Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Exploring the Representational Spaces of States, Values, and Goals in Cognitive Maps

DISSERTATION

for obtaining the doctoral degree
of sciences (Dr. rer. nat.)
at Universität Hamburg
Faculty for Psychology and Human Movement
Institute of Psychology
Cognitive Neuroscience of Learning and Change

submitted by: Nir Moneta

Hamburg, 2025

Doctoral Examination Board:

Chairperson: Prof. Dr. Tanja Lincoln

1st dissertation reviewer: Prof. Dr. Nicolas Schuck

2nd dissertation reviewer: Prof. Dr. Miriam Klein-Flügge

1st disputation reviewer: Prof. Dr. Sebastian Gluth

2nd disputation reviewer: Prof. Dr. Mona Garvert

Date of the disputation: 24.04.2025

Acknowledgments

The last five years have been amazing and challenging. Completing a PhD in these difficult times is a privilege I do not take for granted. I am extremely grateful to everyone who supported me in the process.

Most of all I want to thank Nico Schuck for his support throughout our 7 years journey together. Support that was much more than merely funding, but also opened doors and collaborations and helped me go out to the academic world more prepared. Between lockdowns, endless bureaucracies, broken scanners/projectors, and relocating the lab - Nico always did the best he could to support, much more than many others would do. Thank you, Nico, for the journey behind us, looking forward to the journey ahead.

I am grateful to all the collaborators for inspiring conversations throughout the years, especially Mona and Charley who always took the time to explain, support, and challenge when needed. I am also thankful for helpful feedback on the different parts of the thesis: Angela Langdon and Yael Niv (1st project), Chris Summerfield, Ondrej Zika and Alexa Ruel (3rd project), and Luianta Vera (general introduction and discussion). I am also grateful to the Einstein Center for Neuroscience Berlin for funding 3 years of my PhD. A special thank you to Gregor Caregnato for help with participant recruitment, Sonali Beckmann and Nadine Taube, and numerous members of the Neurocode lab for help with data acquisition, especially Lydia Brundisch and Neele Elbersgerd, and of course all participants for their participation and endurance throughout the different experiments.

A heartfelt gratitude to the Neurocode lab, past and present, for a great time together, inspiring conversations, and support.

Most importantly, I want to thank Malte, my partner. Without him, none of this would have been remotely possible. I could not have dreamed of having a more supportive and loving partner. Thank you.

Table of Contents

List of Figures 1

General abstract 2

General Introduction 5

The role of values in goal-directed behavior . . . 5

Values should guide behavior . . . 6

Not all values should guide behavior . . . 9

Task states might determine which values guide behavior . . . 10

Cognitive maps and goal-directed behavior . . . 12

Cognitive maps as a dynamic representation of the environment . . . 13

The role of values in cognitive maps . . . 15

Rethinking the neural code for values in vmPFC/OFC . . . 17

Representational spaces across the vmPFC/OFC and the hippocampal formation . . . 18

Thesis rationale 20

Project 1: Task state representations in vmPFC mediate relevant and irrelevant value signals and their behavioral influence 23

Introduction . . . 24

Results . . . 27

Behavioral results . . . 28

fMRI results . . . 33

Discussion . . . 45

Methods . . . 49

Project 2: Reward morphs non-spatial cognitive maps in humans 77

Introduction . . . 78

Participants, tasks, and design . . . 83

Results . . . 84

behavioral results . . . 84

Simulations of reward-induced representational change, univariate signals . . . 93

Preliminary fMRI univariate results . . . 97

Simulations of reward-induced representational change, multivariate signals . . . 100

Preliminary multivariate fMRI results . . . 103

Discussion . . . 106

Methods . . . 112

Project 3: Representational spaces in orbitofrontal and ventromedial prefrontal cortex: Task states, values, and beyond **128**
Highlights 129
Main Text 129
Computational and neural underpinnings of value-based decision making 129
All you need is value? 131
Context matters: how tasks shape choice and neural value signals 132
From Context to Task States: how cognitive maps influence values 136
A complex and versatile code in OMPFC 139
A neural network perspective on value and state representations 141
Value and state representations in deep RL models 143
Beyond standard deep RL: flexibility through long-term memory, meta learning and model-based RL 145
Concluding remarks and future perspectives 147

General Discussion **151**
Summary of the different projects 151
Task structure and states govern value representations 153
Are neural representations sub-optimal at separating different value functions? 155
Value generalization triggers representational changes to cognitive maps 157
vmPFC/mOFC: The intersection of values, states, and cognitive maps 160
Compression and competition of alternative goals within the cognitive map 161
Binding it all together: integrating goals and values in dynamic cognitive maps 162
Concluding remarks 167

Supplementary information for 1st project **168**

Supplementary information for 2nd project **202**

References **204**

Declarations related to no another doctoral examination & statutory declaration related to the dissertation **238**

List of Figures

1 Project 1: Task and Design 27
2 Project 1: Behavioral results 30
3 Project 1: RSA analyses show that vmPFC encodes both relevant as well as irrelevant expected values given the current task context 34
4 Project 1: Expected value and context signals co-reside within vmPFC 37
5 Project 1: vmPFC representations of task context, expected value and irrelevant expected value interact 41
6 Project 1: vmPFC representations of context and value jointly guide behavior . . . 44
7 Project 2: Task and Design 82
8 Project 2: Computational model to capture choice bias in the cognitive map 86
9 Project 2: Reward-induced change in behavioral acuity 89
10 Project 2: Generalized reward-induced change in behavioral acuity to non-rewarding areas 91
11 Project 2: Simulation of univariate effects for place-cells-like representation of the cognitive map 94
12 Project 2: Preliminary univariate results support the pulling hypothesis 98
13 Project 2: Simulation of multivariate effects for place-cells-like representations . . 102
14 Project 2: Preliminary multivariate results support the pulling hypothesis 105
15 Project 2: Summary of representational changes in the hippocampus 109
17 Project 3: Values and task states in orbitofrontal and ventromedial prefrontal cortex 133
18 Project 3: Interlinked vmPFC representations of task-state and expected value . . . 135
19 Project 3: Context encoding modulates irrelevant value signals 136
20 Project 3: Principles of deep RL models and emergent representations in late hidden layers 142
21 General discussion: Potential mechanisms for competing goals and their representational spaces 163

General abstract

Maximizing current or future expected rewards is one of the most common goals of many decisions we make. Decisions are always made within the context of a given task. To optimally achieve our goals, we must combine knowledge of the structure of the environment with the current goal to optimally predict different potential rewards. Even the same exact choice can provide different rewards depending on the task at hand. How does the brain combine task demands, environmental structure, and reward maximization in the service of decision-making? Prior works have shown that the ventromedial prefrontal cortex (vmPFC) and adjacent orbitofrontal cortex (OFC) are known to contain signals corresponding to anticipated outcomes of decisions, known as expected value signals, that inform our choices. The hippocampal formation is known to maintain a representation of the environment and potential courses of action in it, known as a cognitive map. This thesis comprises three projects that explore the interaction between task structure, value representation, and cognitive maps within the vmPFC, OFC, and hippocampal formation.

In the first project (Moneta et al., Nature Communications, 2023), we investigate how the vmPFC flexibly switches between different value representations in a task-dependent manner. Thirty-five participants completed a random dot-motion task in which either stimulus color or motion predicted rewards. Multivariate MRI analyses showed that vmPFC signals contain a rich representation that includes the current task state or context (motion/color), the expected value associated with the state, and crucially, the irrelevant expected value of the alternative context. We also find that irrelevant value representations in vmPFC compete with relevant value signals, interact with task-state representations, and relate to behavioral signs of value competition. Our results shed light on vmPFC's role in decision-making, bridging between its role in mapping observations onto the task states of a mental map, and computing expected values for multiple potential states.

In the second project (Moneta et al., in prep.), taking a broader perspective on task structure, we examine how non-spatial cognitive maps are affected by value generalization. A non-spatial

cognitive map refers to mental representations of abstract relationships between different states, helping an agent navigate decision-making by encoding how various choices or actions relate to one another. Animal work has shown that neural representations of spatial cognitive maps are affected by reward. In this project, we tested if similar effects generalize to non-spatial maps in humans. Seventy-two participants (38 of which underwent MRI scanning) performed two sessions of a perceptual discrimination task, before and after extensive reward learning. In all sessions, stimuli varied along two perceptual dimensions, forming a continuous two-dimensional cognitive map. After reward learning, performance in the perceptual discrimination task improved among previously rewarding stimuli. The effect of reward also generalized to areas of the cognitive map that were never rewarding. The precise pattern of changes in perceptual similarity judgments is consistent with the idea that reward learning leads to increased psychological distance between stimuli in the rewarding area, and decreased spacing in neighboring areas. Simulations show that a shift of representational fields towards the rewarded location, akin to a gravitation pulling, can explain the behavioral changes. In line with this, preliminary fMRI data analysis shows evidence for such gravitational pull in the hippocampus and, to some degree, in the medial OFC representations. Future analyses including additional regions in the hippocampal formation and the prefrontal cortex are planned. These results indicate that reward affects non-spatial cognitive maps and suggests accompanying neural representational changes.

In the third project (Moneta, Grossman, Schuck, Trends in Neurosciences, 2024), we review recent literature connecting value and state representations in the OFC/vmPFC, proposing that these regions integrate stimulus, context, and outcome information into a unified representational space. Comparable encoding principles emerge in late layers of deep reinforcement learning models, where single nodes exhibit similar forms of mixed-selectivity, which enables flexible readout of relevant variables by downstream neurons. Based on these lines of evidence, we suggest that outcome-maximization leads to complex representational spaces that are insufficiently characterized by linear value signals that have been the focus of most prior research on the topic. We also discuss major outstanding questions concerning the role of OFC/vmPFC in

learning across tasks, encoding of task-irrelevant aspects, and the role of hippocampus-PFC interactions.

Collectively, these projects shed light on the dynamic relationship between task states and value representations, their integration into a cognitive map, and the representational capacities of the OFC/vmPFC and the hippocampal formation. The findings provide new insights into the neural mechanisms guiding behavior and suggest future research directions.

General Introduction

This thesis explores how the brain integrates value signals, task states, and cognitive maps to support goal-directed behavior. It consists of three interconnected projects investigating the role of the orbitofrontal cortex (OFC), adjacent and overlapping ventromedial prefrontal cortex (vmPFC), and hippocampus in representing states, cognitive maps, and values.

The general introduction provides the conceptual foundation for the thesis. It starts with examining how values influence goal-directed behavior and when they should guide decisions and introduces task states as a potential framework for guiding which values are applicable. It then explores cognitive maps, their relation to task states, their role as flexible environmental representations, and their interaction with values. Lastly, the introduction takes a step back for a broader look into potential theoretical accounts of how values are represented in the brain and interact with other task aspects. After presenting the three projects, the general discussion combines insights from the three projects, addressing the bi-directional influence of task structure and values and the potential roles of vmPFC/OFC and hippocampus in integrating values with states and cognitive maps. The thesis concludes with reflections on the flexibility of goal-directed representational spaces, neuronal mechanisms that might support such flexibility, and future directions.

The role of values in goal-directed behavior

In this thesis, reward refers to a concrete, often tangible outcome of a specific choice, such as the nutritional benefit of eating an apple. Rewards are typically external, measurable events that reinforce behavior. In contrast, value is a more abstract concept reflecting the desirability or expected outcome of an option. Value integrates not only potential rewards but also subjective considerations, such as personal preferences, long-term goals, and situational factors. Expected value (EV) is a more specific term that refers to the expected outcome of a choice, calculated as the sum of possible future rewards weighted by their probabilities. In mathematical terms, EV is objective, as it is a probabilistic calculation. In psychological terms, and this thesis, the focus is usually on more subjective expected values, as they reflect an individual's beliefs about the world

and future outcomes. For example, when choosing between a fresh apple and a chocolate bar, the expected value might consider not only immediate taste but also the belief of future health benefits or indulgence satisfaction. As detailed below, this subjectivity is an important part of understanding value-based behavior and its neural underpinning, as objective and subjective values are not always aligned and other factors (beyond the objective outcome) might influence the desirability of choices.

In reinforcement learning theory, a state represents the current configuration of the environment, encompassing all relevant information needed for decision-making (Sutton & Barto, 1998). States help identify which features are predictive of future rewards (Schuck et al., 2018; Wilson et al., 2014). Within this thesis, context refers to specific features (or feature types) that are predictive of reward and are often used to describe subsets of states. For example, during grocery shopping, in the context of searching for an apple to snack on, textures might be the relevant feature type to focus on, and the specific textures of the apples (e.g. crisp or mealy) are the specific reward-predictive states. When choosing an apple, the expected value of the apple is then the sum of all past experiences we had with apples of similar texture.

Values should guide behavior

Grocery shopping in the supermarket can be a very complex task. Walking down the different aisles, we need to not only make sense of the constant and changing flow of stimuli around us but also make complex choices between them. What guides us in choosing between the different items? A long-believed guiding principle for decision-making is to maximize potential future rewards resulting from our decisions (Peasgood, 2014; Samuelson, 1947; Silver et al., 2021; Sutton & Barto, 1998).

The idea of outcome-maximization goes back centuries to Expected Utility Theory (Daniele Bernoulli, 1738; Peasgood, 2014; Samuelson, 1947), which states that decisions aim to maximize the expected value of a utility function that represents our subjective preference. There has been much work on identifying underlying value functions or decision rules that could potentially guide different types of decisions (Gigerenzer & Gaissmaier, 2011; Kahneman & Tversky, 1979).

Even after centuries of economic, and later neuroeconomic, research, there remains much debate on the shape and nature of such functions, even questioning if they are continuous in nature (Hayden & Niv, 2021). One key shared belief most of these theories have in common is that people would want to choose the option that yields the highest potential outcome and that values are a crucial part of the computations required to do so.

From a neuroscientific perspective, the ventromedial prefrontal cortex (vmPFC) and adjacent orbitofrontal cortex (OFC) are two key areas involved in value processing in humans, primates and rodents (e.g. Padoa-Schioppa & Assad, 2006; Plassmann et al., 2007; Rich & Wallis, 2016), for reviews see (Bartra et al., 2013; Clithero & Rangel, 2014; O'Doherty et al., 2001; Padoa-Schioppa & Assad, 2006). The idea that these regions are crucial for value computation is broadly supported by lesion studies (Ballesta et al., 2020; Fellows, 2007; Hogeveen et al., 2017; Vaidya & Fellows, 2020). Anatomically, vmPFC and OFC can be seen as distinct regions (Rolls et al., 2023) comprised of many subdivisions (Cavada et al., 2000). However, when taking a broader functional perspective, although some studies hint at functional differences between these regions in primates (e.g. Bouret & Richmond, 2010; Castegnetti et al., 2021; M. Z. Wang et al., 2022), the lack of consensus on terminology within and between species makes it challenging to fully separate them. This thesis therefore does not make a harsh distinction between the two regions and aims to focus on their functional roles in guiding decision-making across species (e.g. Levy & Glimcher, 2012; Öngür & Price, 2000).

Value-based decision-making is not a static process; it is dynamic and is influenced by external factors such as learning, experience, and contextual cues. The same apple could have a higher value expectation if we are at a time of trying to keep a strict diet, or a lower value expectation if the last apple we bought was rotten. Values assigned to options are not fixed but instead adapt over time (D. Lee et al., 2012; O'doherty, 2004; Rushworth et al., 2012; Sutton & Barto, 1998). One of the leading and most prominent theories of value-based decision-making is reinforcement learning theory (Sutton & Barto, 1998). A main focus point of reinforcement learning is reward prediction errors which serve to update value representations dynamically. Reward prediction

errors are defined as the differences between expected and received outcomes. In simple terms, the repeated experience of receiving a more satisfying outcome than expected reinforces the subjective value of that option in future decisions, while disappointing outcomes decrease the subjective valuation. This process of value updating was found to be associated with the dopaminergic system and its interactions with the vmPFC/mOFC and striatum (Daw et al., 2006; Frank & Claus, 2006; Montague et al., 1996), which further strengthen their role in dynamically guiding behavior. Current estimates of potential outcomes based on past experiences are often referred to as cashed values. Computationally, keeping cached values allows efficient and rapid decision-making since we don't need to compute them from scratch and retrieve all past experiences every time we encounter a choice.

If values need to be constantly updated and changed, integrated with previous knowledge and future predictions, then regions that are in charge of their representation might do more than only represent a single final value estimate. Indeed, the vmPFC and mOFC are also involved in integrating the expectations from different reward-predicting features of the same object (Basten et al., 2010; Kahnt et al., 2011; Pelletier & Fellows, 2019; Shenhav et al., 2018). At the same time, these regions are involved in optimizing other objective functions such as maximizing confidence (Barron et al., 2015; De Martino et al., 2013; Gherman & Philiastides, 2018; Lebreton et al., 2015), even when it is orthogonal to classic expected values (Shapiro & Grafton, 2020). Additionally, internal states, such as tiredness, modulate choices and also affect values in the brain (Pastor-Bernier et al., 2021; Yoshimoto et al., 2022). The interaction with internal states emphasizes the dynamics of value computation, much like how shopping at the supermarket while hungry can lead to a very different selection of items.

Brought together, these findings highlight the complexity of value-guided behavior, as decision-making is not only a function of immediate reward maximization but rather shaped by multiple interacting cognitive and neural processes.

Not all values should guide behavior

Often we might have more than one goal in an environment, such as shopping for different meals in the same trip to the supermarket. Items on the shelves then might elicit different values, depending on the goal in mind. The same apple might have a different value expectation, depending if the goal is to snack or to bake a cake. Decisions thus are always made within the context of a given task and context is crucial for optimal behavior (see e.g. Hayden & Niv, 2021; Juechems & Summerfield, 2019; Knudsen & Wallis, 2022; Miller et al., 2019; Palminteri & Lebreton, 2021). Even when taking into account all the internal factors and past experiences, the resulting value expectations are not absolute, but rather modulated by the current context. Context could represent immediate influences, such that buying an umbrella on a rainy day might be more valuable than on a sunny day (Berns et al., 2007). The influence of context can also be long-lasting. For example, after participants were extensively trained to make decisions within specific contexts, they were presented with choices between contexts. Participants showed a bias to choose options that were optimal *within their context*, even if they were not the most valued option presented. This bias shows that expected values are context-dependent in a way that can even lead to violation of the principle of value maximization (Bavard & Palminteri, 2023; Bavard et al., 2018; Molinaro & Collins, 2023a; Palminteri et al., 2015). Another important aspect of context is that it is not always clearly observable. When we walk around the supermarket to bake a pie, we don't have the goal explicitly presented in front of us at every given moment. We need to 'keep in mind' that we are here to buy ingredients for a pie and not to bake bread. Partial observability emphasizes goal uncertainty in decision-making, and solving it is key for optimal goal-directed behavior. The OFC/vmPFC was found to represent task context, especially when it is partially observable (e.g. Bradfield & Hart, 2020; Chan et al., 2016; Costa et al., 2023; Elliott Wimmer & Büchel, 2019; Farovik et al., 2015; Muhle-Karbe et al., 2023; Schuck et al., 2016; Wilson et al., 2014; Zhou et al., 2019).

Knowing and keeping track of the context is thus crucial for optimal goal-directed behavior. This raises the question of to what extent context plays a role in neural value representations. Early

studies suggested that OFC value signals are independent of any other task aspect such as sensory features, motor aspects, or other choice options (Padoa-Schioppa & Assad, 2006, 2008; Tremblay & Schultz, 1999). With time, the picture of how the OFC and adjacent vmPFC contribute to decision-making has become more complex. When tasked to only focus on some features and ignore others, lesions of the OFC hinder the ability to ignore irrelevant choice options (Noonan et al., 2017) suggesting the OFC might have a crucial part in the process of identifying context-dependent values. Even when looking at different options within the same context, value signals are normalized within their context within macaque OFC single-cell recordings (Conen & Padoa-Schioppa, 2019; Padoa-Schioppa, 2009), human fMRI (Nelli et al., 2023), and modeling work (Zimmermann et al., 2018), suggesting that values in the OFC are encoded within a certain context and not in an absolute manner.

Brought together, these findings highlight the importance of considering context when evaluating which values should guide behavior. Context can be an influential factor in shaping our perceptions and decisions by guiding which features are important and how values should be evaluated. But what exactly is ’context’? How do we understand it, and how can we define it?

