions Abolish Orbitofrontal-Dependent Reversal Impairments. Neuron, 54(1). https://doi.org/10.1016/j.neuron.2007.02.014

Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., & Hirsch, J. (2006). Predictive Codes for Forthcoming Perception in the Frontal Cortex. Science, 314(5803). https://doi.org/10.1126/science.1132028

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 2011 14:12, 14(12). https://doi.org/10.1038/nn.2957

Tegelbeckers, J., Porter, D. B., Voss, J. L., Schoenbaum, G., & Kahnt, T. (2023). Lateral orbitofrontal cortex integrates predictive information across multiple cues to guide behavior. Current Biology, 33(20). https://doi.org/10.1016/j.cub.2023.09.033

Tsuchida, A., Doll, B. B., & Fellows, L. K. (2010-12-15). Beyond Reversal: A Critical Role for Human Orbitofrontal Cortex in Flexible Learning from Probabilistic Feedback. Journal of Neuroscience, 30(50). https://doi.org/10.1523/JNEUROSCI.1958-10.2010

Tsujimoto, S., Genovesio, A., & Wise, S. P. (2009). Monkey Orbitofrontal Cortex Encodes Response Choices Near Feedback Time. Journal of Neuroscience, 29(8). https://doi.org/10.1523/JNEUROSCI.5777-08.2009

Valentin, V. V., Dickinson, A., & O'Doherty, J. P. (2007). Determining the Neural Substrates of Goal-Directed Learning in the Human Brain. Journal of Neuroscience, 27(15). https://doi.org/10.1523/JNEUROSCI.0564-07.2007

van Holstein, M., Aarts, E., van der Schaaf, M. E., Geurts, D. E. M., Verkes, R. J., Franke, B.,…Cools, R. (2011). Human cognitive flexibility depends on dopamine D2 receptor signaling. Psychopharmacology 2011 218:3, 218(3). https://doi.org/10.1007/s00213-011-2340-2

Waelti, P., Dickinson, A., Schultz, W., Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. Nature 2001 412:6842, 412(6842). https://doi.org/10.1038/35083500

Wallis, J. D., & Miller, E. K. (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task – PubMed. The European journal of neuroscience, 18(7). https://doi.org/10.1046/j.1460-9568.2003.02922.x

Walton, M. E., Behrens, T. E., Buckley, M. J., Rudebeck, P. H., & Rushworth, M. F. (2010). Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. Neuron, 65(6). https://doi.org/10.1016/j.neuron.2010.02.027

Weiskopf, N., Hutton, C., Josephs, O., & Deichmann, R. (2006). Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: A whole-brain analysis at 3 T and 1.5 T. NeuroImage, 33(2). https://doi.org/10.1016/j.neuroimage.2006.07.029

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal Cortex as a Cognitive Map of Task Space. Neuron, 81(2). https://doi.org/10.1016/j.neuron.2013.11.005

Wise, R. A. (2004). Dopamine, learning and motivation. Nature Reviews Neuroscience 2004 5:6, 5(6). https://doi.org/10.1038/nrn1406

Witkowski, P. P., Park, S. A., & Boorman, E. D. (2022). Neural mechanisms of credit assignment for inferred relationships in a structured world. Neuron, 110(16), 2680-+. https://doi.org/10.1016/j.neuron.2022.05.021

Wunderlich, K., Smittenaar, P., & Dolan, Raymond J. (2012/08/08). Dopamine Enhances Model-Based over Model-Free Choice Behavior. Neuron, 75(3-4). https://doi.org/10.1016/j.neuron.2012.03.042

Yin, H. H., Knowlton, B. J., Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. Nature Reviews Neuroscience 2006 7:6, 7(6). https://doi.org/10.1038/nrn1919

Yoshida, W., & Ishii, S. (2006). Resolution of Uncertainty in Prefrontal Cortex. Neuron, 50(5). https://doi.org/10.1016/j.neuron.2006.05.006

Öngür, D., & Price, J. L. (2000). The Organization of Networks within the Orbital and Medial Prefrontal Cortex of Rats, Monkeys and Humans. Cerebral Cortex, 10(3). https://doi.org/10.1093/cercor/10.3.206

Supplementary Information

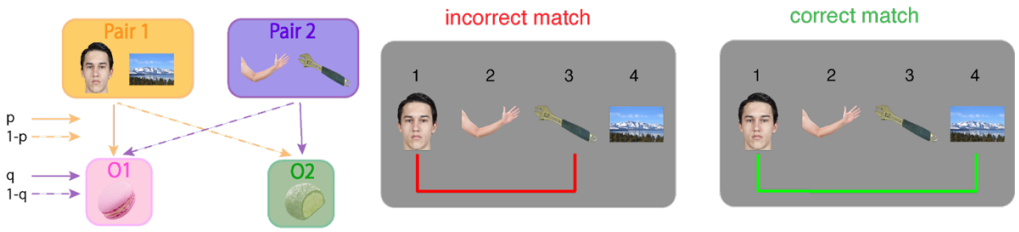


Fig.S1 Matching Task, Related to STAR Methods

Matching task. After completing the forced-choice trials in each run, participants performed a matching task to identify the 2 correct linked pairs they had learned from the forced-choice trials. Four stimuli were randomly displayed on the screen, each labeled with a number from 1 to 4. Participants were instructed to match stimuli of the same linked pair by consecutively pressing the numbers associated with each stimulus. For example, they would match “Face” and “Scene” by pressing 1 and 4 consecutively. Correct matches were indicated by a green line connecting the stimuli, while incorrect responses were indicated by a red line. If participants made an incorrect match, they were asked to repeat the matching task until correct matches were achieved.

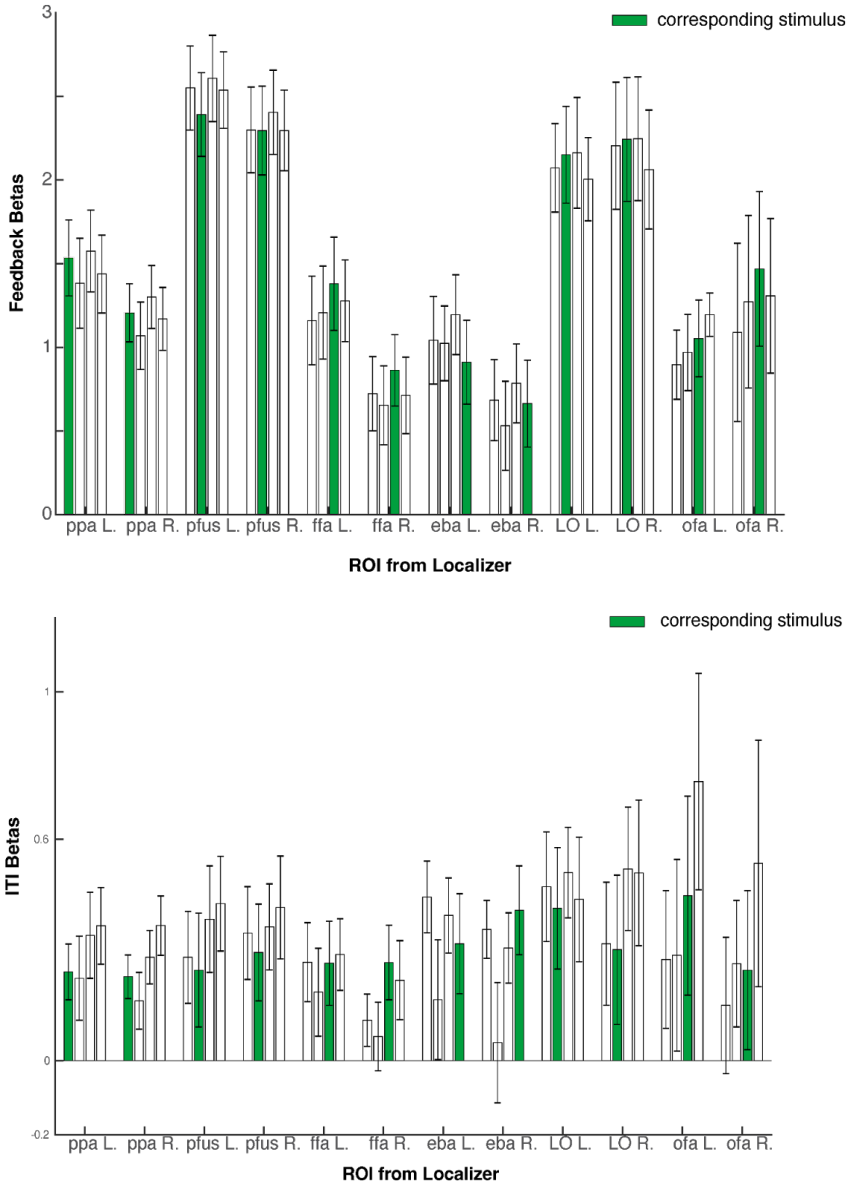


Fig.S2 Localizer Results Show No Evidence for Choice Reactivation in Sensory Cortices during either Feedback or Post-feedback periods.