Task states might determine which values guide behavior

States, as previously mentioned, encompass all the observable and non-observable information necessary to predict decision outcomes. Presented with a choice between two sorts of apples at the supermarket, without knowledge of the context, multiple potential states might co-activate, such as ’all the information needed to bake a cake’ or ’all the information needed to snack’. Cognitive control and attention processes support such arbitration between goal-relevant and goal-irrelevant information (Frömer & Shenhav, 2021; MacLeod, 1991; Monsell, 2003; Shenhav et al., 2018) by enhancing features that are relevant given the current task context (Corbetta & Shulman, 2002; Stokes et al., 2013) and shape which features influence expected value representations in vmPFC (Castegnetti et al., 2021; Frömer et al., 2019; Leong et al., 2017; Niv et al., 2015; Rudebeck & Murray, 2014). Computational modeling and multivariate imaging studies suggest that the vmPFC/OFC not only encode values but also track the structure of the

environment, potentially segmenting experiences into distinct task states that can then guide future decisions (Chan et al., 2021; Nassar et al., 2019; Schuck et al., 2016; Wikenheiser & Schoenbaum, 2016). Recent evidence suggests that states are not merely passive containers of information but actively shape neural representations in value-based decision-making. In particular, the vmPFC and OFC appear to encode latent state representations that integrate contextual cues to guide behavior (Baram et al., 2021; Cromwell et al., 2018; Farovik et al., 2015; Schuck & Niv, 2019; Schuck et al., 2016; Z. Zhang et al., 2017). These regions are thought to maintain an abstract, high-dimensional representation of task states, dynamically updating them based on experience and prediction errors (Niv, 2019; Wilson et al., 2014). This flexibility allows for rapid adaptation to environmental changes, ensuring that decision-making remains optimal even when faced with uncertainty (Momennejad et al., 2017; Piray & Daw, 2021).

States are thus a good framework to understand what context is and help determine which features should predict values, and as a result which values should guide behavior. However, such state-separation might not be as clean as traces of processing of irrelevant features have been found in several cortical regions, including areas responsible for task execution (V. Li et al., 2018; Mante et al., 2013; Schuck et al., 2015; Shahar et al., 2019; Takagi et al., 2020). At the same time, vmPFC was shown to be critical to evaluating counterfactual choices (e.g. ”what might have been” Levens et al., 2014), and to integrate counterfactual signals during decision-making, allowing for the comparison between actual outcomes and potential alternatives (Tobia, Guo, Schwarze, Böhmer, et al., 2014). Similarly, not only task-relevant but also task-irrelevant valuation has been shown to influence cognitive control (Anderson, 2013; Frömer et al., 2021) as well as activity in vmPFC (Lebreton et al., 2009), suggesting multiple states might simultaneously contribute to guiding behavior.

Brought together, when thinking about the computations and representational space required for optimal decision-making, values are not enough. They must come with the right context, as buying a metal chair is only valuable if you want to sit, not if you need to light a fire. It is thus likely that context would directly influence and interact with value representations to guide which

are the correct values to guide behavior. The first project in this thesis focuses on the relationship between states, relevant and irrelevant values, and their underlying neural interaction.

Cognitive maps and goal-directed behavior

Structural knowledge of the world around us, and specific experiences we had in it, play a crucial role in guiding our behavior. To efficiently handle a shopping trip at the supermarket, we need more than just the goal in mind. We need to keep track of where we are at every moment, which items we already saw, where we are headed and even potentially make predictions of where desired items might be located. Humans and animals can adapt their behavior to achieve goals and have the ability to generalize from specific past experiences with the same goal (Daw et al., 2005; Gershman et al., 2014). They can also draw on experiences that are only somewhat connected or even envision new decisions and their outcomes, even if those were never experienced (Barron et al., 2013; Gupta et al., 2010; Schacter et al., 2007).

Already decades ago, inspired by rats making unexperienced shortcuts in complex mazes to reach rewards, Tolman coined the term ’cognitive map’ referring to the internal organization and representation of the world around us (Tolman et al., 1946; Tolman & Honzik, 1930b). In their broad sense, cognitive maps can be seen as the structured internal representations of relationships among different states (Behrens et al., 2018). Decades of research revealed a set of medial-temporal and medial-prefrontal representations that could construct such a spatial cognitive map. The most prominent discovery was hippocampal place cells which fire at specific locations of the map (Moser et al., 2008; O’Keefe & Dostrovsky, 1971) followed by entorhinal grid cells which fire at different locations forming a hexagonal grid (Doeller et al., 2010; Gardner et al., 2022; Hafting et al., 2005). Alongside place and grid cells, more and more cell ’types’ were discovered in the hippocampal formation, such as cells encoding goals (Sarel et al., 2017), head direction (Taube et al., 1990), distance to borders (Solstad et al., 2008) and even cells dedicated to encoding locations of rewards (Gauthier & Tank, 2018). Together, these cells are seen as forming a coordinate system, incorporating knowledge from the environment for optimal navigation (Kaplan et al., 2017; Sharpe et al., 2019; Theves et al., 2019).

Much of the focus on cognitive maps and their underlying representations has been on navigating physical space (Behrens et al., 2018; O'Keefe, 1978). Recent work, however, suggests that this system is not restricted to physically navigating in space and that the same system is also engaged when imagining navigation (Bellmund et al., 2016; Horner et al., 2016; Neupane et al., 2024). Even more strikingly, the same system can represent much more abstract concepts and dimensions, such as visual features (Constantinescu et al., 2016; Theves et al., 2019, 2020), odors and sounds (Aronov et al., 2017; Bao et al., 2019; Radvansky & Dombeck, 2018; Terada et al., 2017), choice probabilities and magnitudes (Bongioanni et al., 2021) and even decisions along two-dimensional social hierarchies (Park et al., 2020, 2021). Cognitive maps thus allow for the spatial and non-spatial organization of experiences and knowledge, and play a crucial role in goal-directed behavior (Behrens et al., 2018; Bellmund et al., 2018; Epstein et al., 2017; O'Keefe, 1978; Schuck et al., 2016; Stachenfeld et al., 2017; Tolman, 1948; Wilson et al., 2014).

As described above, states can be conceptualized as foundational units upon which agents base their actions to maximize rewards. Cognitive maps, on the other hand, can be thought of as internal representations that encode the relationships between different states. The interplay between states and cognitive maps is crucial for understanding how agents learn and make decisions (OC Jordan et al., 2020; Piray & Daw, 2021). The integration of reinforcement learning principles with the concept of cognitive maps offers a comprehensive framework for understanding how agents navigate complex environments (Behrens et al., 2018; Momennejad et al., 2017; Stachenfeld et al., 2017; Stoewer et al., 2022). If states are related to cognitive maps, we might be able to draw parallels between findings on states to enhance our understanding of cognitive maps.

Cognitive maps as a dynamic representation of the environment

To optimally guide our behavior, cognitive maps might exhibit a similar dynamic nature as state representation. One specific similarity cognitive maps might draw from state representations, is filtering and sorting feature relevancy for optimal behavior. When optimizing a representation to achieve a goal, disregarding or compressing certain feature dimensions could be beneficial for

achieving a goal (Flesch et al., 2022; Muhle-Karbe et al., 2023). For example, the shape of the leaves shouldn't influence the apple's taste and therefore a trained shopper might reduce the representation of different sorts of apples on the shelf to a single gradient of color, ignoring the leaves. Previous work indeed suggests that neural representations are reduced in dimension in proportion to feature relevance (Mack et al., 2020). At the same time, contextually irrelevant information is not completely filtered out, even after extensive amounts of practice (Mante et al., 2013; Takagi et al., 2021). We might even keep track and associate value to task aspects that have never been related to our goal (Ben-Artzi et al., 2023; Shahar et al., 2019). It is as if we know the color of the apple predicts its taste, but still keep track of the shape of its leaves and even the hand we use to pick it from the shelf. This suggests that while optimizing cognitive maps for goal achievement, goal-irrelevant features might persist, potentially impacting behavior. Recent findings indeed found that the hippocampus, while encoding complex transition probabilities between stimuli, also simultaneously represented their semantic relations, even though those were irrelevant to the task at hand (Zheng et al., 2024). These results raise the possibility that also on the cognitive map level, previous goals and task-irrelevant aspects might still influence the broad representation of the decision space, i.e. the cognitive map.

Different goals might require different optimizations of the decision space, which, in turn, elicit various representational changes to the cognitive map (Muhle-Karbe et al., 2023). For example, after tasting many apples, an agent might learn that red apples are sweet and green apples are sour. To maximize sweetness, the agent might need to find the optimal shade of red. Reinforcement learning offers methods such as function approximation to address this problem, allowing for the representation of continuous features like color (Sutton & Barto, 1998). Rather than defining discrete states like 'choosing sweet red apples' or 'choosing sour green apples', function approximation enables the agent to treat color as a continuous variable. This approach generalizes learned value functions across similar states using mathematical functions (such as linear or neural network models), which allows the agent to estimate values for states it has not directly encountered by learning the relationships between different states and their outcomes. If

states have a role in emphasizing the right value, as hypothesized above, then when thinking broadly on the cognitive map, value generalization over unseen states could prompt changes to the cognitive map to emphasize rewarding areas. For instance, it might cause a gradual increase in representation in specific rewarding areas, e.g. enhancing acuity around the optimal sweetest red-shade goal. Moreover, such representational changes might not be restricted to the rewarded area, but could also influence the entire map, affecting the entire ’color’ axis (Schaffner et al., 2023).

Outside (and adjacent to) reinforcement learning, the idea of how continuous representations are clustered into discrete categories has a long history. However, there is conflicting evidence for the influence of applying discrete categories to the perception of continuous stimuli, sometimes causing an increase and sometimes a decrease in discriminability (e.g. Goldstone, 1994; Harnad, 2003; Liberman et al., 1957; McMurray, 2022; Thalmann et al., 2024). These findings raise the question of how continuous cognitive maps are influenced by goals and how continuous and categorical structures interact over time.

Brought together, cognitive maps allow individuals to organize knowledge of the environment to guide their behavior. One mechanism they might use is organizing information into current states and even predicting future states. Having a more dynamic, experience-driven representation of cognitive maps (and states) can help us better achieve different goals. Being able to predict rewards is a crucial part of goal-directed behavior (Samuelson, 1947; Silver et al., 2021; Sutton & Barto, 1998). Therefore the interplay of states, cognitive maps, and values is important to understand how humans and animals navigate complex environments to achieve their goals.

The role of values in cognitive maps

Many approaches thus far focused on the separation between value and the structure of the environment. One prominent example is the two-step task, a widely used experimental paradigm to study the interaction of structure and reward. In this task, participants make sequential choices, with each decision influencing subsequent states and potential rewards. This setup distinguishes between model-free learning, which relies on directly cached values from past rewards associated

with specific stimuli or actions, and model-based learning, which involves building a cognitive map of the task structure to predict future outcomes (Daw et al., 2011). The successor representation is another computational framework that bridges model-free and model-based learning through the separation of value and task structures (Dayan, 1993; Momennejad et al., 2017). Successor representation encodes the expected future states, not values, an agent will visit, combining the temporal and spatial structure of the environment. Conceptually, SR can be thought of as a ’value-free’ representation of the environment which allows flexibility to update value prediction when those change. Neuroscientific research shows support for the role of successor representations in the brain’s navigation and decision-making systems (Momennejad et al., 2017; Stachenfeld et al., 2017; Stoewer et al., 2022; Wittkuhn et al., 2022).

Since states incorporate additional information beyond values, they can help guide decisions for more optimal future outcomes. The approach to separate value and structure can be very beneficial to understanding the dependency of value on states. Both the concept of ’states’ as well as successor representations share the notion that the learned structure of the environment should influence and emphasize the relevant value-predicting features. This flexibility is crucial for optimal behavior and correct generalizations. For example, learning the structure of the supermarket while buying ingredients for a cake, will also help us when we buy products for a BBQ. However, it raises the question of whether such influences are not bi-directional. Could values directly shape the representation of the structures of the environment? If one area of the supermarket has our favorite sweets, which we value more than the hardware materials located elsewhere. Will the representation of the supermarket remain agnostic to this aspect? It is possible that over time, repeated exposures to reward in the same location would also influence the underlying representation of the structure, suggesting they might not be as separated.

What role do values play in cognitive maps? One way to think of values is as another dimension along which information is organized, similar to other task dimensions (Bongioanni et al., 2021; Nitsch et al., 2024). Areas in the medial temporal lobe, associated with representing cognitive maps, are also involved in representing rewards (LeGates et al., 2018; Wirth et al., 2009;

Zeithamova et al., 2018). However, as mentioned above, values might play a bigger role in shaping cognitive maps and their underlying representations. Supporting this, studies found that the activity of place and grid cells are directly influenced by reward (Gauthier & Tank, 2018). Place cells can conditionally respond only in a certain context, a larger number of place cells represent areas around reward versus non-reward locations, and even grid cells change in firing rate and location too (Boccara et al., 2019; Dupret et al., 2010; Moita et al., 2003; Sosa & Giocomo, 2021). This raises the possibility that representations of value and other aspects of the map are not fully independent and might influence each other (e.g. Garvert et al., 2023). The second project aims to investigate how knowledge of reward re-organizes cognitive maps which in turn guide behavior.

Rethinking the neural code for values in vmPFC/OFC

Cognitive maps and states exhibit flexible dynamics to organize knowledge and guide our decisions. Identifying the (right) values is a crucial part of this process. At the same time, values are strongly dependent on other task-related factors such as context. Task context might not only shape how values are computed (1st project), but values themselves might reshape how the task is represented (2nd project). This interplay raises a broader long-investigated question: how are values represented in the brain? Are values represented independently or are they a part of a more encompassing representational space? To try and shed light on these questions, in the last project, we reviewed recent literature focused on the vmPFC and mOFC as key regions found to represent value signals (Bartra et al., 2013; Clithero & Rangel, 2014; O'Doherty et al., 2001; Padoa-Schioppa & Assad, 2006).

One prominent idea, inspired by early economic theory (e.g. Samuelson, 1947), is that the vmPFC encodes expected values of different features in a common currency, i.e. a stable desirability scale that guides decisions (Chib et al., 2009; Fehr & Rangel, 2011; Levy & Glimcher, 2012). This notion is supported by evidence that vmPFC value signals can be decoded across tasks with different goals, suggesting that values might be task-independent (Castegnetti et al., 2021; Frömer et al., 2019; Gross et al., 2014; Howard et al., 2015; Kobayashi & Hsu, 2019; McNamee et al.,

2013; Westbrook et al., 2019; Yao et al., 2023; Z. Zhang et al., 2017). However, at the same time, as predicted by the first project and previously mentioned findings, context might interact with these signals, suggesting that values may not be encoded independently, but rather in relation to other task information.

Evidence suggests that neurons in vmPFC encode summary statistics of the current task such as previous offers and outcomes, or the location of the currently attended offer (Mehta et al., 2019). Recording studies have also shown that the same neurons in OFC often encode multiple variables at once, a phenomenon known as mixed selectivity (Rigotti et al., 2013). For instance, the same neurons in macaque OFC can represent both spatial and reward information, even when those are unrelated (Yoo et al., 2018). This raises a challenge to the notion of a single, task-independent value signal: if value representations are interwoven with other task-related signals, even within single neurons, how should we conceptualize their role in the representational space that guides our decisions? The question of how context and values might co-vary was one of the focus points of the first project. The last project takes a broader perspective, synthesizing recent literature to reconcile common-currency accounts of value representation with frameworks emphasizing state representations in vmPFC/mOFC. Drawing from encoding principles observed in late layers of deep reinforcement learning models, where individual nodes exhibit similar mixed selectivity, we explore how value representations may be embedded within a broader representational space. We suggest that outcome maximization leads to complex representational spaces, focused but not restricted to vmPFC/mOFC, that are insufficiently characterized by simple linear value signals.

Representational spaces across the vmPFC/OFC and the hippocampal formation

Although the third project focused mainly on vmPFC/OFC, the hippocampal formation and its interaction with vmPFC/OFC also play an important role in representing task spaces to guide behavior. The hippocampus has a well-established role in long-term memory and cognitive maps (Behrens et al., 2018; Sosa & Giocomo, 2021) and is crucial for model-based planning (Miller et al., 2017). At the same time, state representations in the vmPFC/OFC can be understood as encoding one's current location within a cognitive map (Niv, 2019; Schuck et al., 2016). These

regions share strong anatomical and functional connectivity (Öngür & Price, 2000) and have been shown to represent spatial locations simultaneously in rodents (Tang et al., 2021; Zielinski et al., 2019). In humans, their co-activation has been also linked to constructing complex mental imagery, such as scenes (Monk et al., 2021), usage of tools for specific goals (Castegnetti et al., 2021), and autobiographical memory recall (McCormick et al., 2020). Together, the hippocampal formation and the OFC contribute to constructing spatial and abstract cognitive maps, and trajectories within them, to guide decision-making (Park et al., 2020). Supporting this idea, hippocampal replay, i.e. reactivation of sequences of neural activity corresponding to prior experiences (for review, see Wittkuhn et al., 2021), has been associated with the representation of task states in the OFC (Schuck & Niv, 2019) and the OFC was shown to dynamically updates hippocampal representations to optimize choice inference (Garvert et al., 2023). Together, these suggest a collaborative mechanism between the hippocampus and OFC in consolidating task-related information which likely plays a key role in flexible, goal-directed behavior. While this thesis only briefly touches on this idea in the third project, and the current results of the second project are more focused on these regions separately, we hope that further work on the second project will provide additional insights into more regions in the hippocampal formation and how hippocampus-OFC interactions guide behavior.

Thesis rationale

Understanding how task goals, value representations, and cognitive maps work together to guide our behavior is a fundamental question in cognitive neuroscience. Are values computed independently of task structure, or do they emerge within a structured representation of the environment? Does exposure to reward shape how we organize task-relevant information, and if so, how? How do regions such as the vmPFC, adjacent OFC, and hippocampus integrate these elements into a joint representational space to guide us to our goals?

This thesis attempts to shed some light on these long-standing questions by investigating the bidirectional relationship between task structure and value representation, focusing on their neural substrates. It approaches this from three angles: how does task structure shape value representations; how value generalization influences task structure; and how these elements may be integrated into a unified representational space.

Project 1: How does task structure shape value representations?

In the first project, we ask whether the vmPFC represents values in a strictly task-relevant manner or whether it also encodes other aspects of the task simultaneously. Specifically, does the vmPFC encode only the values that are relevant for the current choice, or does it also maintain information about alternative, context-inappropriate values? In the case of the latter, what can this tell us about the underlying structure of value representations in the vmPFC?

To test this, we designed a task in which either stimulus color or motion predicted reward, arbitrating between two possible contexts (or: states). We hypothesized that during choice, vmPFC will represent different values that occur in different task contexts, i.e. values appropriate in the current context, as well as other, context-inappropriate and therefore choice-irrelevant values. We predicted that (1) vmPFC will represent values in a state-dependent manner. If that is the case, we would expect that its signal and behavior would be mainly guided by the expected value of potential choices within each context: the relevant context and the counterfactual choice one would have made in an alternative state. Furthermore, if the vmPFC encodes both values and context, we would expect to find (2) overlapping signals of value and context that interact with

feature relevance to shape behavior.

Our findings suggest that task state representations govern value computations in vmPFC, such that when faced with multiple potential states, a comparison is triggered in this region between the expected values according to each state.

Project 2: How does value generalization influence cognitive maps?

If other task representations, such as states, influence value representations, can the reverse also be true? Does value influence broader task representations, such as cognitive maps? And specifically in cases where value is not observed but generalized from similar learning experiences? As mentioned above, long exposure to specific goals or rewards can cause changes to perceptual discriminability. Such changes might reflect broader influences on the representation of the cognitive map. In the second project, we investigate how reward exposure affects non-spatial cognitive maps in humans. We ask if exposure to reward can cause changes in how task-relevant dimensions are represented in a perceptual task on the same and the next day.

To address these questions, across four MRI sessions, participants performed a perceptual discrimination task before and after extensive reward learning. We hypothesized two levels of generalization: (1) across time, such that areas associated with reward would show increased behavioral and neural acuity even in later sessions when reward was absent; and (2) across space, where the effects of reward would induce a systematic reorganization of the cognitive map, affecting areas that were never directly rewarded.

We simulated potential mechanisms underlying changes in representational spaces due to reward, considering three main possibilities: (1) Partial remapping, in which a random subset of representational fields centers around reward. (2) Global (value-driven) remapping, where all representational fields across the space remap with respect to distance from reward, causing the signal to scale by distance to reward (equivalent to generalized value map), and (3) a reward-driven ”gravitational pull” where representational fields shift based on proximity to reward. Behavioral and preliminary fMRI results suggest a pulling mechanism, with evidence of this effect in both the hippocampus and medial OFC. This project is ongoing and therefore all the

results and conclusions are preliminary and should be taken with a grain of salt.

Project 3: How are task and value representations organized into a joint representational space in the OFC/vmPFC?

After investigating the bidirectional influence of task structure on value representation and vice versa, the final project asks how task structure, states, and values are integrated to guide our behavior. We reviewed recent literature while focusing mainly on the vmPFC/OFC. We suggest a strong connection between value and state representations and argue that the OFC/vmPFC integrates stimulus, context, and outcome information into one representational space to guide behavior. We explore theoretical frameworks, including deep neural networks, that provide potential analogies for how these representations are structured. We suggest a shift of focus from treating value as an independent, overarching signal, to a framework that emphasizes state representations as the organizing principle of signals in these regions and their relation to behavior. This suggests that values are only one part of a more complex manifold of representations that all reside within the same region and guide our behavior.

A unified framework of how task, values, and cognitive maps shape neural representations to guide behavior

Bringing these projects together, this thesis presents a perspective of bidirectional influences of task and value representations. The first project demonstrates that task state representations shape how value is computed in vmPFC. The second project shows that value exposure, in turn, reshapes task representations. Finally, the third project proposes that these elements are not independent but rather part of a single representational space in the OFC/vmPFC, which flexibly integrates task and value information to guide current and future behavior.

These findings highlight the interaction between state-dependent representations, value computation, and their role in shaping cognitive maps. Future directions include investigating the precise mechanisms governing these interactions, exploring how generalizable these findings are across different task domains, and further disentangling the roles of OFC, vmPFC, and hippocampus in dynamically structuring and guiding behavior.

Project 1: Task state representations in vmPFC mediate relevant and irrelevant value signals and their behavioral influence

Work with Mona M. Garvert, Hauke R. Heekeren, H. R., & Nicolas W. Schuck.

Published at: Nature Communications (2023) DOI: https://doi.org/10.1038/s41467-023-38709-w

Introduction

Decisions are always made within the context of a given task. Even a simple choice between two apples will depend on whether the task is to find a snack, or to buy ingredients for a cake. In other words: The same objects can yield different outcomes in different task contexts. This could complicate the computations underlying retrieval of learned values during a decision, since outcome expectations from the wrong context might exert influence on the neural representation of the available options.

Which reward a choice will yield in a given task context is at the core of many decisions (e.g. Kahneman & Tversky, 1979). Ventromedial prefrontal cortex (vmPFC) represents this so-called expected value (EV) in a variety of species (Bartra et al., 2013; Clithero & Rangel, 2014; O'Doherty et al., 2001; Padoa-Schioppa & Assad, 2006; Plassmann et al., 2007; Rich & Wallis, 2016), and thereby is crucial in determining choices (Ballesta et al., 2020). Several investigations have also shed light on how the brain maps from complex sensory input to expected values, and the associated cognitive control processes. It is known, for instance, that the brain's attentional control network enhances the processing of features that are relevant given the current task context or goal (Corbetta & Shulman, 2002; Stokes et al., 2013), which in turn helps shape which features influence expected value representations in vmPFC (Castegnetti et al., 2021; Frömer et al., 2019; Leong et al., 2017; Niv et al., 2015; Rudebeck & Murray, 2014). Moreover, vmPFC seems to also represent expected value of different features in a common currency (Chib et al., 2009; McNamee et al., 2013); and is involved in integrating reward expectations from different features of the same object (Basten et al., 2010; Kahnt et al., 2011; Pelletier & Fellows, 2019; Shenhav et al., 2018). It remains unclear, however, how context-irrelevant value expectations of available features, i.e. rewards that would be obtained in a different task-context, might affect vmPFC signals, and how such "undue" influence relates to wrong choices.

This is particularly relevant because we often have to do more than one task within the same environment, such as shopping in the same supermarket for different purposes. Cognitive control processes are known to arbitrate between relevant and irrelevant information (MacLeod, 1991;

Monsell, 2003), and it has been suggested that they also gate the flow of information within the value network (Frömer & Shenhav, 2021; Shenhav et al., 2018). But although cognitive control does gate relevant information, it is also known that task-switching leads to less than perfect separation between task contexts/goals (Monsell, 2003) and results in processing of task-irrelevant aspects (MacLeod, 1991). Several studies found traces of the distracting features in several cortical regions, including areas responsible for task execution (V. Li et al., 2018; Mante et al., 2013; Schuck et al., 2015; Shahar et al., 2019; Takagi et al., 2020). Similarly, not only task-relevant but also task-irrelevant valuation has been shown to influence cognitive control (Anderson, 2013; Frömer et al., 2021) as well as activity in vmPFC (Lebreton et al., 2009) and posterior parietal cortex (Grueschow et al., 2015). We therefore hypothesized that during choice the vmPFC will represent different values that occur in different task contexts, i.e. values appropriate in the current context, as well as other, context-inappropriate and therefore choice-irrelevant values. Importantly, unlike in standard cognitive control settings, we asked whether the above mentioned control during value-based choice involves the arbitration between the expected values that would result from the counterfactual choices one would have made in another context.

If that is the case, the neural representation of context might play a major role in gating context-dependent values in vmPFC. Previous work has shown that vmPFC is involved in representing such context-signals (Chan et al., 2016; Schuck et al., 2016, 2018; Wimmer & Büchel, 2019), which suggests that its role goes beyond representing attention-filtered values. Note that knowing the current context alone will not immediately resolve which value of two presented options should be represented, similar to how knowing what you are shopping for (cake or snack) will not answer which of the available apples you should pick. We therefore hypothesized that vmPFC would have a role that goes beyond only encoding the task context, namely that it would also be involved in the arbitration between context-dependent values, meaning that a stronger activation of the relevant task-context will also enhance the representation of task-relevant values. Such a multifaceted representation of multiple values and

task contexts within the same region would reconcile work that emphasizes the role of choice value representations in vmPFC and orbitofrontal cortex (OFC) (Ballesta et al., 2020; Bartra et al., 2013; Clithero & Rangel, 2014; O'Doherty et al., 2001; Padoa-Schioppa & Assad, 2006; Plassmann et al., 2007; Rich & Wallis, 2016) with work which emphasizes the encoding of other aspects of the current task (Constantinescu et al., 2016; Doeller et al., 2010; Mack et al., 2020; Schlichting et al., 2015; Schoenbaum & Roesch, 2005), in particular of so-called task states (Chan et al., 2016; Schuck et al., 2016, 2018; Wimmer & Büchel, 2019), within the same region (see also, Farovik et al., 2015; Zhou et al., 2019). More specifically, we propose that context/task state representations influence value computations in vmPFC, such that a state representation triggers a comparison between the values of options as they would be expected in the represented state/context. In consequence, the value of the option that would be best in the activated state will become represented, and partial co-activation of different possible states could therefore lead to value representations that can refer to different choices (the value of the apple best for snacking and the value of the apple best for baking, even if those are different apples). An alternative view in which state representations do not impact value computations would assume that activated values would always refer to the choice one is going to make in the present context (how valuable the apple chosen for snaking would be for baking).

We investigated these questions using a multi-feature choice task in which different features of the same stimulus predicted different outcomes, and a task-context cue modulated which feature was relevant. We show that participants compute both value expectations of the relevant context as well as value expectations of an additional, explicitly cued-to-ignore, irrelevant context. Behavioral analyses indicated a choice conflict modulated by the possible expected values of the relevant and irrelevant context. Multivariate fMRI signals in a vmPFC value ROI were sensitive to (a) relevant values, (b) contextually irrelevant values and (c) the identity of the current context. We also found that increased representation of irrelevant values during choice were accompanied by a decreased representation of the relevant values, indicating a value competition in vmPFC. This competition was modulated by the task-context signal found in vmPFC. Lastly, we found

that neural indicators of context, values and the competition between them were linked to increased choice conflict. We suggest that information within the vmPFC is organized into a complex multi-faceted representation in which multiple values of the same choice under different task-contexts are co-represented and compete in guiding behavior, while a context (or state) signal might act as a moderator of this competition.

Results

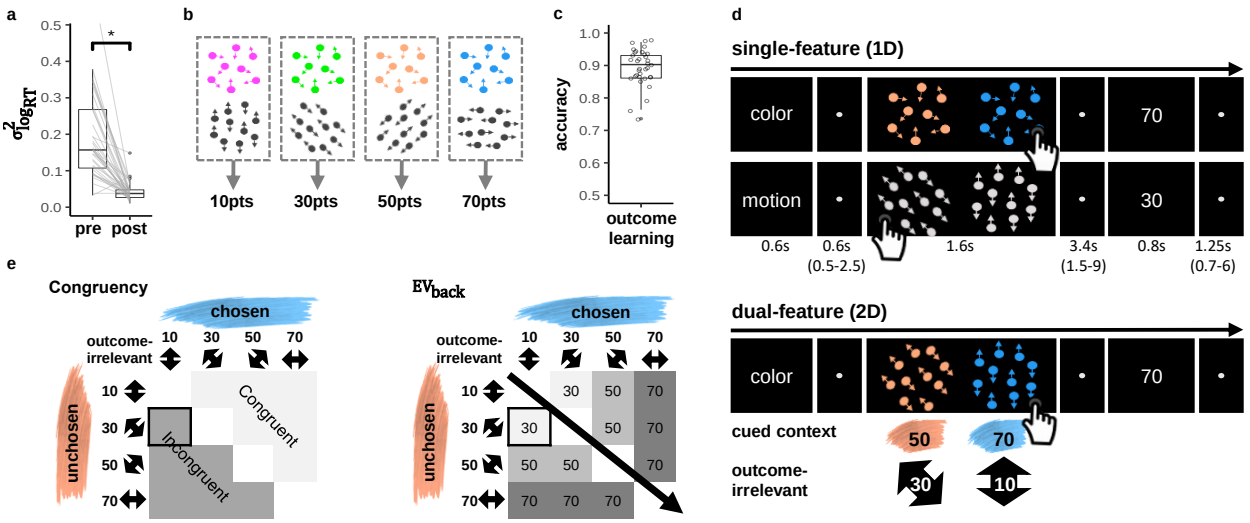


Figure 1

Task and Design a. Prior to value-learning, a participant-specific staircasing procedure adjusted color and motion parameters such that variance of reaction times across different color and motion features (y-axis) was reduced (paired t-test, p<.001, n=35). Box covers interquartile range (IQR), mid-line reflects mean, whiskers the range of the data (until +-1.5*IQR), and solid points represent outliers beyond whiskers. **b.** After staircasing, specific rewards were assigned to each of the four color and four motion directions, such that one feature from each context was associated with the same reward/value. Feature-value mapping was counterbalanced across participants. **c.** Participants achieved near ceiling accuracy in choosing the highest valued feature after training (μ=.89, σ=.06, n=35). Boxplot as in **a**. **d.** Single-feature (1D, top) and dual-feature (2D, bottom) trials both started with a cue of the relevant context ("Color" or "Motion", 0.6s), followed by a fixation (0.5s-2.5s, μ=0.6s) and a choice between two clouds (1.6s). In 1D trials, each cloud only had one relevant feature (colored dots, but random motion, or directed motion, but gray dots), while in 2D trials each cloud had a motion and a color feature. Participants were explicitly asked to select the option yielding the highest outcome in the cued context and ignore irrelevant features. Then followed another fixation (1.5s-9s, μ=3.4s) and the value associated with the chosen cloud's feature of the cued context (outcome, 0.8s). The next trial started after another fixation (0.7s-6s, μ=1.25s). **e.** Experimental manipulation of irrelevant values in 2D trials. For each relevant feature pair (e.g. blue and orange), all possible context-irrelevant feature-combinations were included in the task, except same feature on both sides. Congruency (left): trials were termed congruent when irrelevant features favored the same choice as the relevant features, otherwise incongruent. EV_{back} (right): trials were also characterized by the hypothetical expected value of contextually-irrelevant features, i.e. the maximum value of both irrelevant features. NB that both aspects did not have any impact on outcomes and were irrelevant for the task at hand and that EV, EV_{back} and Congruency were orthogonal by design. Highlighted cell reflects example trial in **d**, bottom.

Behavioral results

Thirty-five right-handed young adults (18 women, $\mu_{age} = 27.6, \sigma_{age} = 3.35$, see Methods for exclusions) were asked to judge either the color (context 1) or motion direction (context 2) of moving dots on a screen (random dot motion kinematogramms, e.g. Pilly & Seitz, 2009). Four different colors and motion directions were used. Before entering the MRI scanner, participants performed a stair-casing task in which participants first received a cue that instructed them which feature (a color or direction) will be the target of the current trial. Then participants had select the matching stimulus from two random dot motion stimuli (see Fig.S1c). In this task, motion-coherence and the speed which dots changed from grey to a target color were adjusted such that the different stimulus features could be discriminated equally fast, both within and between contexts (i.e. Color / Motion, Fig.S1c). As intended, this led to significantly reduced differences in reaction times (RTs) between the eight stimulus features, within and between contexts (paired t test on RT variance before and after the staircasing: $t_{(34)} = 7.29, p < .001$, Fig.1a), also when tested for each button separately ($t_{(34)} =$ Left: 6.52, Right: 7.70, ps$< .001$, Fig.S1d).

Only then, participants learned to associate each color and motion feature with a fixed number of points (10, 30, 50 or 70 points), whereby one motion direction and one color each led to the same reward (counterbalanced across participants, Fig.1b). To this end, participants made choices between clouds that had only one feature-type, while the other feature type was absent or ambiguous (clouds were grey in motion-only clouds and moved randomly in color clouds). To encourage mapping of all features on a unitary value scale, choices in this part (and only here) also had to be made between contexts (e.g. between a green and a horizontal-moving cloud). Participants achieved near-ceiling accuracy in choosing the cloud with the highest valued feature ($\mu = .89, \sigma = 0.06$, t-test against chance: $t_{(34)} = 41.8, p < .001$, Fig.1c), also when tested separately for color, motion and across context ($\mu = .88, .87, .83, \sigma = .09, .1, .1$, t-tests against chance: $t_{(34)} = 23.9, 20.4, 19.9$, ps$< .001$, respectively, Fig.S1e). Once inside the MRI scanner, one additional training block ensured changes in presentation mode did not induce

feature-specific RT changes (Anova on mean RT for each feature: $F_{(7,202)} = 1.06$, $p = 0.392$). These procedures made sure that participants began the main task with firm knowledge of feature values; and that RT differences would not reflect perceptual differences, but could be attributed to the associated values. Additional information about the pre-scanning phase can be found in Methods and in Fig.S1.

During the main task, participants had to select one of two dot-motion clouds. In each trial, participants were first cued whether a decision should be made based on color or motion features, and then had to choose the cloud that would lead to the largest number of points. Following their choice, participants received the points corresponding to the value associated with the chosen cloud's relevant feature. To reduce complexity, the two features of the cued task-context always had a value difference of 20, i.e. the choices on the cued context were only between values of 10 vs. 30, 30 vs. 50 or 50 vs. 70. One third of the trials consisted of a choice between single-feature clouds of the same context (henceforth: 1D trials, Fig.1d, top). All other trials were dual-feature trials, i.e. each cloud had a color and a motion direction at the same time (henceforth: 2D trials, Fig.1d bottom), but only the context indicated by the cue mattered. Thus, while 2D trials involved four features in total (two clouds with two features each), only the two color or two motion features were relevant for determining the outcome. The cued context stayed the same for four to seven trials. Importantly, for each comparison of relevant features, we varied the values of the irrelevant context, such that each relevant value was paired with all possible irrelevant values (Fig.1e). While the irrelevant context in a trial did not impact the outcome, it might nevertheless influence behavior. Specifically, the hypothetical outcomes as they would occur in the irrelevant context could favor the same side as the relevant one, or not (Congruent vs Incongruent trials, see Fig.1e left), and have larger or smaller values compared to the relevant features (Fig.1e right).

We investigated the impact of these factors on RTs in correct 2D trials, where the extensive training ensured near-ceiling performance throughout the main task ($\mu = 0.91, \sigma = 0.05$, t-test against chance: $t_{(34)} = 48.48$, $p < .0001$, Fig.2a). RTs were log transformed to approximate normality and analysed using mixed effects models with nuisance regressors for choice side

(left/right), time on task (trial number), differences between attentional contexts (color/motion) and number of trials since the last context switch (all nuisance regressors had a significant effect on RTs, Type II Wald χ^2 test, all ps< 0.03). We used hierarchical model comparison to assess the effects of (1) the objective value of the chosen option (or: EV), i.e. points associated with the features on the cued context; (2) the maximum points that could have been obtained if the irrelevant features were the relevant ones (the expected value of the background, henceforth:

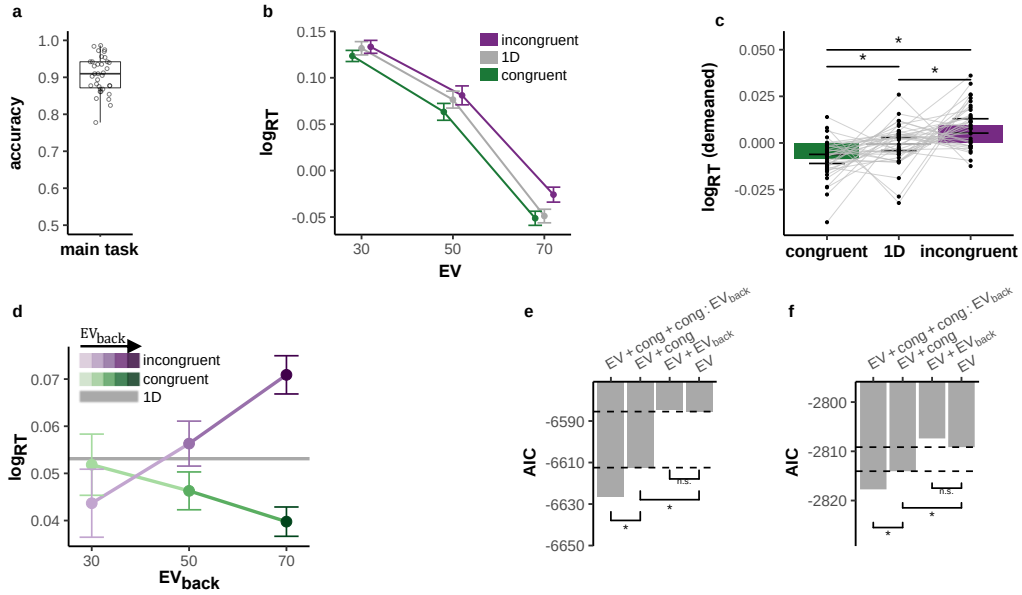


Figure 2

Behavioral results a. Participants performed near-ceiling throughout the main task, $\mu = 0.905, \sigma = 0.05$ (n=35). Box covers interquartile range (IQR), mid-line reflects mean, whiskers the range of the data (until +-1.5*IQR), and solid points represent outliers beyond whiskers. **b.** Participants reacted faster to higher Expected Values (EV, x-axis) and slower to incongruent (purple) compared to congruent (green) trials. RTs for 1D trials shown in gray. Error bars represent corrected within subject SEMs (Cousineau et al., 2005; Morey et al., 2008). **c.** Comparison of log RTs by trial condition. Incongruent trials were slower than 1D trials (paired t-test: p=.013), and 1D trials slower than congruent trials (paired t-test: p =.017; paired t-test congruent vs incongruent: p<.001). Error bars represent corrected within subject SEMs (Cousineau et al., 2005; Morey et al., 2008). P-values FDR-corrected, n=35. **d.** The Congruency effect was modulated by $\mathrm{EV_{back}}$, i.e. the more participants could expect to receive from the ignored context, the slower they were when the contexts disagreed and respectively faster when contexts agreed (x axis, shades of colours). Likelihood-ratio test (LRR) to asses improved model fit: p<.001, n=35. Gray horizontal line depicts the average RT for 1D trials across subjects and EV. Error bars as above. **e.** Hierarchical comparison of 2D trial log-RT models showed that inclusion of a Congruency main effect (p<.001, see **c**), yet not $\mathrm{EV_{back}}$ (p=.27), improved model fit. However, including an additional Congruency $\times$ $\mathrm{EV_{back}}$ interaction improved model fit even more (p<.001, see **d**). P-values from LR tests as above, stars indicate p<.05, n=35. **f.** We replicated the behavioral results in an independent sample of 21 participants outside the MRI scanner. Including Congruency (p =.009) but not $\mathrm{EV_{back}}$ (p=.63), improved model fit. Including an additional Congruency $\times$ $\mathrm{EV_{back}}$ interaction explained the data best (p =.017). P-values/stars as in **e**.