Group-averaged results of the localizer: ROIs for each subject were obtained from the localizer (see Methods). Top graph: Feedback betas were modeled as 2-second boxcars from the onset

of each trial without added parametric modulators. The value of each bar represents the activity during feedback in each ROI, averaged across participants. Each trial was labeled according to the category of the chosen stimulus, with green bars indicating the averaged feedback betas for trials where participants selected the corresponding type of stimulus for an ROI. For example, the green bar in “ppa L.” represents the activity within the left parahippocampal place area during feedback for trials where “scene” was selected. Within each ROI from left to right, the bars correspond to “scene,” “tool,” “face,” and “body”. No correlation between ROIs and the type of chosen stimuli were detected ($p < .05$). Bottom graph: the same as top graph, except that ITI betas were modeled as boxcars including jitters.

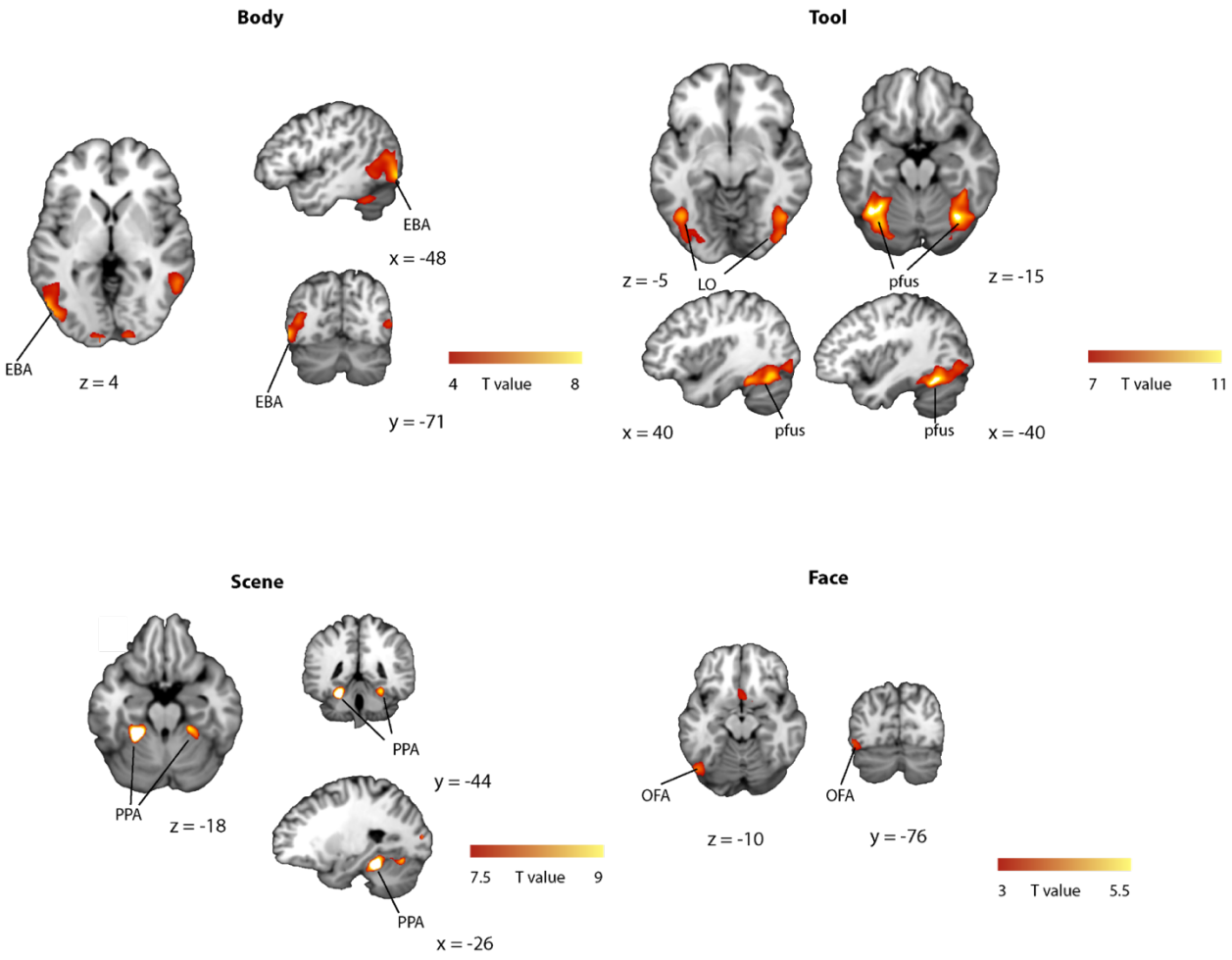


Fig.S3 Group-averaged on category-selective activation from the localizer.

Group-level activation of category-selective activation obtained from functional ROIs from localizer data. ROIs were created in native space for each participant. ROIs are constructed using contrasts with one or more categories to generate optimal category-selective regions. ROI for body parts (body > tools): extrastriate body area (EBA); ROI for tool (tool > body, scene, face): posterior fusiform gyrus; ROI for scene (scene > face): parahippocampal place area (PPA); ROI for face (face > scene): occipital face area (OFA).

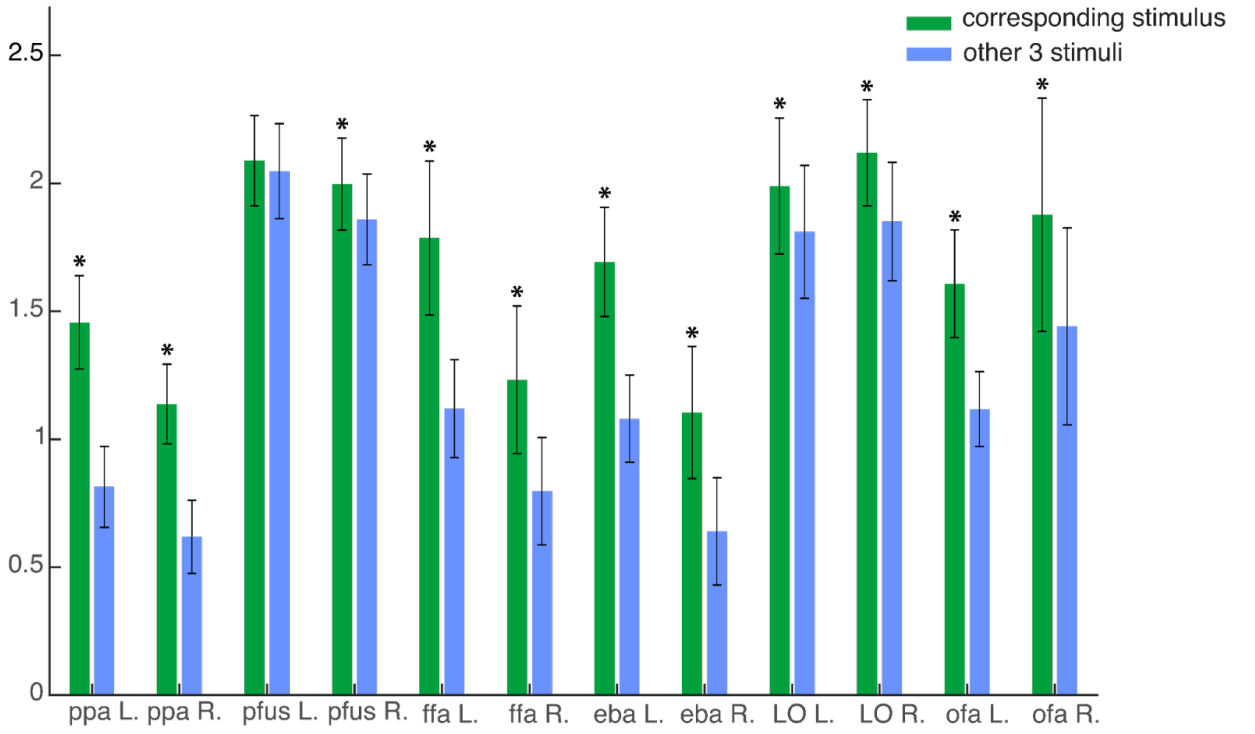


Fig.S4 Activation of functional ROIs from the localizer at cue on free-choice trials from the main experiment.