EV_{back}, Fig 1e right), and (3) whether the irrelevant features favored the same side as the relevant ones or not (Congruency, Fig.1e left). Any effect of the latter two factors would indicate that outcome associations that were irrelevant in the current context nevertheless influence behavior, and therefore could be represented in vmPFC.

We found that participants reacted faster in trials that yielded larger rewards and slower in incongruent compared to congruent trials (likelihood-ratio test to asses improved model fit, EV: $\chi^2_{(1)} = 1374.6, p < .001$, Congruency: $\chi^2_{(1)} = 29.0, p < .001$, Fig.2b-c). Moreover, compared to 1D trials, participants were slower to respond to incongruent trials and faster to respond to congruent trials (paired t-tests: $t_{(34)} = -2.79, p = .013$, $t_{(34)} = 2.5, p = .017$ respectively, FDR-corrected, see Fig.2b-c). Crucially, we found that Congruency interacted with the expected value of the other context: larger EV_{back} increased participants' speed on congruent trials and had the opposite effect on incongruent trials (LR-test: $\chi^2_{(1)} = 18.19, p < .001$, Fig.2d). These effects show that even when participants chose accurately based on the relevant context, the information of the irrelevant context was not completely filtered. The expected value of a 'counterfactual' choice resulting from consideration of the irrelevant context mattered: the outcome such a choice could have led to influenced reaction times. A full model description including effect sizes and confidence intervals can be found in SI table S2.

Neither adding a main effect for EV_{back} nor the interaction of EV $\times$ EV_{back} improved model fit (LR-tests: $\chi^2_{(1)} = 1.21, p = .27$, $\chi^2_{(1)} = .01, p = 0.9$ respectively), indicating that neither the presence of larger irrelevant values alone, nor their similarity to the relevant values influenced participants' RTs. Additionally, the lower valued irrelevant feature did not show comparable effects and did not interact with Congruency (LR-test to baseline model: $\chi^2_{(1)} = 0.92, p = .336$, with interaction: $\chi^2_{(1)} = 2.76, p = .251$). Replacing EV_{back} with a parameter of overall value of the irrelevant features did not improve the fit (which could be understood as an overall distraction of the irrelevant context, AIC of model with $EV_{back}\times$ Congruency: -6626.649, AIC of model with Overall Value $\times$ Congruency: -6619.878, Fig.S3). These results further support that it is specifically the expected reward of the ignored context that played a role in participants' RT.

All major RT effects hold when running the models nested within levels of EV, Block Context or switch (Fig.S2). Moreover, the number of trials since context switch did not interact with our main effect (LR-test with added term for Congruency$\times$EV$_{\text{back}}\times$ switch:$\chi^2_{(1)} = 3.70$, $p = .157$) and our main RT effects still hold when we excluded the first 2 trials after the context switch (LR-tests: Congruency, $\chi^2_{(1)} = 8.12$, $p = .004$, Congruency$\times$EV$_{\text{back}}$, $\chi^2_{(1)} = 16.61$, $p < .001$). We note that an interaction of EV $\times$ Congruency indicated stronger Congruency effect for higher EV (LR-test with added term: $\chi^2_{(1)} = 4.34$, $p = .037$, Fig.2b), but did not replicate in the replication sample (see below, $\chi^2_{(1)} = 0.23$, $p = .63$). Details of other significant effects and alternative models considering for instance within-cloud or between-context value differences can be found in Fig.S3 and Fig.S4 respectively.

We replicated these findings in an additional sample of 21 participants (15 women, $\mu_{age} = 27.1, \sigma_{age} = 4.91$) that were tested outside of the MRI scanner (LR-tests: Congruency, $\chi^2_{(1)} = 6.89$, $p = .009$, EV$_{\text{back}}$, $\chi^2_{(1)} = .23$, $p = .63$, Congruency$\times$EV$_{\text{back}}$, $\chi^2_{(1)} = 5.69$, $p = .017$, Fig.2e).

We next modeled choice accuracy in 2D trials using the same analysis approach and nuisance variables (see methods and Fig.S5) and found the same effects as the RT models: (1) Higher accuracy for higher EV (LR-test: $\chi^2_{(1)} = 14.61$, $p < .001$) (2) decreased performance on incongruent trials with (3) higher error rates occurring on trials with higher EV$_{\text{back}}$ (LR-tests: $\chi^2_{(1)} = 66.12$, $p < .001$, $\chi^2_{(1)} = 6.99$, $p = .03$, respectively, Fig.S5).

In summary, these results indicated that participants did not merely perform a value-based choice among features on the currently relevant context. Rather, both reaction times and accuracy indicated that participants also retrieved the values of irrelevant features and computed the resulting counterfactual choice. We next turned to test if the neural code of vmPFC would also incorporate such counterfactual choices, and if so, how the representation of the relevant and irrelevant contexts and their associated values might interact.

fMRI results

Outcome-relevant and outcome-irrelevant values co-exist within the vmPFC. We derived a value-sensitive vmPFC ROI following common procedures in the literature (e.g. Bartra et al., 2013; Clithero & Rangel, 2014) (see Fig.3a, and Methods) and tested whether both relevant and irrelevant expected values are reflected in multivariate vmPFC patterns using RSA. To estimate value-related activity patterns within the vmPFC mask, we fitted a General Linear Model (GLM) with one separate regressor for each combination of EV and $\mathrm{EV}_{\mathrm{back}}$, irrespective of the context (cross-validated, 1D trials modeled separately). After multivariate noise normalization and mean pattern subtraction (see Walther et al., 2016) we computed the Mahalanobis distance between each combination of regressor. This resulted in one 9×9 Representational Dissimilarity Matrix (RDM, Fig.3 and Methods) per subject, which we analyzed using mixed effects models (Gamma family with a inverse link, Magnusson et al., 2017). We first asked whether EV was reflected in the RDMs, as expected given that we used a functionally defined value ROI. Indeed, adding a main effect for EV dissimilarity (0 when two regressors share the same EV, 1 otherwise) improved model fit compared to a null model (LR-test: $\chi^2_{(1)} = 10.89$, $p < .001$, Fig.3b). Next, we asked if the activity patterns from trials with the same $\mathrm{EV}_{\mathrm{back}}$ were more similar than patterns reflecting different $\mathrm{EV}_{\mathrm{back}}$. Strikingly, adding a main effect of $\mathrm{EV}_{\mathrm{back}}$ dissimilarity (0 when sharing $\mathrm{EV}_{\mathrm{back}}$ and 1 otherwise) further improved model fit (LR-test with added term: $\chi^2_{(1)} = 247.67$, $p < .001$, Fig.3c.).

We then reasoned that the neural codes of expected values should also reflect value-differences in a gradual manner. We therefore asked whether pattern similarity was not only increased if two trials had the same value (e.g. comparing ’30 to ’30’, Fig.3d. purple cells), but also higher when the values in two trials had a difference of 20 (e.g. ’30’ to ’50’, Fig.3d. turquoise) compared to a value difference of 40 (e.g. ’30’ to ’70’, Fig.3d. yellow). Indeed we found that adding main effects for the value difference of EV as well as $\mathrm{EV}_{\mathrm{back}}$ improved model fit (VD_{EV}:LR-test compared to a null model: $\chi^2_{(1)} = 12.34$, $p < .001$, $\mathrm{VD}_{EV_{\mathrm{back}}}$: LR-test with added term: $\chi^2_{(1)} = 256.98$, $p < .001$, Fig.3c-d.). Note that the full model with both value difference effects

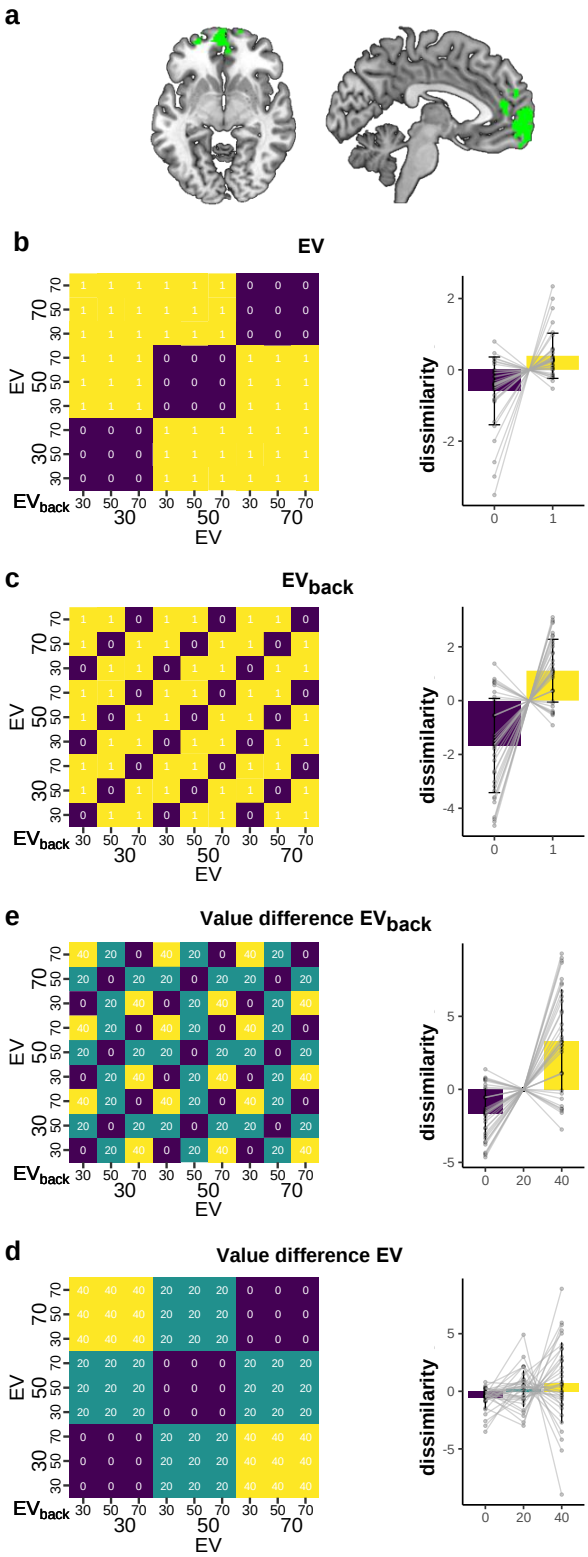


Figure 3

RSA analyses show that vmPFC encodes both relevant as well as irrelevant expected values given the current task context. a. vmPFC region used in all analyses (green voxels), defined functionally as the positive effect of a univariate value regressor thresholded at p_{FDR}<.0005 (one sided t-test, see methods). Note that no information regarding the contextually irrelevant values was used to construct the ROI. Axial slice (left) at x=-6; Sagittal slice (right) at z=-6. **b.** Left: Model RDM, each cell represents one combination of EV and $\mathrm{EV_{back}}$, see axes. Colors reflect whether a combination of trials had the same EV (purple) or not (yellow). Right: Dissimilarity of vmPFC activation patterns for trials with the same vs different EV. Dissimilarity was lower in trials that share the same expected value (EV, p<.001, n=35). **c.** Model RDM (left) testing whether irrelevant expected value ($\mathrm{EV_{back}}$) affected similarity in vmPFC. We found less dissimilarity for trails with the same $\mathrm{EV_{back}}$ (p<.001, n=35, right). **d.** Left: Model RDM that tested whether patterns similarity was influenced by the size of EV differences (0: purple, 20: turquoise, 40: yellow). Right: Average dissimilarity associated with the varying levels of value difference, indicating that larger EV differences between trials were related to higher pattern dissimilarity (p<.001, n=35). **e.** The same effect was found with respect to $\mathrm{EV_{back}}$ where patterns that share the same $\mathrm{EV_{back}}$ (irrespective to EV) also showed a decrease in dissimilarity (p <.001, n=35). Data shown in bar plots are demeaned by trial-frequency in the design to match the mixed effect models (see methods and Fig.S6). Error bars in panels b-e represent corrected within-subject SEMs (Cousineau et al., 2005; Morey et al., 2008). P-values in panels b-e reflect likelihood-ratio test of improved model fit, see main text.

resulted in a better (lower) AIC score than the model with both main effects of the EVs (AIC=165231 and AIC=165241, respectively, Fig.S6) indicating that the value similarity effect is not merely driven by the diagonal. Full models including effect sizes and confidence intervals can

be found in SI Table S5 and Table S6.

Hence, neural patterns in vmPFC were affected by contextually-relevant as well as irrelevant value expectations. Notably, the values of irrelevant features were computed despite being counterfactual (not related to the choice), and co-existed with well known expected values signals in vmPFC.

vmPFC value and context signals co-exist and are positively related. We next turned to investigate how the neural value representations of EV, $\mathrm{EV}_{\mathrm{back}}$ and context interacted with each other on a trial-wise level. We therefore trained a multivariate multinomial logistic regression classifier on the fMRI images acquired approx. 5 seconds after stimulus onset in same vmPFC ROI used above. An expected value classifier was trained on behaviorally accurate 1D trials, where no irrelevant values were present (henceforth: Value classifier, Fig.4a, left; leave-one-run-out training; see methods). For each testing example, the classifier assigned the probability of each class given the data (classes are the expected outcomes, i.e. '30','50' and '70', and probabilities sum up to 1, Fig.4a, right). Crucially, it had no information about the task context of each given trial (training sets were up-sampled to balance w.r.t. color/motion contexts, see methods). We first validated that the classifier was sensitive to values, as expected given the nature of the ROI. Indeed, the class with the maximum probability corresponded to the objective outcome significantly more often than chance, both when when tested on held out 1D and 2D trials as well as when tested only on 2D trials ($\mu_{all} = .35, \sigma_{all} = .029$, $t_{(34)} = 2.89$, $p = .003$, $\mu_{2D} = .35, \sigma_{2D} = .033$, $t_{(34)} = 2.20$,$p = .017$, respectively. Fig.4b). Similar to the RSA analysis, we reasoned that the similarity between the values assigned to the classes will be reflected in gradual probability differences . Specifically, we expected not only that the probability associated with the correct class be highest (e.g. '70'), but also that the probability associated with the closest class (e.g. '50') would be higher than the probability with the least similar class (e.g. '30', Fig.4c). Indeed we found that similar values elicited similar probabilities (LR-test of linear relation between value difference and class probability: $\chi^2_{(1)} = 12.74$, $p < .001$, full analysis can be found in Fig.S7) Additional control analyses indicated that our value classification results were

not the result of a bias caused by overlap of perceptual features between training and test (Fig.S8).

A major feature of our task was that which value expectation was relevant depended on the task context. We therefore hypothesized that vmPFC would also encode the task context, although this is not directly value-related (the average values of both contexts were identical). We thus trained a second classifier on the same data from the EV-sensitive ROI on the same accurate 1D trials, but this time to identify if the trial was 'Color' or 'Motion' (Fig.4d, left). The classifier had no information as to what was the EV of each given trial, and training sets were up-sampled to balance the EVs within each set (see methods). The classifier performed above chance for decoding the correct context, again both when tested on held out trials from all conditions as well as when tested only on 2D trials (t-test against chance: $t_{(34)} = 3.93$, $p < .001$, $t_{(34)} = 3.2$, $p = .001$, respectively. Fig.4e). Moreover, the context was still decodable when keeping the perceptual input identical between the two classes (i.e. testing on 2D trials with fixed value difference of the irrelevant values of 20, since the value difference of the relevant context was always 20, $t_{(34)} = 2.73$, $p = .0005$).

We first hypothesized that if vmPFC is involved in signaling both context and values, then the strength of context signal might relate to the strength of the contextually relevant value. A corresponding mixed effects analysis indeed found that the probability the context classifier assigned to the correct class (henceforth: $\mathrm{P}_{context}$) had a positive effect on the decodability of EV (henceforth: P_{ev}, LR-test compared to null model: $\chi^2_{(1)} = 9.12$, $p = 0.002$, Fig.4f). In other words, the better we could decode the context, the higher was the probability assigned to the correct EV class.

In summary, we found that the Context is represented within the same region as the EV, and that the strength of its representation is directly linked to the representation of EV. The link between Context and relevant EV signals suggest that the Context signal might play a role in governing which values dominate vmPFC.

Competition of vmPFC EV and $\mathrm{EV}_{\mathrm{back}}$ signals is moderated by a context representation.

One main hypothesis was that contextually-irrelevant values might influence neural codes of

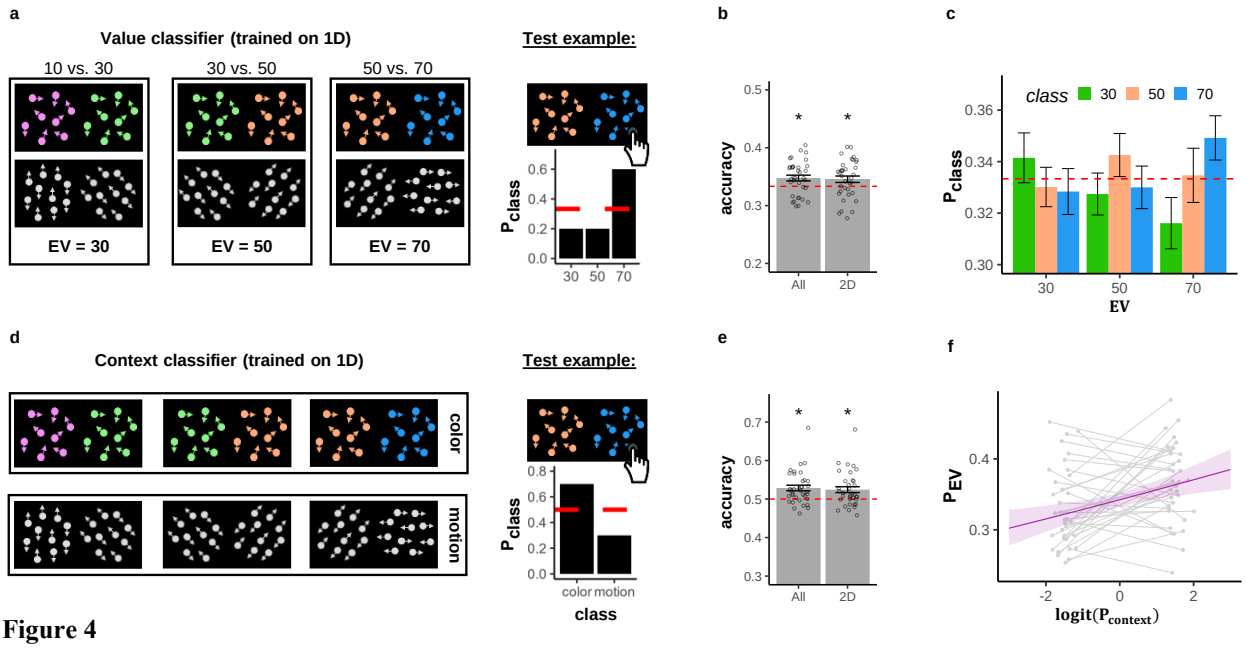


Figure 4

Expected value and context signals co-reside within vmPFC. a. A logistic classifier was trained on behaviorally accurate 1D trials to predict the true EV from vmPFC patterns ("Value classifier", left). We analysed classifier correctness and predicted probability distribution (right). shown in **b** and **c**. **b.** The Value classifier assigned the highest probability to the correct class (objective EV) significantly more often than chance for all trials (p =.003, n=35), also when tested on generalizing to 2D trials alone (p=.017, n=35). **c.** The probabilities the classifier assigned to each class (y-axis, colors indicate the different classes, see legend) split by the objective EV of the trials (x axis). As can be seen, the highest probability was assigned to the class corresponding to the objective EV of the trial (i.e. when the color label matched the X axis label). n=35, for individual data points see Fig. S7. **d.** A second logistic classifier was trained on the same data to distinguish between task contexts (color vs motion), irrespective of the EV ("Context" classifier). **e.** The Context classifier assigned the highest probability to the correct class (objective Context) significantly more often than chance for all trials (p<.001, n=35), also when tested on generalizing to 2D trials alone (p = .001, n=35). **f.** Increased evidence for the objective EV (P_{EV}, y-axis) was associated with stronger context signal in the same ROI (x-axis, where probabilities z-scored and logit-transformed, LR-test compared to null model: p=.002, N=35). Plotted are model predictions and gray lines represent individual participants (mean of the top/bottom 20% of trials). Error bands represent the 89% confidence interval. P values in panels b and e reflect one sided t-test against chance. Error bars in panels b, c and e represent corrected within-subject SEMs (Cousineau et al., 2005; Morey et al., 2008).

expected value in vmPFC, and therefore should interact with EV probabilities decoded from vmPFC in a trial-wise manner. Similar to our analyses above, we used mixed effects models to test whether the Value classifier's probability of the correct class (P_{EV}) was influenced by $\mathrm{EV}_{\mathrm{back}}$ and/ or Congruency of a given 2D trial. This analysis revealed that $\mathrm{EV}_{\mathrm{back}}$ had a negative effect on P_{EV} (LR-test compared to null model $\chi^2_{(1)} = 5.96$, $p = .015$, Fig.5b), meaning that larger irrelevant expected value led to weaker representation of the relevant one (measured by lower probability of the objective EV, P_{EV}). Importantly, this effect cannot be attributed to attentional

effects caused by perceptual input, since replacing $\mathrm{EV}_{\mathrm{back}}$ with a regressor indicating the presence of its corresponding perceptual feature in the training class, as highest or lowest value, did not provide a better model fit (AICs: -1229.2,-1223.3, respectively, see Fig.S8 for details). Adding the minimum value of the irrelevant context of the trial also did not improve the fit, indicating that it is specifically the highest of the two irrelevant features driving this effect (LR-test with added term: $\chi^2_{(1)} = 0.63$, $p = .43$). We found no evidence for a $\mathrm{EV}_{\mathrm{back}} \times \mathrm{P}_{context}$ interaction (LR-test with added term: $\chi^2_{(1)} = 0.012$, $p = .91$). Our RSA analysis also provided further support for this effect, where we found that $\mathrm{EV}_{\mathrm{back}}$ also had a negative effect on the EV similarity, i.e. higher dissimilarity for higher $\mathrm{EV}_{\mathrm{back}}$ (Type II Wald χ^2 test: $\chi^2_{(1)} = 36.6$, $p < .001$, see Fig.S6). Similarly, high $\mathrm{EV}_{\mathrm{back}}$ also disrupted the similarities between of the probabilities of the value classifier (LR-test: $\chi^2_{(1)} = 6.16$, $p = .013$, see Fig.S7). A number of control analyses also indicated the validity of finding: Interestingly, and unlike in the behavioral models, we found that neither Congruency nor its interaction with EV or $\mathrm{EV}_{\mathrm{back}}$ influenced P_{EV} ($\chi^2_{(1)} = 0.035$, $p = .852$,$\chi^2_{(1)} = 0.48$, $p = .787$, $\chi^2_{(1)} = .99$, $p = .317$, respectively, Fig.5c), and a match of value expectations of both contexts (i.e. EV=$\mathrm{EV}_{\mathrm{back}}$) led no change of P_{EV} ($\chi^2_{(1)} = 0.45$, $p = .502$, see methods). We also found no effect of time since switch on the decodability of EV (Type II Wald χ^2 test: $\chi^2_{(1)} = 0.85$, $p = .36$, Fig S9, but see discussion on limitations). Alternative models of P_{EV}, e.g. including within-option or between-context value differences, or alternatives for $\mathrm{EV}_{\mathrm{back}}$ (Fig.S9).

The decrease in value decodability due to high irrelevant value expectations could reflect a general disturbance of the value retrieval process caused by the distraction of competing values. Alternatively, the encoding of $\mathrm{EV}_{\mathrm{back}}$ could directly compete with the representation of EV – reflecting that the relevant and irrelevant value expectations might be represented using similar neural codes (note that the classifier was trained in the absence of task-irrelevant values, i.e. the objective EV of 1D trials). In order to test this idea, we looked at the Value classifier probabilities in trials where EV $\neq$ $\mathrm{EV}_{\mathrm{back}}$. This allowed us to interpret the class probabilities of our Value classifier as either signifying EV (P_{EV}), $\mathrm{EV}_{\mathrm{back}}$ ($\mathrm{P}_{\mathrm{EV}_{\mathrm{back}}}$) or a value that was expected in neither

case ($\mathrm{P}_{\mathrm{other}}$, Fig.5d). We then examined the correlation between each pair of classes. To prevent any disadvantage of the 'other' class, we included only trials in which the the 'other' value's associated feature appeared on the screen (relevant or irrelevant). Note that the three class probabilities for each trial sum up to 1 and hence are strongly biased to correlate negatively. Yet, P_{EV} and $\mathrm{P}_{\mathrm{EV}_{\mathrm{back}}}$ had a significantly more negative correlation than P_{EV} and $\mathrm{P}_{\mathrm{other}}$ ($\rho = -.56$, $\sigma = .22$, $\rho = -.40$, $\sigma = .25$ respectively, paired t-test: $t_{(34)} = -2.77$, $p = .017$, Fig.5e). This shows that when the probability assigned to the EV decreased, it was accompanied by a stronger increase in the probability assigned to $\mathrm{EV}_{\mathrm{back}}$, akin to a competition between both types of expectations. Formally, we show that adding $\mathrm{P}_{\mathrm{EV}_{\mathrm{back}}}$ to the model predicting P_{EV} results in a smaller AIC than when adding $\mathrm{P}_{\mathrm{other}}$ (-574 vs -473, respectively. Fig.5f), likelihood-ratio-test for a model with $\mathrm{P}_{\mathrm{EV}_{\mathrm{back}}}$: $\chi^2_{(1)} = 144.34$, $p < .001$, and with $\mathrm{P}_{\mathrm{other}}$: $\chi^2_{(1)} = 43.83$, $p < .001$).

The previous analysis only informs us about the overall correlation of probabilities across the entire experiment. To investigate the trial-wise dynamics of the neural representation within vmPFC, we trained an additional classifier to detect the $\mathrm{EV}_{\mathrm{back}}$ on behaviorally accurate 2D trials. Although this classifier suffers from some caveats (see Methods, Fig.S6a-c and below for details), we reasoned that trialwise probability fluctuations are unbiased, and proceeded to ask if the probability the $\mathrm{EV}_{\mathrm{back}}$ classifier assigned to the correct class ($\mathrm{P}^{2D}_{\mathrm{EV}_{\mathrm{back}}}$) might relate to encoding of the relevant value as indicated by the Value classifier (i.e., P_{EV}). Importantly, both classifiers were trained on independent data ($\mathrm{EV}_{\mathrm{back}}$ classifier on 2D, and Value classifier on 1D trials), but in both cases on behaviorally accurate trials, i.e. trials where participants choose according to EV, as indicated by the relevant context. This model showed that an increase in neural representation of $\mathrm{EV}_{\mathrm{back}}$, when measured independently ($\mathrm{P}^{2D}_{\mathrm{EV}_{\mathrm{back}}}$), reduced EV decodability on a trial-wise basis (lowered AIC score from -1223.6 to -1225.0, but note that in the LR-test $\chi^2_{(1)} = 3.45$, $p = 0.063$, Fig.5d). Most remarkably, the effect of Context, $\mathrm{P}_{context}$, interacted with the effect of $\mathrm{P}^{2D}_{\mathrm{EV}_{\mathrm{back}}}$, such that when the context signal was stronger, the negative effect of irrelevant value signals on relevant value signals was weaker (i.e. $\mathrm{P}_{context}$ affected the association between $\mathrm{P}^{2D}_{\mathrm{EV}_{\mathrm{back}}}$ and P_{EV}, LR-test: $\chi^2_{(1)} = 5.22$, $p = 0.022$, Fig.5e). In other words, the stronger the relationship between

Context and EV representations, the less vmPFCs irrelevant value signal competed with its value representations, akin to a shielding effect. The same analysis also confirmed our previous finding that the strength of context encoding affected value encoding (effect of P_{context}, LR-test: $\chi^2_{(1)} = 9.99, p = .002$). Note that the above analysis was complicated by the frequency differences between different EV_{back} classes, which we controlled by running the model of P_{EV} with random effects nested within levels of EV_{back} for each subject, i.e. any effect found is not influenced by the (biased) mean difference between the probabilities assigned to each of those levels (intuitively, this is similar to running each correlation separately within each level of EV_{back}). Full models including effect sizes and confidence intervals can be found in tables S3 and S4.

In summary, we showed the neural representation of EV was reduced in trials with higher expected value of the irrelevant context, and weakened EV representations were accompanied by an increase in neural representations of such irrelevant value expectation, in the same vmPFC region. The effect occurred irrespective of action-conflict between the relevant and irrelevant values (unlike participants' behaviour). Most strikingly, the negative influence of EV_{back} representation on EV decodability was mediated by a neural context signal, i.e. when the link between Context and EV increased, the effect of EV_{back} representations diminished. As will be discussed later in detail, we consider this to be evidence for parallel processing of two task aspects in this region, EV and EV_{back}.

Neural representation of EV, EV_{back} and Context guide choice behavior. Finally, we investigated how vmPFC's representations of EV, EV_{back} and Context influence participants' behavior. We first investigated this influence on choice accuracy. Note that the two contexts only indicate different choices in incongruent trials, where a wrong choice could be a result of a strong influence of the irrelevant context. Motivated by our behavioral analyses that indicated an influence of the irrelevant context on accuracy, we asked whether $P_{EV_{\text{back}}}$ was different on behaviorally wrong or incongruent trials. We found an interaction of accuracy $\times$ Congruency (Type II Wald χ^2 test: $\chi^2_{(1)} = 4.51, p = .034$, Fig.6a) that indicated increases in $P_{EV_{\text{back}}}$ in accurate congruent trials and decreases in wrong incongruent trials. Hence, on trials in which

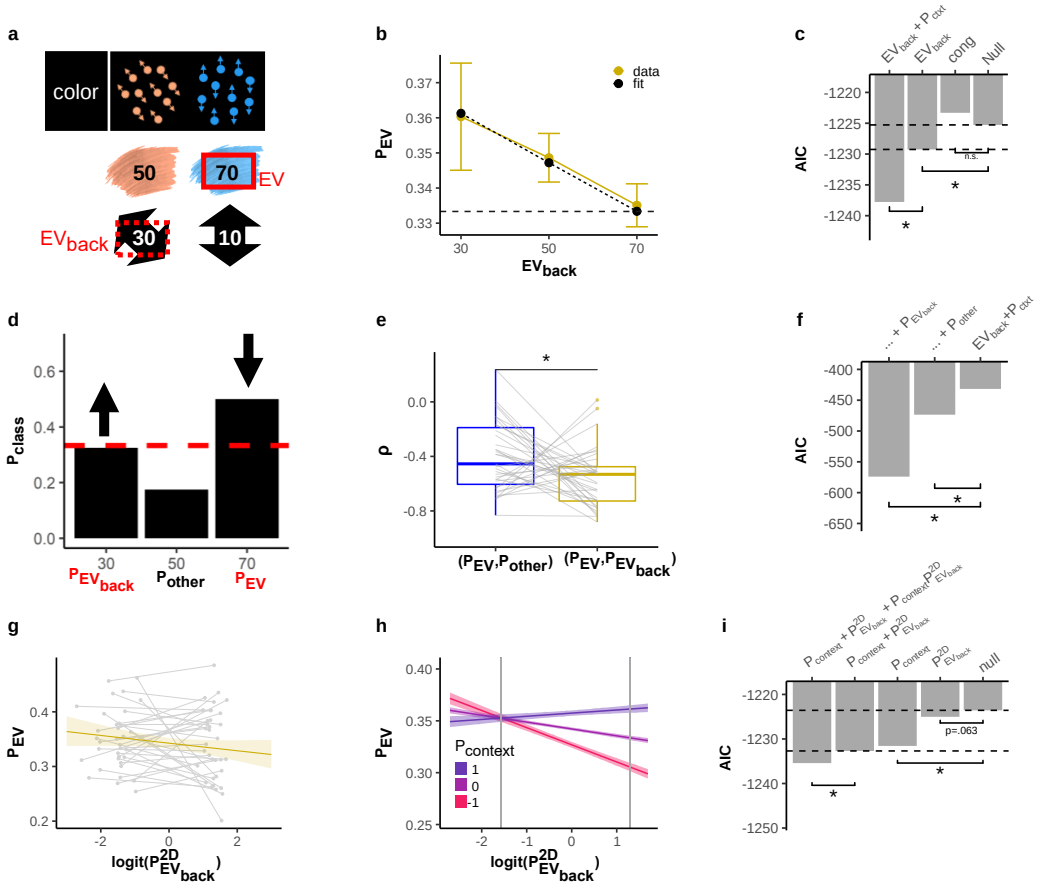


Figure 5

vmPFC representations of task context, expected value and irrelevant expected value interact a. Exemplary 2D color trial (top), its relevant outcomes (middle, color-based), and hypothetical/irrelevant outcomes (motion-based, bottom). The maxima of relevant and irrelevant outcomes are termed EV and $\mathrm{EV_{back}}$, respectively. **b.** Higher $\mathrm{EV_{back}}$ was related to decreased decodability of EV (P_{EV}) in behaviorally accurate trials, likelihood-ratio (LR)test: p=.015, n=35. Color see legend. Error bars represent corrected within-subject SEM. For supporting RSA evidence see Fig.S6d. (Cousineau et al., 2005; Morey et al., 2008). **c.** Modelling the probabilities assigned to the true EV class (P_{EV}) showed an effect of $\mathrm{EV_{back}}$ (p=.015) but not Congruency (p=.852). Including $\mathrm{EV_{back}}$ and context decodability (P_{context}) yielded the best fit (p=.001). P-values reflect LR-tests. **d.** Illustration that Value classifier class probabilities in panel a's example could reflect the true EV ($\mathrm{P_{EV}}$), the $\mathrm{EV_{back}}$ ($\mathrm{P_{EV_{back}}}$) or neither EV or $\mathrm{EV_{back}}$ ($\mathrm{P_{Other}}$). **e.** The correlation between $\mathrm{P_{EV}}$ and $\mathrm{P_{EV_{back}}}$ (yellow) was significantly more negative than the correlation between $\mathrm{P_{EV}}$ and $\mathrm{P_{Other}}$ (blue, paired t-test, p=.017, n=35). Box covers interquartile range (IQR), mid-line reflects median, whiskers the range of the data (until +-1.5*IQR), and solid points represent outliers beyond whiskers. **f.** Comparing models of $\mathrm{P_{EV}}$ confirmed that adding $\mathrm{P_{EV_{back}}}$ improved fit more than adding $\mathrm{P_{Other}}$ (AIC: -574 vs. -473), LR test with each individual effect: p<.001. n=35. **g.** The neural representations of relevant EV ($\mathrm{P_{EV}}$, y-axis) and the irrelevant EV ($\mathrm{P^{2D}_{EV_{back}}}$, x-axis, z-scored and multinomial-logit-transformed) were marginally negatively associated (LR-test: p=.063, n=35). Error bands represent 89% confidence interval and gray lines individual participants' top/bottom 20%. **h.** Increased evidence for a Context representation ($\mathrm{P}_{context}$) correlated with less EV/$\mathrm{EV_{back}}$ competition (i.e. weaker effect of $\mathrm{P^{2D}_{EV_{back}}}$ on $\mathrm{P_{EV}}$ when $\mathrm{P}_{context}$ was stronger, LR-test with interaction term: p=.022). Lines reflect model predictions, error bands represent 89% CI and vertical lines show group means of the top/bottom 20% of data (averaged first within participant, for individual lines, see Fig.S10). NB that $\mathrm{P}_{context}$ was split into three levels for visualization; in our model it was continuous. **i.** Comparing models of $\mathrm{P_{EV}}$ (nested within $\mathrm{EV_{back}}$ levels) revealed that adding either $\mathrm{P^{2D}_{EV_{back}}}$ or $\mathrm{P}_{context}$ improved model fit (panels g and h, p=.063 and p=.022), as well as their interaction $\mathrm{P}_{context} \times \mathrm{P^{2D}_{EV_{back}}}$ (LR-test with interaction compared to only $\mathrm{P}_{context}$: p=.022, and only $\mathrm{P^{2D}_{EV_{back}}}$: p=.029, n=35). Note that $\mathrm{P_{EV_{back}}}$ (panels a-f) indicates the Value classifier' class probabilities of the the $\mathrm{EV_{back}}$ class, whereas $\mathrm{P^{2D}_{EV_{back}}}$ (panels g-i) indicates the $\mathrm{EV_{back}}$ classifier's $\mathrm{EV_{back}}$ class probabilities (the former was trained on 1D, the latter on 2D trials). Stars in **c**, **e**, **f**, **i** represent threshold of p<.05.

participants erroneously chose the option with higher valued irrelevant features, $P_{EV_{back}}$ was increased. Focusing only on behaviorally accurate trials, we found no effect of EV or Congruency on $P_{EV_{back}}$ (Type II Wald χ^2 tests: $\chi^2_{(1)} = 0.07$, $p = .794$, $\chi^2_{(1)} = 0.00$, $p = .987$ respectively). This effect is preserved when modeling only wrong trials (Type II Wald χ^2 test of Congruency: $\chi^2_{(1)} = 4.36$, $p = .037$).

Motivated by the different predictions for congruent and incongruent trials, we next turned to model these trial-types separately. When focusing on incongruent trials we found that a weaker representation of the relevant context was marginally associated with an increased error rate (negative effect of $P_{context}$) on accuracy, indicating an increased representation of the wrong context, LR-test: $\chi^2_{(1)} = 3.66$, $p = .055$, Fig.6b). Moreover, we found that the joint increases of the wrong context and its associated irrelevant expected value representation (EV_{back}) strengthend this effect, i.e. adding a $P_{context} \times P_{EV_{back}}$ term to the model of error rates improved model fit (LR-test: $\chi^2_{(1)} = 6.33$, $p = .012$, Fig.6b; NB that we found no main effects of EV or EV_{back} LR-tests: $\chi^2_{(1)} = 0.28$, $p = .599$, $\chi^2_{(1)} = 0.0$, $p = .957$, respectively). We next turned to congruent trials, where a wrong choice should not be associated with activation of the wrong context since both contexts indicate the same choice. Indeed, there was no influence of $P_{context}$ on accuracy in Congruent trials (LR-test: $\chi^2_{(1)} = 0.0$, $p = .922$). However, strong representation of either relevant or irrelevant EV should lead to a correct choice. Indeed, we found that both an increase in $P_{EV_{back}}$ and (marginally) in P_{EV} had a positive relation to behavioral accuracy ($\chi^2_{(1)} = 3.5$, $p = .061$, $\chi^2_{(1)} = 6.48$, $p = .011$, respectively, Fig.6c).

Finally, we investigated reaction times of behaviorally accurate trials. In line with the results presented above, we found that participants who had a weaker influence of Context activity on their EV representation, also had a stronger RT Congruency effect ($r = -.39$,$p = .022$ Fig 6d). Next, we hypothesized that increased conflict between EV and EV_{back} representations of should influence RT. Indeed, all neural signatures of EV/EV_{back} conflict correlated with the Congruency-related RT effect: the more negative a participant's correlation between P_{EV} and $P_{EV_{back}}$ was, the stronger her RT Congruency effect ($r = -.45$,$p = .008$, Fig.6e); a more negative

association between EV_{back} and P_{EV} was linked to a stronger EV_{back} modulation of the RT Congruency effect ($r = .43$,$p = .01$, Fig.6f); finally, the same was true when considering the strength of the effect of the neural representation of EV_{back} ($P^{2D}_{EV_{back}}$) on the neural EV signal in relation to the above behavioral marker ($r = 35$,$p = .004$, Fig.6g). In other words, the negative influence of irrelevant EV and its neural representation on relevant EV signal, related to the interactive effect of $EV_{back} \times$ Congruency on RTs (i.e. slower RT for incongruent and faster for congruent trials).

In sum, choice accuracy was negatively related to the the representation of irrelevant contexts and its associated value only in incongruent trials (i.e. when it mattered), while in congruent trials neural representations of EV and EV_{back} contributed to accuracy. RT analyses showed that markers of (a) weaker representational link between context and EV and (b) stronger conflict between EV_{back} and EV were both associated with a stronger influence of the counterfactual choice on their RT. Brought together these findings show that the representations of EV, EV_{back} and Context in vmPFC don't only interact with each other, but guide choice behavior as reflected in accuracy as well as RT in behaviorally accurate trials.