Functional ROIs identified from the localizer were used to evaluate whether their activation levels were higher for the corresponding stimulus during cue periods on free-choice trials in the main experiment on a per participant basis. For instance, the activation of the left PPA for each participant was compared between when "scene" was displayed on the screen and its averaged activation when "face," "body part," and "tool" were shown. 11 out of 12 ROIs demonstrated significantly higher activation ($p < 0.05$) for the corresponding stimulus in the main task, confirming the validity of the ROIs identified from the localizer. Scene: left ppa $t(24) = 6.66$, $p = 0.000001$; right ppa $t(24) = 5.13$, $p = 0.00003$. Tool: left pfus $t(25) = 0.618$, $p = 0.542$; right pfus $t(26) = 2.12$, $p = 0.0437$; left LO $t(25) = 2.543$, $p = .0175$; right LO $t(26) = 3.537$. Face: left ffa

t(23) = 4.676, p = 0.000104; right ffa t(19) = 3.376, p = 0.00317; left ofa t(16) = 3.372, p = 0.00389; right ofa t(11) = 2.491, p = 0.0300. Body part: left eba t(25) = 5.388 p = 0.000014; right eba t(20) = 3.439, p = 0.00260.

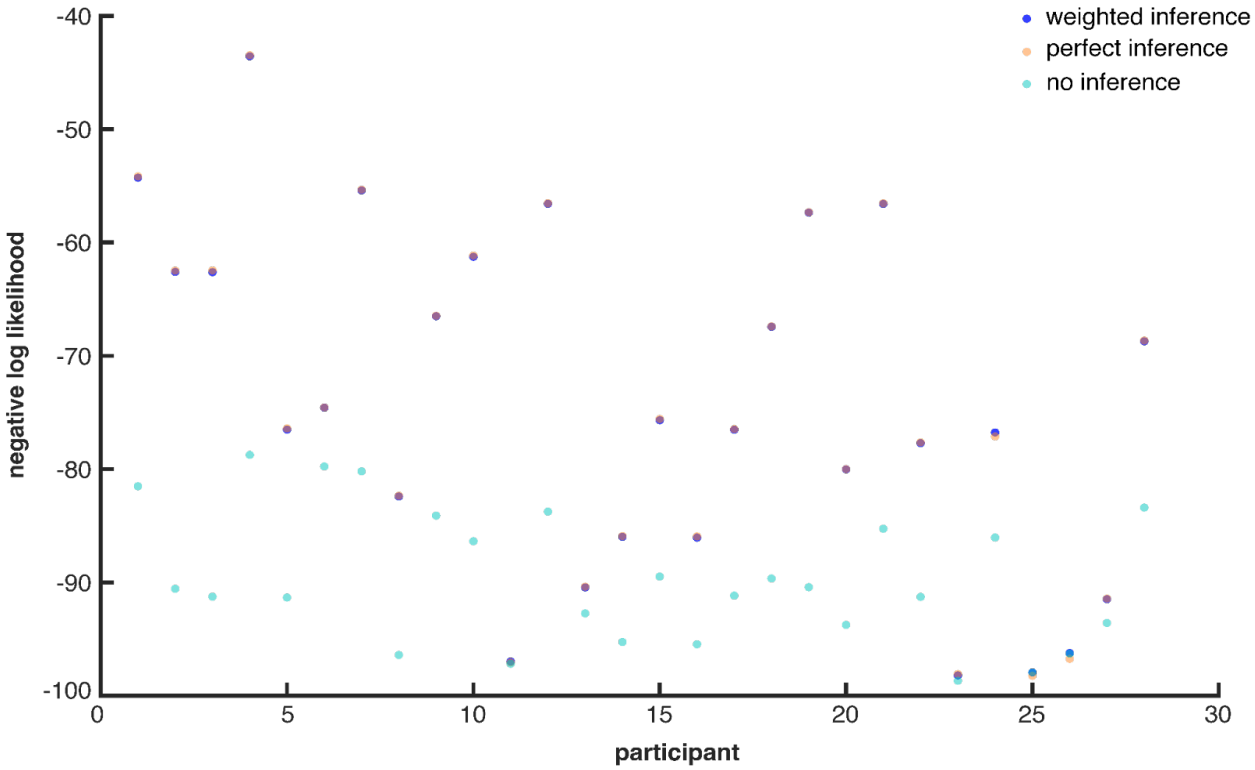


Fig.S5 Additional model fitting results. Related to STAR Methods

Negative log-likelihoods of the 3 behavioral models by participant. This is evidence that the weighted inference model explains the behavioral data almost equally well as the perfect inference model, and that participants are well trained for the task.

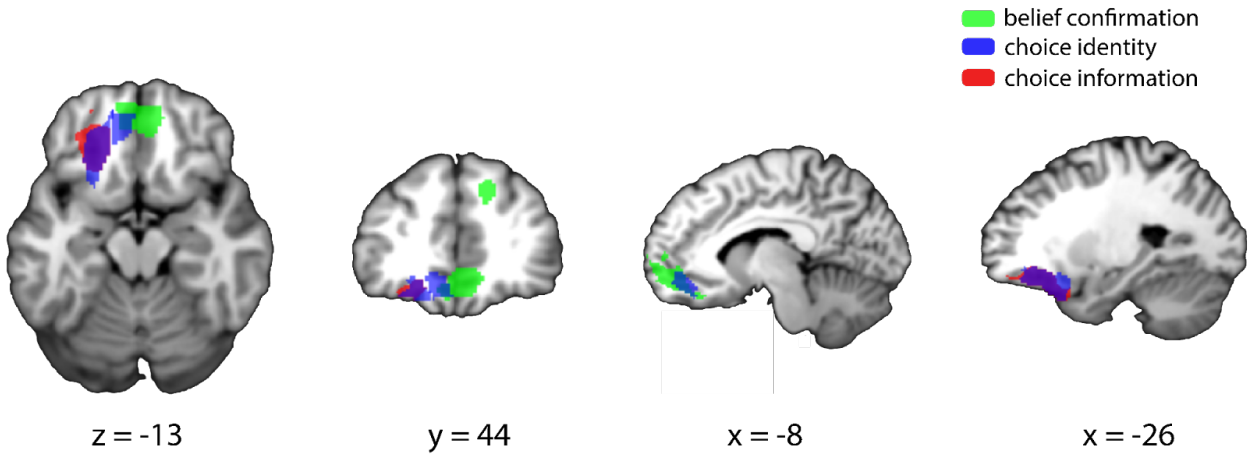


Fig.S6 Combined results on the locations of OFC from 3 analyses.

Related to Figure 2.2-2.4. From left to right: horizontal, coronal, and sagittal views of the OFC clusters surviving small-volume corrections (pTFCE < 0.05) from a priori ROIs in analyses of choice identity and choice information (Figure 2.3). The OFC clusters surviving whole-brain TFCE (pTFCE < 0.05) in belief confirmation from univariate results (Figure 2.2) are also shown. The left lateral OFC contains decodable information on both choice information and choice identity, while the cOFC is involved only in choice identity. The mPFC is involved in belief confirmation, with some overlap with the cOFC. Note that belief confirmation also involves the right OFC, as depicted in Figure 2.2.

Table S1. Distributions of individual-specific parameter estimates for each behavioral model.

Related to STAR methods

a

				Peak (MNI)			
Region	*Voxels*	*p-value*	*p-value TFCE*	*t-stat (max)*	*x*	*y*	*z*
L. pre-SMA ★	126	$4.23x10^{-5}$	$9.00x10^{-3}$	4.62	-10	20	44
R. pre-SMA ★	145	$5.24x10^{-5}$	$1.14x10^{-2}$	4.54	10	14	50
L. anterior insula ★	67	0.000337	0.007	3.84	-32	18	-2
R. anterior insula ★	60	0.00179	0.0136	3.42	32	22	-4

b

				Peak (MNI)			
Region	*Voxels*	*p-value*	*p-value TFCE*	*t-stat (max)*	*x*	*y*	*z*
medial PFC	1218	$8.99x10^{-9}$	$8.99x10^{-9}$	-7.88	10	52	-6
R. OFC	223	$6.67x10^{-5}$	$6.67x10^{-5}$	-4.45	34	34	-18
L. hippocampus	399	$7.23x10^{-5}$	$7.23x10^{-5}$	-4.42	-30	-16	-18
R. hippocampus	369	$3.85x10^{-4}$	$3.85x10^{-4}$	-3.79	26	-8	-20
posterior cingulate	579	$9.39x10^{-6}$	$9.39x10^{-6}$	-5.18	8	-54	14
R. middle temporal lobe	1025	3.78x10-06	$3.78x10^{-6}$	-5.52	54	-60	2
L. middle temporal gyrus	175	0.000259	0.000259	-3.94	-56	-62	0
L. middle occipital lobe	257	0.000416	0.000416	-3.76	-46	-78	22
R. middle temporal lobe	191	0.000273	0.000273	-3.92	50	-74	22

Table S2. Univariate activation in response to DKL **at feedback, related to Figure 2.2:**

a) *Univariate activation at the time of feedback given* DKL *update (positive* DKL*), ★ indicates TFCE correction with an anatomically independently defined ROI* ***b)*** *Univariate activation at the Time of Feedback Given* DKL *confirmation (negative* DKL*)*

				Peak (MNI)			
Region	*Voxels*	*p-value*	*p-value TFCE*	*t-stat (max)*	*x*	*y*	*z*
Ventral Tegmental Area	70	0.000394	0.048	3.78	8	-18	-18

Table S3. Between-subject correlation of DKL at feedback and significant logistic regression coefficients for the inferred condition (1-back and 2-back)

a				*Peak (MNI)*			
Region	*Voxels*	*p-value*	*p-value TFCE*	*t-stat (max)*	*x*	*y*	*z*
L. lOFC ★	95	0.000306	0.0136	3.89	-22	36	16
L. putamen ★				5.37	-28	0	-6
vta substantia nigra pars compacta ★				4.27	12	-18	-12
b				*Peak (MNI)*			
L. OFC ★	79	0.000115	0.013	4.28	-18	38	-16
L. cOFC ★	31	0.000412	0.007	4.63	12	42	-18
R. Amygdala ★	11	0.0021	0.017	3.13	18	-2	-18
Midbrain ★	45	0.000137	0.0148	4.18	-4	-24	-10

Table S4. Decoding of general information for choice and identity-specific code for choice, related to Figure 2.3. ***A)*** *Decoding of general information of choice, ★ indicates TFCE correction with an anatomically independently defined ROI* ***b)*** *Decoding of identity-specific code for choice, ★ indicates TFCE correction with an anatomically independently defined ROI*

a				*Peak (MNI)*			
Region	*Voxels*	*p-value*	*p-value TFCE*	*t-stat (max)*	*x*	*y*	*z*
R. frontopolar ★	41	0.000145	0.0232	4.16	24	56	-2
L. mPFC ★	11	0.0027	0.031	3.02	0	52	-12

DLPFC	80	0.000216		4.01	-6	48	32
b				*Peak (MNI)*			
L. LOFC★	**need an ROI**	0.000462	0.0934	3.72	-44	40	-20

Table S5. Decoding of general information for inferred stimulus and identity-specific code for inferred stimulus, related to Figure 2.4. ***A)*** *Decoding of general information of inferred stimulus, ★ indicates TFCE correction with an anatomically independently defined ROI* ***b)*** *Decoding of identity-specific code for inferred stimulus, ★ indicates TFCE correction with an anatomically independently defined ROI*

				Peak (MNI)			
category	*region*	*voxel*	*p-value*	*t-stat (max)*	*x*	*y*	*z*
body	L. extrastriate body area (EBA)	443	$2.34x10^{-8}$	7.49	-48	-82	-2
	R. extrastriate body area (EBA)	206	0.0015	3.27	52	-72	-4
tool	R. posterior fusiform	415	$3.38x10^{-11}$	10.35	48	-64	-10
	L. posterior fusiform	448	$2.05x10^{-11}$	10.59	-34	-68	-14
	R. lateral occipital cortex (LO)	759	$2.92x10^{-9}$	8.35	32	-84	4
	L. lateral occipital cortex (LO)	812	$3.88x10^{-9}$	8.23	-36	-84	8
scene	L. parahippocampal place area (PPA)	332	$2.01x10^{-12}$	11.75	-26	-44	-18
	R. parahippocampal place area (PPA)	191	$2.75x10^{-10}$	9.38	22	-38	-16
face	L. occipital face area	290	$8.21x10^{-6}$	5.23	-50	-70	-4

Table S6. Group-averaged on category-selective activation from the localizer. Group-level activation of category-selective activation obtained from functional ROIs from localizer data. Related to Figure S3.

Chapter 3. Hierarchical Causal Levels Shape and Support Learning and Abstraction

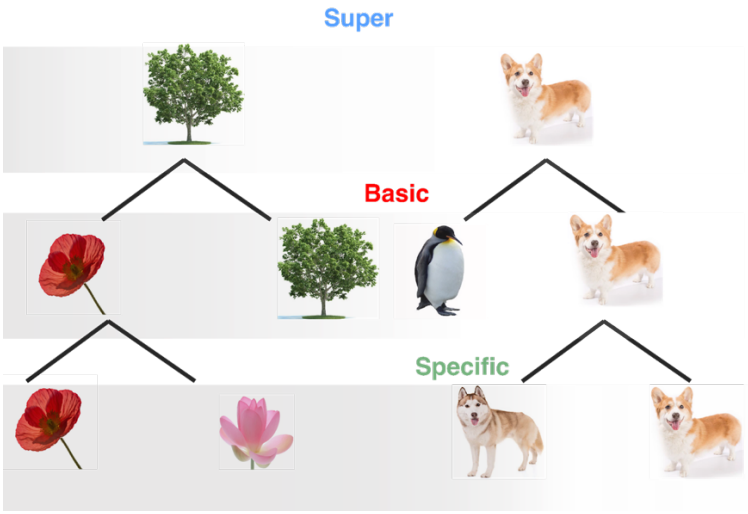
Introduction

In the previous study, we investigated the neural mechanisms underlying one type of causal learning, where different actions are guided by a shared underlying cause, allowing learners to infer the outcomes of possible actions after learning the stimulus-outcome contingency of observed actions. We identified neural reinstatements for both observed and inferred causal choices during feedback, demonstrating that these reinstatements support flexible learning in a structured environment. However, real-world environments are often characterized by multiple hierarchies of latent cause, unlike our previous study, which only included two simple hierarchical levels of latent cause: two systems and two options under each system. Here is an example of how learners must assign credit given multiple hierarchies. For example, after a successful first date from a dating app, a user might try to understand why it went well by assigning credit at different levels. At the superordinate level, the success could be attributed to men in general, suggesting that the user has positive interactions with the male population overall. At the basic level, the success might be credited to a specific category of men, such as those with similar educational backgrounds or shared interests, indicating that these factors played a role in the compatibility. Finally, at the specific level, the user might assign credit to the individual man they went out with, recognizing that the unique qualities and personality of that particular person were key to the successful date. This hierarchical approach helps the user understand what aspects contributed most to the positive outcome.The number of hierarchical levels varies according to the context. This chapter elaborates on the behavioral findings of an ongoing study that will include a neuroimaging portion for future research. Specifically, we asked whether humans learn and perform abstraction differently according to which

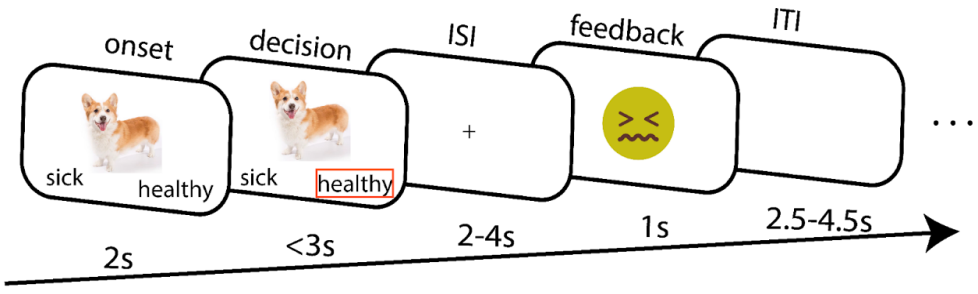
hierarchical level is relevant for credit assignment. Ultimately, this project aims to examine if there exists a neural reinstatement of the appropriate hierarchical level during feedback given belief updates of both the governing hierarchical level and stimulus-outcome contingency. Additionally, future research will explore how the neural representations of hierarchical levels could be utilized during abstraction or the prediction of novel stimuli in credit assignment.

Methods and Results

A



B



C

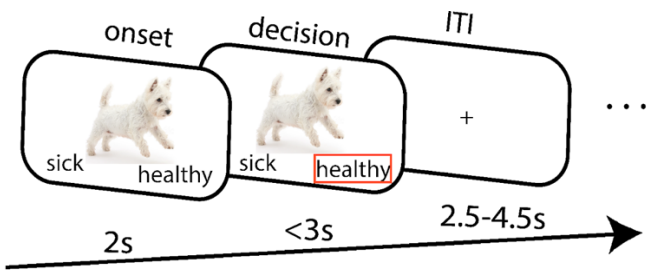


Figure 3.1. Learning task design

A. Illustration of the 3 hierarchical casual levels: super, basic, and specific. Each category within a level shares the same contingency for outcomes. For example, at the super level, all items belonging to the category of "animal" are either all sick or all healthy. Similarly, at the basic level, all items belonging to the category of "bird" are either all sick or all healthy. At the specific level, a category is defined as an independent item. Thus, if the corgi is sick, it does not infer the husky's outcome since husky and corgi are separate categories at the specific level.