No univariate evidence for effects of irrelevant values on expected value signals in vmPFC. The above analyses indicated that multiple value expectations are represented in parallel within vmPFC. Lastly, we asked whether whole-brain univariate analyses could also uncover evidence for processing of multiple value representations. Detailed description of the univariate analysis can be found in Fig.S12. Unlike the multivariate analysis, this revealed no positive modulation of Congruency, EV_{back} or their interaction was observed in any frontal region. A negative effect of was found EV_{back} in the Superior Temporal Gyrus, $p < .001$, Fig.S12c). We also found no region for the univariate effect of Congruency $\times$ EV_{2D} interaction (even at $p < .005$). However, we found a negative univariate effect of Congruency $\times$ EV_{back} in the primary motor cortex at a liberal threshold, which indicated that the difference between Incongruent and Congruent trials increased with higher EV_{back}, akin to a response conflict ($p < .005$, Fig.S12d). These findings contrast with the idea that competing values would have been integrated into a single EV

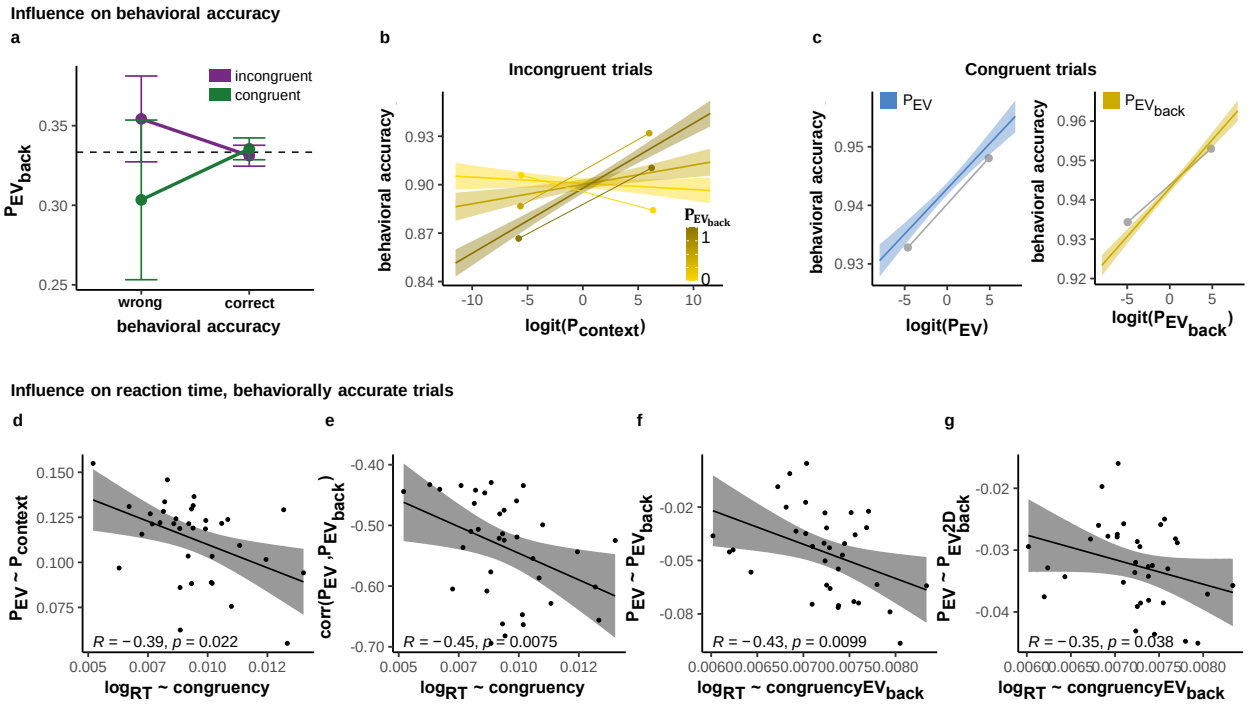


Figure 6

vmPFC representations of context and value jointly guide behavior Panels a-c include all trials whereas panels d-g show only behaviorally accurate trials. **a.** The Value classifier's probability of EV_{back} ($P_{EV_{\text{back}}}$, y-axis) was increased when participants chose the option based on EV_{back}, corresponding to a wrong choice in incongruent trials (purple) and correct choice in congruent trials (green), LR-test vs. null model: p=.034, n=35. Error bars represent corrected within subject SEMs (Cousineau et al., 2005; Morey et al., 2008). **b.** Decrease in behavioral accuracy (y-axis) in incongruent trials was marginally associated with lower context decodability (Context classifier, x-axis, p=.051). This effect was modulated by EV_{back} representation, i.e. stronger in trials with higher $P_{EV_{\text{back}}}$ in vmPFC (shades of gold, p=.012, discretisation only for visualization). P-values represent LR-test with added terms and error bands represent the 89% CI. **c.** Value classifier decodability of EV (blue, left) and EV_{back} (gold, right) were both positively related to behavioral accuracy in congruent trials (ps: .058 and .009, respectively, y axis). Lines are fitted slopes. Grey dots are group means of top and bottom 20% of data (within participant, for individual lines, see Fig.S11). P-values represent LR-test with added terms and error bands represent the 89% confidence interval. **d.** Participants with weaker associations between Context and EV representations (y-axis, Fig.5f), had a stronger Congruency RT effect (x-axis, larger values indicate stronger RT difference between incongruent and congruent trials, i.e. distance between purple and green lines in Fig.2b.) **e.** More negative correlations between EV and EV_{back} representations (y-axis, Fig.4b) were associated with stronger Congruency RT effects (x-axis, see panel d.). **f.** Participants with a stronger (negative) link between P_{EV} and EV_{back} (y-axis, see Fig.5e.) also had a stronger EV_{back} modulation on the Congruency RT effect (x-axis, see distance between purple and green lines in Fig.2d). **g.** Participants with a more negative link between P_{EV} and $P^{2D}_{EV_{\text{back}}}$ (y-axis, more negative indicate stronger decrease, see Fig.5g), had a stronger modulation of EV_{back} on Congruency RT effect (x-axis, see panel f.). Panels d-g present Pearson correlations, p-values represent Spearman's p statistic to estimate a rank-based measure of association (Best & Roberts, 1975; Hollander et al., 2013) and error bands represent 95% confidence interval.

representation in vmPFC, because this account would have predicted a higher signal for Congruent compared to incongruent trials.

Discussion

We investigated how contextually-irrelevant value expectations influence behavior and neural activation patterns in vmPFC. Participants reacted slower when the irrelevant context favored a different choice and faster when it favored the same. This Congruency effect increased with increasing reward associated with the hypothetical choice in the irrelevant context (EV_{back}). fMRI analyses of vmPFC voxels sensitive to the objective, i.e. relevant, expected value (EV) showed that (a) vmPFC contains a multifaceted representation of each trials expected value, irrelevant expected value and context; and that (b) higher irrelevant expected values, or a stronger neural representation of them, impaired the expected value signal, akin to a representational conflict between the two values. This conflict was moderated by the strength of the context signal, such that a stronger context signal was associated with a stronger expected value signal, and a diminished negative effect of the expected value of the irrelevant context. The different facets of vmPFC's representations were linked to participants' behavior in a manner generally consistent with the idea that the representations of the alternative/irrelevant context and its associated value were present within vmPFC and guided behavior. The strength of these representations within vmPFC was related to slower and less accurate choices when the different contexts implied different actions, and faster and more accurate choices when they agreed on the action to be made.

One notable aspect of our experiment was that feature relevance was cued on each trial, and rewards were never influenced by irrelevant features. Nevertheless, participants' behavior was influenced by the expected outcome of the counterfactual choice. This supports the notion that cognitive control based arbitration between relevant and irrelevant features is incomplete (Frömer & Shenhav, 2021; Mante et al., 2013; Takagi et al., 2020). Our neural analyses showed how internal value expectation(s) within vmPFC were shaped by such incomplete suppression: not the ignored context per se influenced vmPFC signals, but rather the computed expected value of the counterfactual choice that would have been made in that context. This was evidenced by the fact that the expected value of the background captured fluctuations in value representations. A control analysis showed that this cannot be explained by the presence of its corresponding

perceptual-feature on the screen. Hence, our results cannot be explained by value-independent attention capture caused by the 'distracting' irrelevant context (Fig.S8), and go beyond previous research on cognitive control, such as the Stroop Task (MacLeod, 1991).

We also asked whether relevant and irrelevant expected values integrate into a single EV, but found neither univariate nor multivariate evidence for this possibility. Specifically, we found no univariate EV_{back} or congruency effects, and no increase in EV decodability when EV equalled EV_{back}. This suggests some differences in the underlying representations of relevant and irrelevant expected values. At the same time, our analysis showed that the value classifier was sensitive to the expected value of the irrelevant context in 2D trials, even though it was trained on 1D trials during which irrelevant values were not present. This suggests that within vmPFC 'conventional' expected values and counterfactual values are encoded using partially, but not completely, similar patterns. Moreover, our results suggest that the EV of each context were activated simultaneously and competed with each other, a competition governed by the context signal. While neural evidence for EV competition did link behavioral evidence of choice conflict, we found no influence of action-congruency on vmPFC signal itself. This suggests that the conflicts between incongruent motor commands might be resolved elsewhere. Univariate analyses revealed that primary motor cortex was sensitive to Congruency, and hence might be the site of conflict resolution, in line with studies that suggest distracting information can be found in task execution cortex in humans and monkeys (Mante et al., 2013; Takagi et al., 2020). The idea that the conflict between multiple values encoded in vmPFC is resolved in motor cortex and is also in line with our interpretation that vmPFC does not integrate both tasks into a single EV representation that drives choice.

Participants repeatedly had to switch between contexts in our task, a process that is well known to engage cognitive control mechanisms (Frömer et al., 2021; Frömer & Shenhav, 2021; MacLeod, 1991; Monsell, 2003; Shenhav et al., 2018). We evaluated to what extent this task switching affected our results and found that behavioral effects hold when excluding the first 2 trials after a context switch, and that the distance from the last switch did not interact with the influence of the

irrelevant values (Fig.S2). Likewise, we found no influence of task switching on multivariate EV effects in vmPFC. Note, however, that due to our design we could not create balanced training sets (with respect to number of trials since context switch) which would be required for a more thorough investigation of the effect of trials since switch on value signals. We therefore conclude that while context switching is part of the investigated phenomenon, its presence alone cannot explain our findings.

Another important implication of our study concerns the nature of neural representations in vmPFC/mOFC, and in particular the relationship between state (Schuck et al., 2016, 2018; Wilson et al., 2014; Wimmer & Büchel, 2019) and value (Bartra et al., 2013; Clithero & Rangel, 2014; O'Doherty et al., 2001; Padoa-Schioppa & Assad, 2006; Plassmann et al., 2007; Rich & Wallis, 2016) codes in this area. In order to compare both aspects, we used a categorical classifier for value as well as states, rather than examining continuous value representations. Nevertheless, we believe that the value similarity analysis (both in the RSA, Fig.3d-e. and classifier probabilities, Fig.S7) additionally shows evidence for such continuous value representations. We specifically chose to focus on the vmPFC region that is commonly investigated in value-based decision research. We therefore defined our ROI in a univariate manner as commonly done in the literature (e.g. Bartra et al., 2013; Clithero & Rangel, 2014)) and studied the multivariate state and value signal within this ROI (e.g. Schuck et al., 2016, 2018) . We found that in addition to (expected) value information, vmPFC/mOFC also represented the context or task-state, which identified relevant information and thereby disambiguated the partially observable sensory state (e.g. Schuck et al., 2016, 2018; Wilson et al., 2014). Note that in our case the task context was agnostic to value (which was balanced across contexts) and specific features, but rather consisted of a superset of the more specific motion direction/color features. Any area sensitive to these more specific states would therefore also show decoding of context as defined here. Another methodological aspect was that we decoded based on timeshifted TR images, rather than deconvolved activity patterns (Mumford et al., 2012) as is common practice in fMRI decoding papers (McNamee et al., 2013; Momennejad et al., 2018; Polyn et al., 2005; Wittkuhn & Schuck,

2021). Decoding level and approach may have implications for the representations that can be uncovered in future research. Overall, our findings are in line with work that has found that EV could be one additional aspect of OFC activity (Zhou et al., 2019), which is multiplexed with other task-related information. Crucially, the idea that state representations integrate different kinds of task-relevant information (Niv, 2019; Schoenbaum & Roesch, 2005) could explain why this region was found to be crucial for integrating valued features when all features of an object are relevant for choice (Pelletier & Fellows, 2019; Schoenbaum & Roesch, 2005), although some work suggests that it might also reflect integration of features not carrying any value (Mack et al., 2020).

To conclude, the main contribution of our study is that we elucidated the relation between task-context and value representations within vmPFC. By introducing multiple possible values of the same option in different contexts, we were able to reveal a complex representation of task structure in vmPFC, with both task-contexts and their associated expected values activated in parallel. The decodability of both contexts and EVs independently from vmPFC, and their relation to choice behavior, hints at integrated computation of these in this region. We believe that this bridges between findings of EV representation in this region to the functional role of this region as representing task-states, whereby relevant and counterfactual values can be considered as part of a more encompassing state representation.

Methods

The study complies with all relevant ethical regulations and was approved by the ethics board of the Free University Berlin (Ref. Number: 218/2018).

Participants

Forty right-handed young adults took part in the experiment (18 women, $\mu_{age} = 27.6, \sigma_{age} = 3.35$) in exchange for monetary reimbursement. Participants were recruited using the participant database of Max-Planck-Institute for Human Development. Beyond common MRI-safety related exclusion criteria (e.g. piercings, pregnancy, large or circular tattoos etc.), we also did not admit participants to the study if they reported any history of neurological disorders, tendency for back pain, color perception deficiencies or if they had a head circumference larger than 58 cm (due to the limited size of the 32-channel head-coil). Gender of participants was self-reported (note that the study was conducted in the German language where there is no clear distinction between sex and gender). We had no reason to suspect any gender differences in the task and therefore did not include this information in the analyses. After data acquisition, we excluded five participants from the analysis; one for severe signal drop in the OFC, i.e. more than 15% less voxels in functional data compared to the OFC mask extracted from freesurfer parcellation of the T1 image (Dale et al., 1999; Klein et al., 2017). One participant was excluded due to excessive motion during fMRI scanning (more than 2mm in any axial direction) and three participants for low performance (less than 75% accuracy in one context in the main task). In the behavioral-replication, 23 young adults took part (15 women, $\mu_{age} = 27.1, \sigma_{age} = 4.91$) and two were excluded for the same accuracy threshold. Due to technical reasons, 3 trials (4 in the replication sample) were excluded since answers were recorded before stimulus was presented and 2 trials (non in the replication) in which RT was faster than 3 SD from the mean (likely premature response). The monetary reimbursement consisted of a base payment of 10 Euro per hour (8.5 for replication sample) plus a performance dependent bonus of 5 Euro on average.

Experimental procedures

Design. Participants performed a random dot-motion paradigm in two phases, separated by a short break (minimum 15 minutes). In the first phase, psychophysical properties of four colors and four motion directions were first titrated using a staircasing task. Then, participants learned the rewards associated with each of these eight features during a outcome learning task. The second phase took place in the MRI scanner and consisted mainly of the main task, in which participants were asked to make decisions between two random dot kinematograms, each of which had one color and/or one direction from the same set. Note there were two additional mini-blocks of 1D trials only, at the end of first- and at the start of the second phase (during anatomical scan, see below). The replication sample completed the same procedure with the same break length, but without MRI scanning. That is, both phases were completed in a behavioral testing room. Details of each task and the stimuli are described below. Behavioral data was recorded during all experiment phases. MRI data was recorded during phase 2. We additionally collected eye-tracking data (EyeLink 1000; SR Research Ltd.; Ottawa, Canada) both during the staircasing and the main decision making task to ensure continued fixation (data not presented). The overall experiment lasted between 3.5 and 4 hours (including the break between the phases). Additional information about the pre-scanning phase can be found in Fig. S1.

Room, Luminance and Apparatus. Behavioral sessions were conducted in a dimly lit room without natural light sources, such that light fluctuations could not influence the perception of the features. A small lamp was stationed in the corner of the room, positioned so it would not cast shadows on the screen. The lamp had a light bulb with 100% color rendering index, i.e. avoiding any influence on color perception. Participants sat on a height adjustable chair at a distance of 60 cm from a 52 cm horizontally wide, Dell monitor (resolution: 1920 x 1200, refresh rate 1/60 frames per second). Distance from the monitor was fixed using a chin-rest with a head-bar. Stimuli were presented using psychtoolbox version 3.0.11 (Brainard & Vision, 1997; Kleiner et al., 2007; Pelli & Vision, 1997) in MATLAB R2017b (*MATLAB version 9.3.0.713579 (R2017b)*, 2017)In the MRI-scanner room lights were switched off and light sources in the

operating room were covered in order to prevent interference with color perception or shadows cast on the screen. Participants lay inside the scanner at distance of 91 cm from a 27 cm horizontally wide screen on which the task was presented a D-ILA JVC projector (D-ILa Projektor SXGA, resolution: 1024x768 , refresh rate: 1/60 frames per second). Stimuli were presented using psychtoolbox version 3.0.11 (Brainard & Vision, 1997; Kleiner et al., 2007; Pelli & Vision, 1997) in MATLAB R2012b (*MATLAB version (R2012b)*, 2017) on a Dell precision T3500 computer running windows XP version 2002.

Stimuli. Each cloud of dots was presented on the screen in a circular array with 7° visual angle in diameter. In all trials involving two clouds, the clouds appeared with 4° visual angle distance between them, including a fixation circle (2° diameter) in the middle, resulting in a total of 18° field of view (following total apparatus size from Pilly & Seitz, 2009). Each cloud consisted of 48 square dots of 3x3 pixels. We used four specific motion and four specific color features.

To prevent any bias resulting from the correspondence between response side and dot motion, each of the four motion features was constructed of two angular directions rotated by 180°, such that motion features reflected an axis of motion, rather than a direction. Specifically, we used the four combinations: 0°-180° (left-right), 45°-225° (bottom right to upper left), 90°-270° (up-down) and 135°-315° (bottom left - upper right). We used a Brownian motion algorithm (e.g. Pilly & Seitz, 2009), meaning in each frame a different set of given amount of coherent dots was chosen to move coherently in the designated directions in a fixed speed, while the remaining dots moved in a random direction (Fig. S1). Dots speed was set to 5° per second (i.e. 2/3 of the aperture diameter per second, following Pilly & Seitz, 2009). Dots lifetime was not limited. When a dot reached the end of the aperture space, it was sent ’back to start’, i.e. back to the other end of the aperture. Crucially, the number of coherent dots (henceforth: motion-coherence) was adjusted for each participant throughout the staircasing procedure, starting at 0.7 to ensure high accuracy (see Pilly & Seitz, 2009). An additional type of motion-direction was ’random-motion’ and was used in 1D color clouds. In these clouds, dots were split to 4 groups of 12, each assigned with one of the four motion features and their adjusted-coherence level, resulting in a balanced

subject-specific representation of random motion.

In order to keep the luminance fixed, all colors presented in the experiment were taken from the YCbCr color space with a fixed luminance of Y = 0.5. YCbCr is believed to represent human perception in a relatively accurate manner (cf. Abbott et al., 2016). In order to generate an adjustable parameter for the purpose of staircasing, we simulated a squared slice of the space for Y = 0.5 (Fig. S1) in which the representation of the dots color moved using a Brownian motion algorithm as well. Specifically, all dots started close to the (gray) middle of the color space, in each frame a different set of 30% of dots was chosen to move coherently towards the target color in a certain speed whereas all the rest were assigned with a random direction. Perceptually, this resulted in all the dots being gray at the start of the trial and slowly taking on the designated color. Starting point for each color was chosen based on pilot studies and was set to a distance of 0.03-0.05 units in color space from the middle. Initial speed in color space (henceforth: color-speed) was set so the dots arrive to their target (23.75% the distance to the corner from the center) by the end of the stimulus presentation (1.6s). i.e. distance to target divided by the number of frames per trial duration. Color-speed was adjusted throughout the staircasing procedure. An additional type of color was 'no color' for motion 1D trials for which we used the gray middle of the color space.