B. Depiction of an example training trial. Participants were presented with a stimulus, and then were asked to choose between "sick" and "healthy" for its outcome. A red box appeared on the decision. Then, participants were presented with the correct outcome of the stimulus. In this example, the corgi's correct outcome was sick.

C. Depiction of an example generalization trial. Participants were presented with a novel stimulus on 7 of the 23 trials within a block and had to decide whether the novel stimulus was sick or healthy by utilizing their knowledge about the estimated hierarchical causal level of that period. In this example, when deciding if the terrier was sick or healthy, participants would use

their estimate of the most probable causal level from previous observations. The terrier would be predicted to be sick if the most probable causal level was basic and other dogs were almost always sick.

Participants (N=10) completed a "hierarchical 3 levels learning task" in which they tracked the probability of 8 stimuli causing sickness (Figure 3.1). The task consisted of 12 blocks, and each block belonged to 1 of the 3 hierarchical levels (Figure 3.1 A): super, basic, or specific. All stimuli belonging to the same category under a level have the same outcome ("sick" or "healthy"). For example, if the block is under the super level, then all animals have the same outcome, and all plants had the same outcome. Therefore, learning the outcome of one stimulus of a category allowed participants to infer the outcome of other stimuli of the same category. The basic level consisted of 4 levels: birds, dogs, flowers and trees. For example, all birds had the same outcome, and all flowers had the same outcome, etc. Finally, under the specific level, each stimulus predicted its outcome.

In every trial, participants were presented with a stimulus for 2 seconds and then asked to choose if the stimulus was sick or healthy. At the beginning of the task, participants did not know which level they were at for a current block, and had not made any observations, thus they needed to guess for the outcomes. After making a decision, participants received feedback labeled by a sick or healthy emoji, indicating if the stimulus presented on the trial was sick or healthy. Over time, participants could infer which hierarchical level they were at by associating stimuli with observed outcomes. They could then use this information on the hierarchical level to guide their decisions, especially on generalization trials (Figure 3.1C). Each block consisted of 16 training trials and 7 generalization trials, where a novel stimulus not included in the 8

training stimuli was presented to participants. Participants needed to decide whether the novel stimulus was sick or healthy based on their estimate of the current hierarchical level. If they believed they were on the super level and from their observations, animals almost always were sick, they would most likely predict a novel animal to be sick. Similarly, if they believed that the current hierarchical level was basic, and had learned that flowers almost always were sick, they would most likely predict a novel flower to be healthy.

Participants were told not to predict a stimulus to be "sick" if a novel stimulus did not belong to any category within a level. For example, let's say that they believed that the current hierarchical level was basic and had learned that birds almost always were sick, and dogs almost always were healthy. If a novel stimulus on a generalization trial showed up to be a whale, which belonged to neither bird nor dog, then participants were most likely to predict it to be "healthy". Another important note is that, under the specific level, since each stimulus predicts its outcome, any novel stimuli in generalization trials under this level should be predicted as "healthy". To elaborate, if participants learned that the current hierarchical level is specific, and that among the training stimuli: sparrow was sick, penguin was healthy, corgi was sick, and dalmatian was healthy. Then if a novel bird or dog shows up on a generalization trial, they should predict it to be "healthy", since the novel stimulus should not depend on the outcome of any training stimuli under the specific level.

The probability of a particular category under a level leading to its outcome was 93.75%, which was nearly deterministic. Participants were not explicitly told when a block finished and moved on to the next block with a different hierarchical level, thus they needed to infer the level by consistently tracking and updating the stimulus-to-outcome contingencies. Participants

were told that a random block would be selected upon completing the task, and their accuracy on the generalization trials within that block would determine how much bonus money they received.

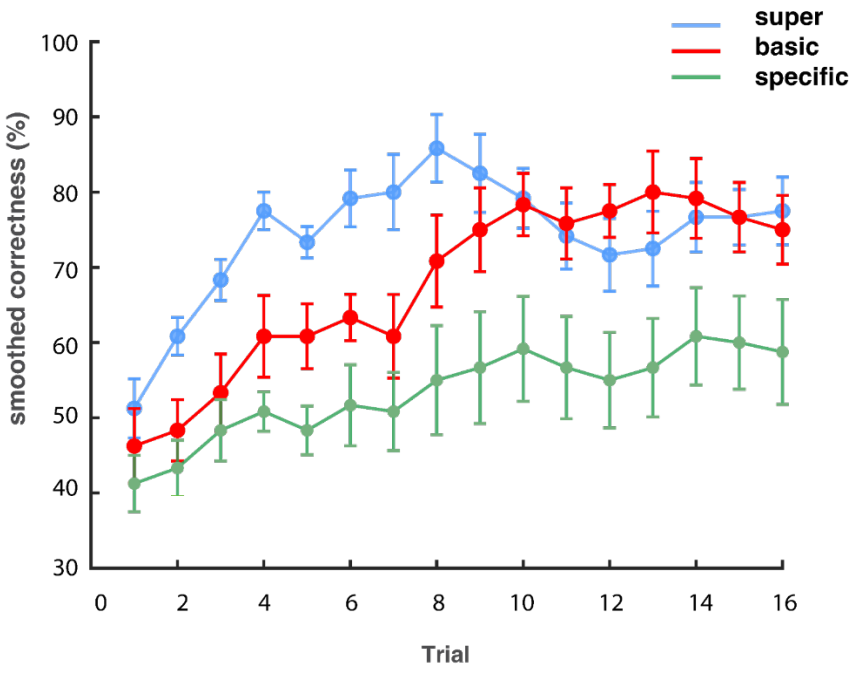


Figure 3.2. Learning over time during training trials Results of the smoothed correctness over time within a block for the training trials. Answers to the training trials were labeled as correct or incorrect and smoothed over a window of 3 trials to reduce noise. A linear mixed effect model was used to test for the fixed effect of hierarchical causal level and trial number on correctness while controlling for the random effects among participants. For visualization, this figure illustrates the mean over the participants for 16 training trials with standard error on every trial.

To investigate whether learning behaviors depend on hierarchical level, we plotted the learning curve over time for training trials within each block. Participants completed 16 training

and 7 prediction trials in each of the 12 blocks. Each training trial was labeled as "correct" or "incorrect" based on whether the participant correctly identified the outcome (“sick” or “healthy”) in response to the stimulus. To reduce noise, including random fluctuations in performance, we smoothed the correctness data for each block. We achieved this by averaging the correctness of each trial with the correctness of the preceding and following trials. For instance, if a participant correctly identified the outcome on trial n-1, incorrectly on trial n, and correctly on trial n+1, the correctness for trial n was calculated as 0.67.

After smoothing the correctness data, we applied a linear mixed-effects model to determine whether correctness could be explained by hierarchical level and/or trial. The model included hierarchical level and trial as fixed effects and participants as random effects to account for inter-individual variability.

$$correctness_{ijkl} = \beta 0 + \beta 1 Level_{j} + \beta 2 Trial_{K} + u_{i} + \epsilon_{ijkl}$$

Where correctness is the learning performance for subject *i*, level *j*, trial *k*, and measurement *l*. Since there are 4 blocks of the same level throughout the task, we had 4 measurements in total per level. As expected, the fixed effect of level negatively predicted correctness (BIC= 742.01, beta1 = -0.1043, t [1917] = -12.896, p = 1.49×10^{-36}). Ordinal rank for trial number within a block (1 to 16) positively predicted correctness (BIC= 742.01, beta2 = 0.0140, t[1917] = 9.799, p = 3.742×10^{-22}). Since level was an ordinal variable with super, basic, and specific being coded as 1, 2, and 3, respectively, our result suggested that correctness for training trials was the highest for the super level and decreased for basic and specific levels. As

the trial number increased within a block, meaning more observations were made, the correctness score also increased.

This linear mixed-effects model accounted for the random effects of participants, considering individual differences in learning, and summarized the relationship between correctness, level, and time. As such, learning for training trials is significantly dependent on the hierarchical level, with the best learning occurring on the super level, followed by the basic level, and the least learning overall on the specific level.

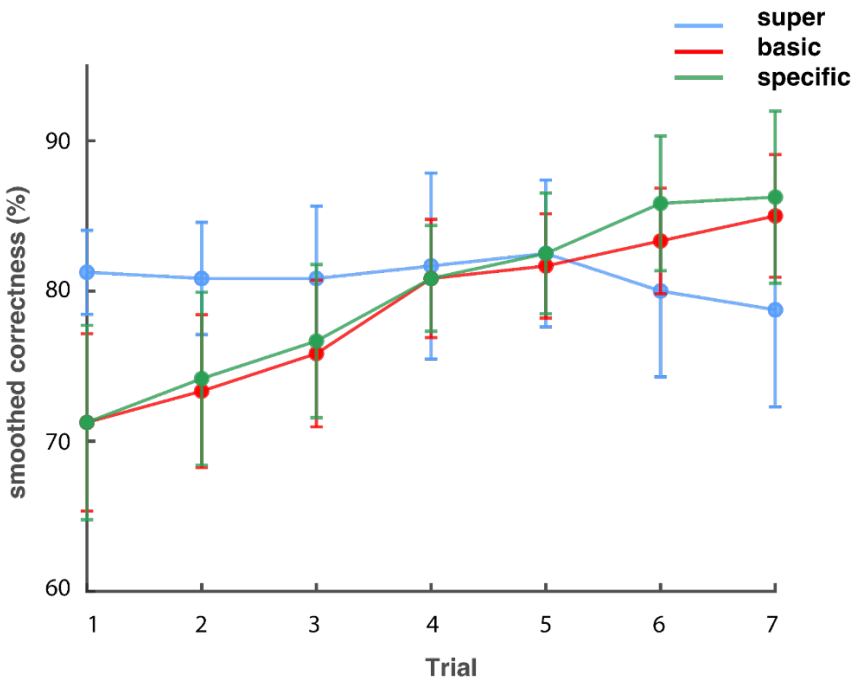


Figure 3.3. Accuracy of abstraction over time during generalization trials Results of the smoothed correctness over time within a block for the 7 prediction trials. Answers to the prediction trials were labeled as correct or incorrect and smoothed over a window of 3 trials to reduce noise. A linear mixed effect model was used to test for the fixed effect of hierarchical causal level and trial number on correctness while controlling for the random effects among participants. Then, another linear mixed effect model was used to test for only the first 3

instances out of the 7 prediction trials to test for the fixed effect of level on smoothed correctness in the early period of learning. For visualization, this figure illustrates the mean over the participants for 7 training trials with standard error on every trial.

We analyzed the learning for generalization trials separately from training trials due to the inherent differences in design between the two types of trials. Firstly, generalization trials did not provide feedback to participants after they decided on the outcome ("sick" or "healthy") of a stimulus. In contrast, training trials offered the correct answer after making a decision. Generalization trials thus reflected the level of abstraction independent of external feedback, effectively indicating participants' beliefs about the hierarchical level they were in at a particular moment. Secondly, each block consisted of 16 training trials but only 7 generalization trials. The reduced number of trials necessitates that generalization trials be analyzed separately. Thirdly, the stimuli in generalization trials differed from those in training trials. In training trials, participants were presented with familiar stimuli, each presented 24 times throughout the task, while stimuli in generalization trials were presented much less frequently.

Using a similar approach to the analysis in Figure 3.2, we smoothed the correctness for the generalization trials within a block. Then we used a linear mixed-effects model to test whether correctness in generalization trials could be explained by hierarchical level and/or trial number. The model equation was the same as the one used for training trials. The fixed effect of a level was not significant (beta1 =-0.00595, t[837] = -0.548, p = 0.584, BIC = 158.71),

whereas the ordinal rank of trial number (1 to 7) within a block positively predicted correctness (beta2 =0.09159, t[837] = 3.594, p = 0.000345, BIC = 158.71). The significant fixed effect of trial number on correctness suggested that participants' accuracy on generalization trials improved over time within a block as they made more observations of the stimulus-outcome relationships in training trials and formed a better estimate of the current hierarchical level of the block. The insignificant fixed effect of the hierarchical level indicated that participants' accuracy on generalization trials did not depend on which of the three levels they were at within a block. Although this might seem contradictory to the result from training trials, a closer look at the mean smoothed accuracy for generalization trials (Figure 3.3) reveals that the accuracy of generalization trials seemed to differ early in the block compared to later. Specifically, in the early stage of a block, learning at the super level exceeded that in basic and the specific levels, but learning for basic and the specific levels improved over time, and learning across the three levels appeared to be similar by the end of the block. It is, therefore, reasonable to examine the early stage of a block separately to determine if learning depends on the level during this period.

We applied the linear mixed-effects model to the first three ordinal ranks of the generalization trials to test if the level could predict correctness in generalization during the early stage of a block. The fixed effect of level on correctness was significant (beta1 =-0.0347, t[837] =-1.976, p = 0.0489, BIC = 131.06). This result suggests that accuracy in generalization trials was highest at the super level early in the block, followed by the basic and the specific levels. Together, our results indicate that the hierarchical level predicted how well participants generalized their knowledge to novel stimuli, but only in the early stage. Additionally, learning

in the basic and the specific levels improved over time, eventually matching the level of learning in the super level.

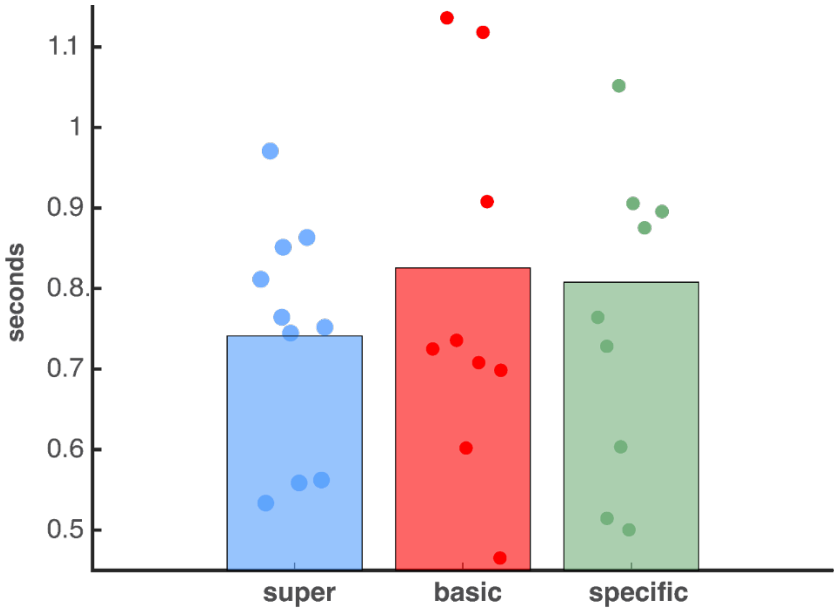


Figure 3.4. Differences in reaction times for decisions in generalization over 3 hierarchical levels Repeated measure ANOVA with-in subjects shows a marginally significant effect of the hierarchical causal level on reaction time ($F[2,478] = 2.550$, $p = 0.0792$). For visualization, only the mean over all generalization trials for participants are shown here. The bars illustrate the mean reaction times of participants, and the dotted circles indicate the reaction time of each participant at different levels, averaged over all generalization stimuli in all blocks.

Since correctness in generalization trials depended on hierarchical level in the early stage of learning, we asked if generalization in the super level was less difficult than basic and

the specific levels. Therefore, we examined the reaction times over the 3 hierarchical levels as a behavioral correlate for difficulty. We hypothesized that reaction time was the lowest for the super level, followed by basic and the specific levels.

One advantage of the task design was the even distribution of the generalization trials across the 3 levels, enabling unbiased comparisons of behavioral or neural correlates according to hierarchical levels. Specifically, out of the 36 generalization stimuli, 24 were presented 3 times, and 12 were presented only 1 time across trials. For those 24 stimuli, each was presented exactly one time on 1 of the 3 levels. As a result, each of the 24 stimuli was presented one time per level. In other words, each of the 24 stimuli stayed constant, while level changed for the 3 times it was presented. This design controlled for any confounds that might stem from differences in stimuli, and limited the predictor to a hierarchical level only, enabling us to examine the effect of level on these 24 stimuli.

We conducted a repeated measure ANOVA with-in subjects to test the effect of hierarchical level on reaction time. The within-subjects factor was the hierarchical level with 3 levels (super, basic, specific). Results suggested marginal significance on the intercept of levels ($F[2,478] = 2.550$, $p = 0.0792$), indicating a marginal effect of level on reaction time, given the same stimulus identity on generalization trials. As depicted in Figure 3.4, the mean reaction time on generalization trials was the least in the super level. Participants responded faster in generalization trials on blocks belonging to the specific level than the basic level. In sum, the hierarchical level has a marginally significant effect on reaction time in generalization trials.