Staircasing task. In order to ensure RTs mainly depended on associated values and not on other stimulus properties (e.g. salience), we created a staircasing procedure that was conducted prior to value learning. In this procedure, motion-coherence and color-speed were adjusted for each participant in order to minimize between-feature detection time differences. As can be seen in Fig. S1, in this perceptual detection task participants were cued (0.5s) with either a small arrow (length $2°$) or a small colored circle ($0.5°$ diameter) to indicate which motion-direction or color they should choose in the upcoming decision. After a short gray (middle of YCbCr) fixation circle (1.5s, diameter $0.5°$), participants made a decision between the two clouds (1.6s). Clouds in this part could be either both single-feature or both dual-features. In dual feature trials, each stimulus had one color and one motion feature, but the cue indicated either a specific motion or a

specific color. After a choice, participants received feedback (0.4s) whether they were (a) correct and faster than 1 second, (b) correct and slower or (c) wrong. After a short fixation (0.4s), another trial started. All timings were fixed in this part. Participants were instructed to always look at the fixation circle in the middle of the screen throughout this and all subsequent tasks. To motivate participants and continued perceptual improvements during the later (reward related) task-stages, participants were told that if they were correct and faster than 1 second in at least 80% of the trials, they will receive an additional monetary bonus of 2 Euros.

The staircasing started after a short training (choosing correct in 8 out of 12 consecutive trials mixed of both contexts) and consisted of two parts: two adjustment blocks an two measurement blocks. All adjustments of color-speed and motion-coherence followed this formula:

$$\theta_i^{t+1} = \theta_i^t + \alpha\theta_i^t \frac{\overline{\mathrm{RT}_i^t} - \mathrm{RT}^0}{\mathrm{RT}^0} \tag{1}$$

where θ_i^{t+1} represents the new coherence/speed for motion or color feature i during the upcoming time interval/block $t+1$, θ_i^t is the level at the time of adjustment, $\overline{RT_i^t}$ is the mean RT for the specific feature i during time interval t, RT_0 is the "anchor" RT towards which the adjustment is made and α represents a step size of the adjustment, which changed over time as described below.

The basic building block of adjustment blocks consisted of 24 cued-feature choices for each context ($4 \times 3 \times 2 = 24$, i.e. 4 colors, each discriminated against 3 other colors, on 2 sides of screen). The same feature was not cued more than twice in a row. Due to time constrains, we could not include all possible feature-pairing combinations between the cued and uncued features. We therefore pseudo-randomly choose from all possible background combinations for each feature choice (unlike later stages, this procedure was validated on and therefore included also trials with identical background features). In the first adjustment block, participants completed 72 trials, i.e. 36 color-cued and 36 motion-cued, interleaved in chunks of 4-6 trials in a non-predictive manner. This included, for each context, a mixture of one building block of 2D trials and half a block of 1D trials, balanced to include 3 trials for each cued-feature. 1D or 2D

trials did not repeat more than 3 times in a row. At the end of the first adjustment block, the mean RT of the last 48 (accurate) trials was taken as the anchor (RT^0) and each individual feature was adjusted using the above formula with $\alpha = 1$. The second adjustment block started with 24 motion-cued only trials which were used to compute a new anchor. Then, throughout a series of 144 trials (72 motion-cued followed by 72 color-cued trials, all 2D), every three correct answers for the same feature resulted in an adjustment step for that specific feature (Eq. 1) using the average RT of these trials ($\overline{RT_i^t}$) and the motion anchor RT^0 for both contexts. This resulted in a maximum of six adjustment steps per feature, where alpha decreased from 0.6 to 0.1 in steps of 0.1 to prevent over-adjustment.

Next, participants completed two measurement blocks identical in structure to the main task (see below) with two exceptions: First, although this was prior to learning the values, they were perceptually cued to chose the feature that later would be assigned with the highest value. Second, to keep the relevance of the feature that later would take the lowest value (i.e. would rarely be chosen), we added 36 additional trials cued to choose that feature (18 motion and 18 color trials per block).

Outcome learning task. After the staircasing and prior to the main task, participants learned to associate each feature with a deterministic outcome. Outcomes associated with the four features on each contexts were 10, 30, 50 and 70 credit-points. The value mapping to perceptual features was assigned randomly between participants, such that all possible color- and all possible motion-combinations were used at least once (4! = 24 combinations per context). We excluded motion value-mapping that correspond to clockwise or counter-clockwise ordering. The outcome learning task consisted only of single-feature clouds, i.e. clouds without coherent motion or dots ‘without’ color (gray). Therefore each cloud in this part only represented a single feature. To encourage mapping of the values for each context on similar scales, the two clouds could be either of the same context (e.g. color and color) or from different contexts (e.g. color and motion). Such context-mixed trials did not repeat in other parts of the experiment.

The first block of the outcome learning task had 80 forced choice trials (5 repetitions of 16 trials:

4 values × 2 Context × 2 sides of screen), in which only one cloud was presented, but participants still had to choose it to observe its associated reward. These were followed by mixed blocks of 72 trials which included 16 forced choice interleaved with 48 free choice trials between two 1D clouds (6 value-choices: 10 vs 30/50/70, 30 vs 50/70, 50 vs 70 × 4 context combinations × 2 sides of screen for highest value). To balance the frequencies with which feature-outcome pairs would be chosen, we added 8 forced choice trials in which choosing the lowest value was required. Trials were pseudo-randomized so no value would repeat more than 3 times on the same side and same side would not be chosen more the three consecutive times. Mixed blocks repeated until participants reached at least 85% accuracy of choosing the higher valued cloud in a block, with a minimum of two and a maximum of four blocks. Since all clouds were 1D and choice could be between contexts, these trials started without a cue, directly with the presentation of two 1D clouds (1.6s). Participants then made a choice, and after short fixation (0.2s) were presented with the value of both chosen and unchosen clouds (0.4s, with value of choice marked with a square around it, see Fig. S1). After another short fixation (0.4s) the next trial started. Participants did not collect reward points in this stage, but were told that better learning of the associations will result in more points, and therefore more money later. Specifically, in the MRI experiment participants were instructed that credit points during the main task will be converted into a monetary bonus such that every 600 points they will receive 1 Euro at the end. The behavioral replication cohort received 1 Euro for every 850 points.

Main task preparation. In preparation of the main task, participants performed one block of 1D trials at the end of phase 1 and then at the start of the MRI session during the anatomical scan. These blocks were included to validate that changing presentation mediums between phases (computer screen versus projector) did not introduce a perceptual bias to any features and as a final correction for post value-learning RT differences between contexts. Each block consisted of 30 color and 30 motion 1D trials interleaved in chunks of 4-7 trials in a non-predictive manner. The value difference between the clouds was fixed to 20 points (10 repetitions of 3 value comparisons × 2 contexts). Trials were pseudo-randomized so no target value was repeated more

than once within context (i.e. not more than twice all in all) and was not presented on the same side of screen more than 3 consecutive trials within context and 4 in total. In each trial, they were first presented with a contextual cue (0.6s) for the trial, followed by short fixation (0.5s) and the presentation of two single-feature clouds of the cued context (1.6s) and had to choose the highest valued cloud. After a short fixation (0.4s), participants were presented with the chosen cloud's outcome (0.4s). The timing of the trials was fixed and shorter than in the remaining main task because no functional MRI data was acquired during these blocks. Participants were instructed that from the first preparation block they started to collect the rewards. Data from these 1D block were used to inspect and adjust for potential differences between the MRI and the behavior setup. First, participants reacted generally slower in the scanner ($t(239) = -9.415$, $p < .001$, paired t-test per subject per feature). Importantly, however, we confirmed that this slowing was uniform across features, i.e. no evidence was found for a specific feature having more RT increase than the rest (ANOVA test on the difference between the phases, $F(7, 232) = 1.007$, $p = .427$). Second, because pilot data indicated increased RT differences between contexts after the outcome learning task we took the mean RT difference between color and motion trials in the second mini-block in units of frames (RT difference divided by the refresh rate), and moved the starting point of each color relative to their target color, the number of frames $\times$ its speed. Crucially, the direction of the move (closer/further to target) was the same for all colors, thus ensuring not to induce within-context RT differences.

Main task. Finally, participants began with the main experiment inside the scanner. Participants were asked to choose the higher-valued of two simultaneously presented random dot kinematograms, based on the previously learned feature-outcome associations. As described in the main text, each trial started with a cue that indicated the current task context (color or motion). In addition, both clouds could either have two features (each a color and a motion, 2D trials) or one feature only from the cued context (e.g., colored, but randomly moving dots).

The main task consisted of four blocks in which 1D and 2D trial were intermixed. Each block contained 36 1D trials (3 EV $\times$ 2 Contexts $\times$ 6 repetitions) and 72 2D trials (3 EV $\times$ 2 Contexts

$\times$ 12 feature-combinations, see fig1c). Since this task took part in the MRI, the duration of the fixation circles were drawn from a truncated exponential distribution with a mean of μ=0.6s (range 0.5s-2.5s) for the interval between cue and stimulus, a mean of μ=3.4s (1.5s-9s) for the interval between stimulus and outcome and a mean of μ=1.25s (0.7s-6s) for the interval between outcome and the cue of the next trial. The cue, stimulus and outcome were presented for 0.6s, 1.6s and 0.8s, respectively. Timing was optimized using VIF-calculations of trial-wise regression models (see Classification procedure section below).

The order of trials within blocks was controlled as follows: the cued context stayed the same for 4-7 trials (in a non-predictive manner), to prevent context confusion caused by frequent switching. No more than 3 repetitions of 1D or 2D trials within each context could occur, and no more than 5 repetition overall. The target did not appear on the same side of the screen on more than 4 consecutive trials. Congruent or incongruent trials did not repeat more than 3 times in a row. In order to avoid repetition suppression, i.e. a decrease in the fMRI signal due to a repetition of information (e.g. Barron et al., 2016; Garvert et al., 2017), no target feature was repeated two trials in a row, meaning the EV could repeat maximum once (i.e. one color and one motion). As an additional control over repetition, we generated 1000 designs according the above-mentioned rules and choose the designs in which the target value was repeated in no more than 10% of trials across trial types, as well as when considering congruent, incongruent or 1D trials separately.

In all mixed effect models, When describing main effects of models, the χ^2 represents Type II Wald χ^2 tests, whereas when describing model comparison, the χ^2 represents the log-likelihood ratio test. Model comparison throughout the paper was done using the 'anova' function. The reason we used χ^2 test is that classification probabilities as well as RSA dissimilarities are not normally distributed (these follow beta and gamma distributions respectively, note that the glmmTMB toolbox also uses χ^2 as its default for these distributions). Regressors were scaled prior to fitting the models for all analyses.

Throughout the behavioral and fMRI analyses we report exact p-values unless they fall below 0.001, in which case we report $p < .001$.

Behavioral analysis

RT data was analyzed in R (R version 3.6.3 R Core Team, 2017, RStudio version 1.3.959 RStudio Team, 2020) using linear mixed effect models (lmer in lme4 1.1-21: Bates et al., 2015). The behavioral model that we found to fit the behavioral RT data best was:

$$\log RT_k^t = \beta_0 + \gamma_{0k} + \beta_1 \mathrm{EV} + \beta_2 \mathrm{Congruency}_t + \beta_3 \mathrm{Congruency}_t \times \mathrm{EV}_{\mathrm{back_t}} + \beta_4 \mathrm{Congruency}_t \times \mathrm{EV}_t + \nu_1 t + \nu_2 \mathrm{side}_t + \nu_3 \mathrm{switch}_t + \nu_4 \mathrm{context}_t \tag{2}$$

where $\log RT_k^t$ is the log reaction time of subject k in trial t, β_0 and γ_{0k} represent global and subject-specific intercepts, ν-coefficients reflect nuisance regressors ($side$ of target object, trials since last context switch and the current $\mathrm{context}$), β_1 to β_4 captured the fixed effect of EV, Congruency, Congruency $\times$ $\mathrm{EV}_{\mathrm{back}}$ and Congruency $\times$ EV, respectively. The additional models reported in the SI included intercept terms specific for each factor level, nested within subject (for EV, Block and Context, see Fig. S2). An exploratory analysis investigating all possible 2-way interactions with all nuisance regressors can be found in Fig. S4.

Investigations of alternative parametrizations of the values can be found in Fig. S3.

Accuracy data was analyzed in R (R version 3.6.3 R Core Team, 2017, RStudio version 1.3.959 RStudio Team, 2020) using generalized linear mixed effect models (glmer in lme4 1.1-21: Bates et al., 2015) employing a binomial distribution family with a 'logit' link function. Regressors were scaled prior to fitting the models for all analyses. No-answer trials of were excluded from this analysis. The model found to fit the behavioral accuracy data best was almost equivalent to the RT model, except for the fourth term involving Congruency $\times$ switch:

$$ACC_k^t = \beta_0 + \gamma_{0k} + \beta_1 \mathrm{EV} + \beta_2 \mathrm{Congruency}_t + \beta_3 \mathrm{Congruency}_t \times \mathrm{EV}_{\mathrm{back}_t} + \beta_4 \mathrm{Congruency}_t \times \mathrm{switch}_t + \nu_1 t + \nu_2 \mathrm{side}_t + \nu_3 \mathrm{switch}_t + \nu_4 \mathrm{context}_t \tag{3}$$

where ACC_k^t is the accuracy (1 for correct and 0 for incorrect) of subject k in trial t and all the rest of the regressors are equivalent to Eq. 2. An exploratory analysis investigating all possible 2-way interactions with all nuisance regressors can be found in Fig. S5. We note that the interaction Congruency $\times$ switch indicates that participants were more accurate the further they were from a context switch point. Out of the nuisance variables, only 'switch' influenced accuracy, Type II Wald χ^2 test in baseline model: $\chi^2_{(1)} = 10.22, p = .001$.

fMRI data

fMRI data acquisition. MRI data was acquired using a 32-channel head coil on a research-dedicated 3-Tesla Siemens Magnetom TrioTim MRI scanner (Siemens, Erlangen, Germany) located at the Max Planck Institute for Human Development in Berlin, Germany. High-resolution T1-weighted (T1w) anatomical Magnetization Prepared Rapid Gradient Echo (MPRAGE) sequences were obtained from each participant to allow registration and brain surface reconstruction (sequence specification: 256 slices; TR = 1900 ms; TE = 2.52 ms; FA = 9 degrees; inversion time (TI) = 900 ms; matrix size = 192 x 256; FOV = 192 x 256 mm; voxel size = 1 x 1 x 1 mm). This was followed with two short acquisitions with six volumes each that were collected using the same sequence parameters as for the functional scans but with varying phase encoding polarities, resulting in pairs of images with distortions going in opposite directions between the two acquisitions (also known as the blip-up / blip-down technique). From these pairs the displacements were estimated and used to correct for geometric distortions due to susceptibility-induced field inhomogeneities as implemented in the the fMRIPrep preprocessing pipeline. In addition, a whole-brain spoiled gradient recalled (GR) field map with dual echo-time images (sequence specification: 36 slices; A-P phase encoding direction; TR = 400 ms; TE1 = 4.92 ms; TE2 = 7.38 ms; FA = 60 degrees; matrix size = 64 x 64; 619 FOV = 192 x 192 mm; voxel size = 3 x 3 x 3.75 mm) was obtained as a potential alternative to the method described above. However, this GR frield map was not used in the preprocessing pipeline. Lastly, four functional runs using a multi-band sequence (sequence specification: 64 slices in interleaved ascending order; anterior-to-posterior (A-P) phase encoding direction; TR = 1250 ms; echo time

(TE) = 26 ms; voxel size = 2 x 2 x 2 mm; matrix = 96 x 96; field of view (FOV) = 192 x 192 mm; flip angle (FA) = 71 degrees; distance factor = 0, MB acceleration factor = 4). A tilt angle of -30 degrees from AC-PC (tilted backwards, or: front side of FOV upwards) was used in order to maximize signal from the orbitofrontal cortex (OFC, see Weiskopf et al., 2006). For each functional run, the task began after the acquisition of the first four volumes (i.e., after 5.00 s) to avoid partial saturation effects and allow for scanner equilibrium. Each run was about 15 minutes in length, including a 20 seconds break in the middle of the block (while the scanner is running) to allow participants a short break. We measured respiration and pulse during each scanning session using pulse oximetry and a pneumatic respiration belt part of the Siemens Physiological Measurement Unit. Full details of the sequences used, as provided by the MRI scanner, are shared in the same repository with the code (see ”MRI_Sequences.pdf”)

BIDS conversion and defacing. Data was arranged according to the brain imaging data structure (BIDS) specification (K. J. Gorgolewski et al., 2016) using the HeuDiConv tool (version 0.6.0.dev1; freely available from https://github.com/nipy/heudiconv). Dicoms were converted to the NIfTI-1 format using dcm2niix [version 1.0.20190410 GCC6.3.0; (X. Li et al., 2016)]. In order to make identification of study participants highly unlikely, we eliminated facial features from all high-resolution structural images using pydeface (version 2.0; available from https://github.com/poldracklab/pydeface). The data quality of all functional and structural acquisitions were evaluated using the automated quality assessment tool MRIQC [for details, (see Esteban et al., 2017), and the MRIQC documentation]. The visual group-level reports confirmed that the overall MRI signal quality was consistent across participants and runs.

fMRI preprocessing. Data was preprocessed using fMRIPrep 1.2.6 (Esteban, Markiewicz, et al., 2018; Esteban, Blair, et al., 2018; RRID:SCR_016216), which is based on Nipype 1.1.7 (K. Gorgolewski et al., 2011; K. J. Gorgolewski et al., 2018; RRID:SCR_002502). Many internal operations of fMRIPrep use Nilearn 0.5.0 (Abraham et al., 2014, RRID:SCR_001362), mostly within the functional processing workflow.

Specifically, the T1-weighted (T1w) image was corrected for intensity non-uniformity (INU)

using `N4BiasFieldCorrection` (Tustison et al., 2010, ANTs 2.2.0), and used as a T1w-reference throughout the workflow. The anatomical image was skull-stripped using `antsBrainExtraction.sh` (ANTs 2.2.0), using OASIS as the target template. Brain surfaces were reconstructed using `recon-all` (FreeSurfer 6.0.1, RRID:SCR_001847, Dale et al., 1999), and the brain masks were estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438, Klein et al., 2017). Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c (Fonov et al., 2009, RRID:SCR_008796) was performed through nonlinear registration with `antsRegistration` (ANTs 2.2.0, RRID:SCR_004757, Avants et al., 2008), using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using `fast` (FSL 5.0.9, RRID:SCR_002823, Y. Zhang et al., 2001).

To preprocess the functional data, a reference volume for each run and its skull-stripped version were generated using a custom methodology of fMRIPrep. A deformation field to correct for susceptibility distortions was estimated based on two echo-planar imaging (EPI) references with opposing phase-encoding directions, using `3dQwarp` (Cox & Hyde, 1997) (AFNI 20160207). Based on the estimated susceptibility distortion, an unwarped BOLD reference was calculated for a more accurate co-registration with the anatomical reference. The BOLD reference was then co-registered to the T1w reference using `bbregister` (FreeSurfer), which implements boundary-based registration (Greve & Fischl, 2009). Co-registration was configured with nine degrees of freedom to account for distortions remaining in the BOLD reference. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using `mcflirt` (FSL 5.0.9, Jenkinson et al., 2002). BOLD runs were slice-time corrected using `3dTshift` from AFNI 20160207 (Cox & Hyde, 1997, RRID:SCR_005927) and aligned to the middle of each TR. The BOLD time-series (including slice-timing correction) were resampled

onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep.

Several confound regressors were calculated during preprocessing: Six head-motion estimates (see above), Framewise displacement, six anatomical component-based noise correction components (aCompCorr) and 18 physiological parameters (8 respiratory, 6 heart rate and 4 of their interaction). The head-motion estimates were calculated during motion correction (see above). Framewise displacement was calculated for each functional run, using the implementations in Nipype (following the definitions by Power et al., 2014). A set of physiological regressors were extracted to allow for component-based noise correction (CompCor, Behzadi et al., 2007). Principal components are estimated after high-pass filtering the BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor, unused) and anatomical (aCompCor). For aCompCor, six components are calculated within the intersection of the aforementioned mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction, and co-registrations to anatomical and template spaces). Gridded (volumetric) resamplings were performed using `antsApplyTransforms` (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Lastly, for the 18 physiological parameters, correction for physiological noise was performed via RETROICOR (Glover et al., 2000; Hutton et al., 2011) using Fourier expansions of different order for the estimated phases of cardiac pulsation (3rd order), respiration (4th order) and cardio-respiratory interactions (1st order) (A. K. Harvey et al., 2008): The corresponding confound regressors were created using the Matlab PhysIO Toolbox (Kasper et al., 2017, open source code available as part of the TAPAS software collection (Version 3.2.0): https://www.translationalneuromodeling.org/tapas. For more

details of the pipeline, and details on other confounds generated but not used in our analyses, see the section corresponding to workflows in fMRIPrep's documentation.