Discussion

To investigate how learning and abstraction could differ across 3 hierarchical levels (super, basic, specific), we conducted several analyses to identify behavioral correlates associated with the effect of level on learning. In the first analysis, we found different learning patterns in training trials across time. Both time within a block and the hierarchical level of the block significantly affected the correctness of decisions in training trials. Learning improved over time within a block and improved as the level increased in the hierarchy. Then, we examined the generalization trials separately from the training trials and found that the overall accuracy of abstraction did not differ significantly across 3 levels. However, in the early stage of learning, abstraction at the super level was much more effective than on basic and specific levels. As more observations were made from obtaining feedback, the accuracy for abstraction improved for basic and specific levels and eventually reached the same accuracy as the super level. Finally, we identified that the speed of abstraction was faster at the super level than at the basic and the specific level.

Learning and abstraction exhibited behavioral similarities and differences across the 3 hierarchical levels. It might be intuitive to think that learning and abstraction are relatively easy at the super level and increase in difficulty as the hierarchy moves down to the basic and the specific level. Our results demonstrated that this is not entirely accurate. Indeed, abstraction and learning seem at their finest on the super level when there are the least number of categories within a level. However, as an agent observed more outcomes from their decisions and updated the stimulus-outcome contingency with a better estimate of the hierarchy level, abstraction in lower hierarchies improved and matched that in high hierarchies. In future neuroimaging analyses, it is wise not to assume measurements related to abstraction as a

constant variable across the entire duration. Learning, on the other hand, exhibited distinct patterns across time for the 3 hierarchical levels. the super level exhibited the most robust learning measured by correctness, followed by basic and specific levels, where the tracking of stimulus-outcome contingency required more working memory loads.

We have preliminarily found how hierarchical level affected learning and abstraction and showed that participants demonstrated learning and understanding of the task. The following steps of the study will include 1) building a behavioral computational model to capture the learning and abstraction behaviors, and to generate estimates on the belief of what hierarchical level one is currently at. 2) conduct a logistic regression analysis to capture how previous choices and outcomes in training trials would affect a current decision and test if participants had integrated previous information to guide their choice 3) conduct the experiment with fMRI to find neural correlates during learning and abstraction for different levels of hierarchy. It will be exciting to see, for a same stimulus in generalization, how different hierarchical levels could affect its neural representation other than its consistent low-level visual representations, as an agent could utilize potential reinstated representations of different hierarchical levels to perform optimal abstraction. For our hypotheses in neural signatures built on previous studies about how different hierarchical levels might support contingent learning, see the general discussion in Chapter 4.

Chapter 4. General Discussion

Research from multiple disciplines has provided insights into how the brain forms and utilizes predictive models to infer causality in complex environments. From the development of the temporal difference learning model in reinforcement learning (Sutton, 1988) to the identification of neural correlates of prediction errors as a key mechanism for information updating in learning (Schultz et al., 1997), significant progress has been made. This dissertation contributes to the broader understanding of causal learning through two studies on credit assignment, highlighting the flexibility of causal learning supported by both behavioral and neural mechanisms. Across the experiments presented, we identified a key neural signature at the time of feedback, where learning signals are updated. Specifically, the brain reinstates both observed and inferred actions in distinct regions of the frontal cortex to facilitate the formation of task-relevant contingencies. Additionally, we uncovered behavioral differences in learning and abstraction across varying hierarchical levels of causality. Collectively, these findings elucidate the neural underpinnings that support adaptive credit assignment and generalization in structured environments.

Causal choice reinstatement in contingent learning

Previous research identified the critical role of the OFC in model-based contingent learning (Boorman et al., 2016; Izquierdo et al., 2004; Jocham et al., 2016; Noonan et al., 2017; Walton et al., 2010; Wilson et al., 2014). In particular, the lateral portion of OFC is important and necessary for the formation of contingencies between precise stimuli to outcomes (Seo & Lee, 2010; Walton et al., 2010). However, it remained unclear what types of information was processed in the lOFC for it to play a causal role in enabling contingent learning. In Chapter 2,

we extended the previous findings by providing evidence in accord with our a priori hypothesis that lOFC contains a representation of choice or action, which is reinstated during feedback upon the receipt of outcome. In addition, we discovered that the brain not only reinstates causal choice or actions, but also contains a code in the frontopolar cortex during feedback for unobserved choice or actions whose contingency to outcome could be inferred. Moreover, we found that the degree of causal learning through inference varies among individuals and is reflected in the VTA signals for belief updates during feedback.

The results in Chapter 2 provide a mechanistic explanation to why the OFC is important in contingent learning and the use of a cognitive map in decision making. Our task design and decoding method implied that the reactivation of causal choice or action encompasses multiple layers of information regarding the chosen stimulus. The lateral portion of the OFC encodes general information about a relevant past choice, including, but not limited to, state, value, identity, and emotional salience. These various aspects of information about the chosen stimulus collectively form a comprehensive representation of choice. Our findings in Chapter 2 provide robust evidence that general information about the choice is reinstated during feedback in the lOFC, supporting the theory that the OFC tracks state-wise information within a cognitive map (Stalnaker et al., 2015; Takahashi et al., 2011; Wilson et al., 2014; Witkowski et al., 2022) The position of an agent in a cognitive map is heavily dependent on the overall information about the choice. Disruptions to the OFC would affect the reinstatement of choice during feedback and perturb the continual updating necessary for forming a cognitive map. In Chapter 2, we also located the ROIs where identity-specific information of the observed choice is reinstated. The BLA nucleus of the amygdala and medial portion of OFC both showed

significant evidence in representing the identity of choice during feedback. Our result added to a growing body of research demonstrating the significance of the amygdala in representing conditioned stimuli of different modalities (Gore et al., 2015; Herry et al., 2008; Livneh & Paz, 2012; Schoenbaum et al., 1998), and extended those results by showing that the activities in the amygdala during feedback besides the decision period are also contingent learning.

One possible limitation of Chapter 2 was the robustness of evidence regarding the identity of inferred choice that is not directly observed during the decision. Although we demonstrated that the ventral lateral portion of the OFC contains decodable information about the identity of inferred choice, the cluster was relatively small and was more lateral than most lateral portions of the OFC that were reported in previous studies. It is wise to interpret this result with caution.

A promising direction for future research, building on our findings in Chapter 2, involves developing methodologies to investigate the informational connectivity (Coutanche & Thompson-Schill, 2013) between the OFC and sensory cortices where visual information is processed. Given that the OFC receives substantial input from both high-level sensory areas and low-level visual regions and plays a crucial role in object recognition (Bar et al., 2006; Carmichael & Price, 1995; Rushworth et al., 2007), it is plausible that the OFC gradually forms a code for choice reinstatement as it acquires more information about the chosen stimulus from posterior visual regions. Once the OFC has obtained sufficient information regarding the visual components of task-relevant stimuli, it may no longer rely on the sensory cortices for updating stimulus-outcome contingencies. In the early stages of learning, the OFC and relevant sensory cortices might carry similar information about the stimulus. However, as learning progresses,

the OFC might diverge from the sensory cortices and instead align with other higher-order regions, such as the VTA for prediction error and the mPFC for common cause, to update state-wise information of observed and inferred choices. To test this hypothesis, future research will require advanced methodologies that enable the analysis of how different ROIs represent and transfer information over time, with time serving as a gradient-like variable. If this theory holds true, it could provide an explanation for why reactivation in category-specific cortices for choice during feedback was not observed.

Hierarchical causal levels affect learning and abstraction

In Chapter 2, we showed that reinstatement of observed and inferred causal choices supports contingent learning in structured environments. In Chapter 3, we further investigated flexible contingent learning in structured environments with multiple hierarchies of causal levels. Navigation in real-world environments often requires the solving of complex credit assignment problems with more than one or two levels of causal hierarchies. Although a previous study revealed the neural representation of a shared underlying cause among stimuli in the mPFC (Witkowski et al., 2022), the structured environment of the task only consisted of two levels of causal hierarchies. In Chapter 3, we elaborated on the design of a hierarchical learning task consisting of three causal levels and presented behavioral findings demonstrating that learning and abstraction are both dependent at the causal hierarchical level. We found that learning was the fastest and the most robust at the superordinate level, in which only a single governing causal level was pertinent in learning the contingency between stimuli and outcome. Learning was slower at the basic level in which there are two causal levels at stake, and the slowest at the specific level with the greatest number of causal levels to update. Importantly, we

demonstrated that abstraction exhibited different trends across time in terms of accuracy when compared to learning. By abstraction, we use the learned knowledge about the most probable hierarchical level to predict outcomes for novel choices that learners encounter for the first time. We find that, in the early learning periods, abstraction at the superordinate level significantly outperformed those at the basic and specific levels. However, as the estimate of the hierarchical level becomes more robust with more observations of outcomes after decisions, the accuracy of abstraction at the basic and specific levels improves significantly and matches that of the superordinate level.

Results in Chapter 3 laid the groundwork for future work, which involves conducting the same task in an fMRI scanner. We have two hypotheses regarding the neural signatures of contingent learning in hierarchically structured environments. First, we hypothesize that either the OFC or the more general prefrontal cortex contains a gradient representing the distribution over possible causal levels. Previous research demonstrated an anterior-posterior gradient in the OFC and frontal regions, with the anterior portion signaling more abstract rewards and anterior frontal regions (McNamee et al., 2013). Within cognitive control, anterior frontal regions are associated with high-level, abstract cognitive processes such as goal setting and strategic planning, while posterior frontal regions are responsible for more immediate aspects of cognitive control, such as initiation of motor responses (Koechlin et al., 2003). This anterior-posterior gradient could be partially explained by the differences in cytoarchitecture, with the more anterior portion of the frontal cortex having more dispersed connections and the posterior portion having more intrinsic connections (Badre et al., 2009).

Previous studies on the size of place cells' firing fields showed that the topographical hippocampal representations can represent environments at multiple scales. (Kjelstrup et al., 2008). The hippocampal axis was also known to represent information from coarse to fine (Brunec et al., 2018). Built on those previous findings, our second hypothesis proposes that in contingent learning, the representation of the same items will shift along a rostral-caudal gradient in the hippocampus depending on the most likely causal level of a stimulus. The hippocampus has long been recognized to facilitate recollection (Eichenbaum et al., 2007), and its representations flexibly change depending on contexts and goals (Crivelli-Decker et al., 2023). Given its critical role in memory and learning, and the topographical feature of its axis, the hippocampus is likely to support generalization to novel stimuli in contingent learning by mapping onto the most probable belief of the hierarchical causal level, with the superordinate causal level at the rostral portion, and the basic and the specific levels towards the medial and caudal portions. As we discussed in Chapter 3, the task design will enable future research to examine this potential function of the hippocampus by using representational similarity analysis to compare the same stimulus in different hierarchical causal levels.

I would like to conclude this dissertation with two open questions that could be worth asking. First, in real-world credit assignment problems, choices and actions are usually highly abstract and cannot be categorized into concrete entities. At the receipt of outcomes, would the reinstatement of abstract choices be the same as concrete choices? Second, during the receipt of outcomes, it is often the case to attribute the cause to multiple actions and choices, each with a different weight of probability leading to the outcome. How could future research design computational models to capture the distributed and diffuse characteristics in this type

of credit assignment? In closing, these questions highlight the complexity of credit assignment problems and the potential for future research to elucidate its neural and behavioral mechanisms, offering promising trajectories for theoretical and practical advancements in learning and decision-making.

References

Badre, D., D'Esposito, M., Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? Nature Reviews Neuroscience 2009 10:9, 10(9). https://doi.org/10.1038/nrn2667

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M.,…Halgren, E. (2006). Top-down facilitation of visual recognition. Proceedings of the National Academy of Sciences, 103(2). https://doi.org/10.1073/pnas.0507062103

Boorman, E. D., Rajendran, Vani G., O'Reilly, Jill X., & Behrens, T. E. (2016). Two Anatomically and Computationally Distinct Learning Signals Predict Changes to Stimulus-Outcome Associations in Hippocampus. Neuron, 89(6). https://doi.org/10.1016/j.neuron.2016.02.014

Brunec, I. K., Bellana, B., Ozubko, J. D., Man, V., Robin, J., Liu, Z.-X.,…Moscovitch, M. (2018). Multiple Scales of Representation along the Hippocampal Anteroposterior Axis in Humans. Current Biology, 28(13). https://doi.org/10.1016/j.cub.2018.05.016

Carmichael, S. T., & Price, J. L. (1995). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. Journal of Comparative Neurology, 363(4). https://doi.org/10.1002/cne.903630408

Coutanche, M. N., & Thompson-Schill, S. L. (2013). Frontiers | Informational connectivity: identifying synchronized discriminability of multi-voxel patterns across the brain. Frontiers in Human Neuroscience, 7. https://doi.org/10.3389/fnhum.2013.00015

Crivelli-Decker, J., Clarke, A., Park, S. A., Huffman, D. J., Boorman, E. D., Ranganath, C.,…Ranganath, C. (2023). Goal-oriented representations in the human hippocampus during planning and navigation. Nature Communications 2023 14:1, 14(1). https://doi.org/10.1038/s41467-023-35967-6

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The Medial Temporal Lobe and Recognition Memory. Annual Review of Neuroscience, 30(Volume 30, 2007). https://doi.org/10.1146/annurev.neuro.30.051606.094328

Gore, F., Schwartz, E. C., Brangers, B. C., Aladi, S., Stujenske, J. M., Likhtik, E.,…Axel , R. (2015). Neural Representations of Unconditioned Stimuli in Basolateral Amygdala Mediate Innate and Learned Responses. Cell, 162(1). https://doi.org/10.1016/j.cell.2015.06.027

Herry, C., Ciocchi, S., Senn, V., Demmou, L., Müller, C., Lüthi, A.,…Lüthi, A. (2008). Switching on and off fear by distinct neuronal circuits. Nature, 454(7204). https://doi.org/10.1038/nature07166

Izquierdo, A., Suda, R. K., & Murray, E. A. (2004). Bilateral Orbital Prefrontal Cortex Lesions in Rhesus Monkeys Disrupt Choices Guided by Both Reward Value and Reward Contingency. Journal of Neuroscience, 24(34). https://doi.org/10.1523/JNEUROSCI.1921-04.2004

Jocham, G., Brodersen, Kay H., Constantinescu, Alexandra O., Kahn, Martin C., Ianni, A. M., Walton, Mark E.,…Behrens, T. E. (2016). Reward-Guided Learning with and without Causal Attribution. Neuron, 90(1). https://doi.org/10.1016/j.neuron.2016.02.018

Kjelstrup, K. B., Solstad, T., Brun, V. H., Hafting, T., Leutgeb, S., Witter, M. P.,…Moser, M.-B. (2008). Finite Scale of Spatial Representation in the Hippocampus. Science, 321(5885). https://doi.org/10.1126/science.1157086

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The Architecture of Cognitive Control in the Human Prefrontal Cortex. Science, 302(5648). https://doi.org/10.1126/science.1088545

Livneh, U., & Paz, R. (2012). Aversive-Bias and Stage-Selectivity in Neurons of the Primate Amygdala during Acquisition, Extinction, and Overnight Retention. Journal of Neuroscience, 32(25). https://doi.org/10.1523/JNEUROSCI.0323-12.2012

McNamee, D., Rangel, A., O'Doherty, J. P., McNamee, D., Rangel, A., & O'Doherty, J. P. (2013). Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. Nature Neuroscience 2013 16:4, 16(4). https://doi.org/10.1038/nn.3337

Noonan, M. P., Chau, B. K. H., Rushworth, M. F., & Fellows, L. K. (2017). Contrasting Effects of Medial and Lateral Orbitofrontal Cortex Lesions on Credit Assignment and Decision-Making in Humans. Journal of Neuroscience, 37(29). https://doi.org/10.1523/JNEUROSCI.0692-17.2017

Rushworth, M. F. S., Behrens, T.E.J. , Rudebeck, P. H., & Walton, M. E. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. Trends in Cognitive Sciences, 11(4). https://doi.org/10.1016/j.tics.2007.01.004

Schoenbaum, G., Chiba, A. A., Gallagher, M., Schoenbaum, G., Chiba, A. A., & Gallagher, M. (1998). Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. Nature Neuroscience 1998 1:2, 1(2). https://doi.org/10.1038/407

Seo, H., & Lee, D. (2010). Orbitofrontal cortex assigns credit wisely. Neuron, 65(6). https://doi.org/10.1016/j.neuron.2010.03.016

Stalnaker, T. A., Cooch, N. K., Schoenbaum, G., Stalnaker, T. A., Cooch, N. K., & Schoenbaum, G. (2015). What the orbitofrontal cortex does not do. Nature Neuroscience 2015 18:5, 18(5). https://doi.org/10.1038/nn.3982

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 2011 14:12, 14(12). https://doi.org/10.1038/nn.2957

Walton, M. E., Behrens, T. E., Buckley, M. J., Rudebeck, P. H., & Rushworth, M. F. (2010). Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. Neuron, 65(6). https://doi.org/10.1016/j.neuron.2010.02.027

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal Cortex as a Cognitive Map of Task Space. Neuron, 81(2). https://doi.org/10.1016/j.neuron.2013.11.005

Witkowski, P. P., Park, S. A., & Boorman, E. D. (2022). Neural mechanisms of credit assignment for inferred relationships in a structured world. Neuron, 110(16), 2680-+. https://doi.org/10.1016/j.neuron.2022.05.021