For univariate analyses, BOLD time-series were re-sampled to MNI152NLin2009cAsym standard space in the fMRIPrep pipeline and then smoothed using SPM (Penny et al., 2011, SPM12 (7771)) with 8mm FWHM, except for ROI generation, where a 4mm FWHM kernel was used. Multivariate analyses were conducted in native space, and data was smoothed with 4mm FWHM using SPM (Penny et al., 2011, SPM12 (7771)). Classification analyses further involved three preprocessing steps of voxel time-series: First, extreme-values more than 8 standard deviations from a voxels mean were corrected by moving them by 50% their distance from the mean towards the mean (this was done to not bias the last z scoring step). Second, the time-series of each voxel was detrended, a high-pass filter at 128 Hz was applied and confounds were regressed out in one action using Nilearn 0.6.2 (later changed to 0.7.0) (Abraham et al., 2014). Lastly, the time-series of each voxel for each block was z scored.

Univariate fMRI analysis

All GLMs were conducted using SPM12 (Penny et al., 2011, SPM12 (7771)) in MATLAB (*MATLAB version 9.3.0.713579 (R2017b)*, 2017). All GLMs consisted of two regressors of interest corresponding to the onsets of the two trial-types (1D/2D, except for one GLM where 2D onsets were split by Congruency) and included one parametric modulator of EV assigned to 1D onset and different combinations of parametric modulators of EV, Congruency, EV_{back} and their interactions (see Fig.S13 for GLM visualization). All parametric modulators were demeaned before entering the GLM, but not orthogonalized. Regressors of no interest reflected cue onsets in Motion and Color trials, stimulus onsets in wrong and no-answer trials, outcome onsets and 31 nuisance regressors (e.g. motion and physiological parameters, see fMRI-preprocessing). The duration of stimulus regressors corresponded to the time the stimuli were on screen. The durations for the rest of the onset regressors were set to 0. Microtime resultion was set to 16 (64 slices / 4 MB factor) and microtime onset was set to the 8 (since slice time correction aligned to middle slice, see fMRI-preprocessing). Data for all univariate analyses were masked with a whole

brain mask computed as intercept of each functional run mask generated from fMRIprep (Dale et al., 1999; Klein et al., 2017). MNI coordinates were translated to their corresponding brain regions using the automated anatomical parcellation toolbox (Rolls et al., 2015, 2020; Tzourio-Mazoyer et al., 2002, AAL3v1) for SPM. We verified the estimability of the design matrices by assessing the Variance Inflation Factor (VIF) for each onset regressor in the HRF-convolved design matrix. Specifically, for each subject, we computed the VIF (assisted by scripts from https://github.com/sjgershm/ccnl-fmri) for each regressor in the HRF-convolved design matrix and averaged the VIFs of corresponding onsets across the blocks. None of the VIFs surpassed a value of 3.5 (a value of 5 is considered a conservative indicator for overly colinear regressors, e.g. Mumford et al., 2015, see Fig.S13 for details). Detailed descriptions of all GLMs are reported in the main text. Additional GLMs verifying the lack of Congruency in any frontal region can be found in Fig.S13.

Functionally defined vmPFC ROI. Our fMRI analyses focused on understanding the representations of expected values in vmPFC. We therefore first sought to identify a value-sensitive region of interest (ROI) that reflected expected values in 1D and 2D trials, following common procedures in the literature (e.g. Bartra et al., 2013). We analyzed the fMRI data using general linear models (GLMs) with separate onsets and EV parametric modulators for 1D and 2D trials (at stimulus presentation with 0s duration) and defined a functional ROI for value representations centered on vmPFC using the the union of the EV modulators for 1D and 2D trials (EV_{1D} + EV_{2D} >0), Fig.3a, $p < .0005$ FDR corrected). Note that this GLM had no information regarding the contextually irrelevant context. The group ROI was generated in MNI space and included 998 voxels. Multivariate analyses were conducted in native space and the ROI was transformed to native space using ANTs and nearest neighbor interpolation (ANTs 2.2.0 Avants et al., 2008) while keeping only voxels within the union of subject- and run-specific brain masks produced by the fMRIprep pipeline (Dale et al., 1999; Klein et al., 2017). The resulting subject-specific ROIs therefore had varying number of voxels (μ = 768.14, σ = 65.62, min = 667, max = 954).

Verifying design trial-wise estimability. To verify that the individual trials are estimatable (for the trial-wise multivariate analysis) and as a control over multi-colinearity (Mumford et al., 2015), we convolved a design matrix with the HRF for each subject with one regressor per stimuli (432 regressors with duration equal to the stimulus duration), two regressor across all cues (split by context) and three regressor for all outcomes (one for each EV). We then computed the VIF for each stimulus regressor (i.e. how predictive is each regressor by the other ones). None of the VIFs surpassed 1.57 across all trials and subjects ($\mu_{VIF} = 1.42, \sigma_{VIF} = .033$, min = 1.34). When repeating this analysis with a GLM in which also outcomes were split into trialwise regressors, we found no stimuli VIF larger than 3.09 ($\mu_{VIF} = 2.64, \sigma_{VIF} = .132$, min = 1.9). Note that 1 is the minimum (best) value and 5 is a relatively conservative threshold for colinearity issues (e.g. Mumford et al., 2015). This means that the BOLD responses of individual trials can be modeled separately and should not have colinearity issues with other stimuli nor with the outcome presentation of each trial.

Multivariate analysis

RDM analyses. RDM was conducted using betas taken from a GLM fit to data in native space (4mm smoothing) with one onset for EV of 1D trials and one onset for each combination or EV and EV_{back} for 2D trials (e.g. one onset for all trials where EV=30 and EV_{back}=30, one onset when EV=30 and EV_{back}=50 etc.). Duration of the onsets was set to 0. Regressors of no interest were identical to the GLMs described in 'Univariate fMRI analysis' section above. For each subject, we extracted the beta values for each run from the above defined functional ROI for each one of the 2D onset regressors. We then performed multvariate noise normalization (normalize each voxel by its residuals, Walther et al., 2016) and mean pattern subtraction (i.e. subtract the mean pattern across conditions for each voxel from each response pattern, Walther et al., 2016). Lastly, we computed the Euclidean distance between each pair of patterns across runs using Nilearn (Abraham et al., 2014). Note that noise-normalized Euclidean distance is equivalent to the Mahalanobis distance (Walther et al., 2016). To prevent biasing the diagonal, we excluded any correlation within a run across conditions (where the diagonal would be 1). This resulted in a

9x9 RDM for each subject and each block comparison. The resulting distances (half the matrix including the diagonal for each subject) were analyzed in R (R version 3.6.3 R Core Team, 2017, RStudio version 1.3.959 RStudio Team, 2020) with Generalized Linear Mixed Models using Template Model Builder (glmmTMB, Brooks et al., 2017) models, employing a gamma distribution family with a 'inverse' link function. When describing main effects of models, the χ^2 represents Type II Wald χ^2 tests, whereas when describing model comparison, the χ^2 represents the log-likelihood ratio test. Model comparison throughout the paper was done using the 'anova' function. Throughout all the analyses, each regressor was scaled prior to fitting the models.

The best explaining model for the main effects of the RDM was:

$$d_{i,j}^{k} = \beta_0 + \gamma_{0k} + \beta_1 \text{Diagonal}_{\text{EV}} + \beta_2 \text{Diagonal}_{\text{EV}_{\text{back}}} + \zeta_{0_{k,\text{frequency}}} \tag{4}$$

where $d_{i,j}^{k}$ is the Mahalanobis distance of combination i and j for subject k, where i and j each represent all possible patterns (i.e. combination of EV and EV_{back}. β_0 and γ_{0k} represent global and subject-specific intercepts. $Diagonal_{\text{EV}}$ is 1 when the EV of pattern i is the same as the EV of pattern j. $Diagonal_{\text{EV}_{\text{back}}}$ is 1 when the EV_{back} of pattern i is the same as EV_{back} of pattern j. $\zeta_{0_{k,\text{frequency}}}$ is an additional intercept for every level of frequency nested within each within each subject level. For details on the effect of frequency, see Fig.S6.

The best explaining model for the value difference effects of the RDM was:

$$d_{i,j}^{k} = \beta_0 + \gamma_{0k} + \beta_1 \text{ValueDifference}_{\text{EV}} + \beta_2 \text{ValueDifference}_{\text{EV}_{\text{back}}} + \zeta_{0_{k,\text{frequency}}} \tag{5}$$

where all parameters are identical to eq. 4 above, only that $ValueDifference_{\text{EV}}$ corresponds to the value difference between the EV of pattern i and the EV of pattern j and $\text{ValueDifference}_{\text{EV}_{\text{back}}}$ is the value difference between the EV_{back} of pattern i and the EV_{back} of pattern j.

Classification procedure. The training set for Value and Context classifiers consisted of fMRI data from behaviorally accurate 1D trials. For each trial, we took the TR corresponding to approx.

5 seconds after stimulus onset ($round(onset + 5)$) to match the peak of the Haemodynamic Response Function (HRF) estimated by SPM (Penny et al., 2011). Training of Value and Context classifiers was done using a leave-one-run-out scheme across the four runs with 1D trials. To avoid bias in the training set after sub-setting only to behaviorally accurate trials (i.e. over-representation of some information) we up-sampled each training set to ensure equal number of examples in the training set for each combination of EV (3), Context (2) and Chosen-Side (2). Specifically, if one particular category was less frequent than another (e.g., more value-30, left, color trials than value-50, left-color trials) we up-sampled that example category by randomly selecting a trial from the same category to duplicate in the training set, whilst prioritising block-wise balance (i.e., if one block had 2 trials in the chunk and another block had only 1, we first duplicated the trial from under-represented block etc.). We did not up-sample the testing set. The $\mathrm{EV}_{\mathrm{back}}$ classifiers were trained on behaviorally accurate 2D trials (5 seconds after stimulus onset) and up-sampled by EV (3), Context (2) and $\mathrm{EV}_{\mathrm{back}}$ (3) (without Chosen-Side as this resulted in excluding many subjects for lack of trials in some training sets). Due to strong imbalance of unique examples of $\mathrm{EV}_{\mathrm{back}}$ in the training sets (see below) we trained 3 one-vs-rest classifiers, each tasked with identifying one level of $\mathrm{EV}_{\mathrm{back}}$. This required to adjust the sample weights in order to account for the higher frequency of the 'rest' compared to the 'one' label.

Decoding was conducted using multinomial logistic regression as implemented in scikit-learn 0.22.2 (Pedregosa et al., 2011), using a C parameter of 1.0, L2 regularization and the lbgfs solver. For each test example (i.e. trial) we obtained the predicted probability per class. To avoid numerical issues in the subsequent modeling of the classifier's predictions, probabilities were constrained to lie within 0.00001 and 0.99999, rather than 0 and 1. In addition to the probabilities, we obtained the balanced classification accuracy (i.e. is the class with the highest probability also the correct class of the test trial). We separately averaged classification for each participant, test fold and label (this ensured controlling for any label imbalance in the testing set).

In the classification analyses we modelled directly the class probabilities estimated by the classifiers with beta regression mixed effects models (Magnusson et al., 2017). For technical

reasons, before modelling the probabilities using linear mixed effects models, we averaged the classifiers probabilities across the nuisance effects, i.e. we obtained one average probability for each combination of relevant and irrelevant values. Crossing each level of EV (three levels) with each level of irrelevant value of the chosen side combined with irrelevant value of the non-chosen side (12 level, see Fig. 1), resulted in 36 combinations per participant. Note that the relevant value of the unchosen cloud was always EV - 20 and therefore we did not include this as a parameter of interest. After averaging, we computed for each combination of values the EV_{back}, Congruency and alternative parameters (see Fig. S9). The main model comparison, as well as the lack of effects of any nuisance regressor, was confirmed on a dataset with raw, i.e. non-averaged, probabilities (see Fig S7 and S9). Because in the one-vs-rest training of EV_{back} classifiers the three class probabilities for each trial were obtained independently, they sum to 1. We therefore first normalized the probabilities for each testing trial.

Probabilities were analyzed in R (R version 3.6.3 R Core Team, 2017, RStudio version 1.3.959 RStudio Team, 2020) with Generalized Linear Mixed Models using Template Model Builder (glmmTMB, Brooks et al., 2017) models, employing a beta distribution family with a 'logit' link function. When describing main effects of models, the χ^2 represents Type II Wald χ^2 tests, whereas when describing model comparison, the χ^2 represents the log-likelihood ratio test. Model comparison throughout the paper was done using the 'anova' function. Throughout all the analyses, each regressor was scaled prior to fitting the models. Lastly, for the analysis of behavioral accuracy (Fig. 6) we also included behaviorally wrong trials.

Additional coding of the analyses in Python (3.7, **Python3-10.5555/1593511**) using NumPy (1.19.5, Harris et al., 2020) and pandas (1.1.5, **reback2020pandas**). Most of the plots were produced using ggplot2 (3.3.5, Wickham, 2016)

Value similarity analyses. asked whether the predicted probabilities reflected the difference from the objective probability class.

The model we found to best explain the data was:

$$P^k_{t,c} = \beta_0 + \gamma_{0k} + \beta_1|\mathrm{EV}_t - c_t| + \beta_2|\mathrm{EV}_t - c_t|\mathrm{EV}_{\mathrm{back}_t} \tag{6}$$

where $\mathrm{P}^k_{t,c}$ is the probability that the Value classifier assigned to class c in trial t for subject k, β_0 and γ_{0k} represent global and subject-specific intercepts, $|\mathrm{EV}_t - Class_{c,t}|$ is the absolute difference between the EV of the trial and the class the probability is assigned to and $|\mathrm{EV}_t - Class_{c,t}|\mathrm{EV}_{\mathrm{back}_t}$ is the interaction of this absolute difference with $\mathrm{EV}_{\mathrm{back}}$. For models nested in the levels of EV, we included $\zeta_{0_{k,\mathrm{EV}}}$, which is the EV-specific intercept nested within each within each subject level. In these models, testing for main effects of $\mathrm{EV}_{\mathrm{back}}$ or Congruency was not sensible because both factors don't discriminate between the classes, but rather assign the same value to all three probabilities from that trial (which sum to 1). More details can be found in Fig. S7

Values, not perceptual features and not attention capture, explain our effects best. For the feature similarity model we substituted $|\mathrm{EV}_t - c_t|$ from eq. 6 with a "similarity" parameter that encoded the perceptual similarity between each trial in the test set and the perceptual features that constituted the training examples of each class of the classifier. For 1D trials, this perceptual parameter was identical to the value similarity parameter ($|\mathrm{EV}_t - c_t|$). This was because from the shown pairs of colors, both colors overlapped between training and test if the values were identical; one color overlapped if the values were different by one reward level (e.g. a 30 vs 50 comparison corresponded to two trials that involved pink vs green and green vs orange, i.e. sharing the color green); and no colors overlapped if the values were different by two levels (30 vs 70). On 2D trials however, due to changing background features and their value-difference variation, perceptual similarity of training and test was not identical to value similarity. Even though both the value similarity and the perceptual similarity parameter correlated ($\rho = .789$, $\sigma = .005$), we found that the value similarity model provided a better AIC score (value similarity AIC: -3898, Feature similarity AIC: -3893, Fig. panel d). Detailed description with examples can

be found in Fig. S7. Crucially, even when keeping the value difference of the irrelevant features at 20, thus limiting the testing set only to trials with feature-pairs that were included in the training, our value similarity model provided a better AIC (-1959) than the feature similarity model (-1956). To test for a perceptual alternative of $\mathrm{EV}_{\mathrm{back}}$ we substituted the corresponding parameter from the model with $\mathrm{Similarity}_{\mathrm{back}}$. This perceptual parameter takes on 1 if the perceptual feature corresponding to the $\mathrm{EV}_{\mathrm{back}}$ appeared in the 1D training class (as highest or lowest value) and 0 otherwise. As described in the main text, none of the perceptual-similarity encoding alternatives provided a better fit than our models that focused on the expected values the features represented.

Modelling the influence of irrelevant values and Context signals on EV representation. The following model of the probability of the objective EV was found to explain the data best:

$$P^k_{t,\mathrm{EV}} = \beta_0 + \gamma_{0k} + \beta_1 \mathrm{EV}_{\mathrm{back}_t} + \beta_2 P^k_{t,\mathrm{Context}} \tag{7}$$

where $\mathrm{P}^k_{t,\mathrm{EV}}$ is the probability assigned to the objective class by the Value classifier (corresponding to EV of the trial t) for subject k, β_0 and γ_{0k} represent global and subject-specific intercepts, $\mathrm{EV}_{\mathrm{back}}$ is the maximum of the two ignored values (or the EV of the contextually irrelevant context) and $\mathrm{P}^k_{t,\mathrm{Context}}$ is the probability assigned to the objective class by the Context classifier (logit-transformed , i.e. $logit(P) = \log \frac{P}{1-P}$, and scaled for each subject). For models nested in the levels of EV, we included $\zeta_{0_{k,\mathrm{EV}}}$ which is EV specific intercept nested within each within each subject level (see Fig. S9). Investigations of alternative parametrizations of the values can be found in Fig. S9. Including an additional regressor that encoded trials in which EV=$\mathrm{EV}_{\mathrm{back}}$ (or: match) which did not improve model fit, and no evidence for an interaction of the match regressor with the $\mathrm{EV}_{\mathrm{back}}$ was found (LR test with added terms: $\chi^2_{(1)} = 0.45, p = .502$, $\chi^2_{(1)} = 0.77, p = .379$, respectively). This might indicate that when value expectations of both contexts matched, there was neither an increase nor a decrease of P_{EV}.

To compute the correlations between each pair of classes we transformed the probabilities for each class using a multinomial logit transform. For example, for class 30 we performed probabilities were transformed with $mlogit(P_{t,30}) = 0.5(\log \frac{P_{t,30}}{P_{t,50}} + \log \frac{P_{t,30}}{P_{t,70}})$. To examine the

relationship between EV and $\mathrm{EV}_{\mathrm{back}}$, we only included 2D trials in which EV $\neq$ $\mathrm{EV}_{\mathrm{back}}$. This allowed us to categorize all three probabilities as either EV, $\mathrm{EV}_{\mathrm{back}}$ or Other, whereby Other reflected the value that was neither the EV, nor the $\mathrm{EV}_{\mathrm{back}}$. To prevent bias we included only trials in which Other was presented on screen (as relevant or irrelevant value). We then averaged across nuisance regressors (see Classification procedure) and computed the correlation across all trials (Spearman rank correlation). Lastly, we Fisher z-transformed the correlations ($0.5 \log \frac{1+\rho}{1-\rho}$) to approximate normality for the t test. To validate these results, we performed an additional model comparison in which we added a term of the logit transformed $\mathrm{P}_{\mathrm{EV}_{\mathrm{back}}}$ or of P_{other} to Eq. 7 ($\beta_2 mlogit(P_{t,\mathrm{EV}_{\mathrm{back}}})$ or $\beta_2 mlogit(P_{t,Other})$,respectively). As reported in the main text, adding a term reflecting $\mathrm{P}_{\mathrm{EV}_{\mathrm{back}}}$ resulted in a smaller (better) AIC score than when we added a term for P_{other} (-567,-475, respectively). This was also preserved when running the analysis including nuisance regressors (see νs in Eq. 2) on the non-averaged data (AICs: -5913.3,-5813.3). We note that subsetting the data the way we did resulted in a strong negative correlation in the design matrix between EV and $\mathrm{EV}_{\mathrm{back}}$ ($\rho = -0.798$, averaged across subjects). Although this should not directly influence our interpretation, we validated the results by using alternative models with effects hierarchically nested within the levels of EV and $\mathrm{EV}_{\mathrm{back}}$ (Averaged data AICs: -560, -463, Raw data AICs: -5906.8,-5804.3)

As previously clarified, $\mathrm{P}^{2D}_{\mathrm{EV}_{\mathrm{back}}}$ was derived from a classifier trained on 2D trials. The number of unique examples for each class of $\mathrm{EV}_{\mathrm{back}}$ differed drastically (due to our design, see Fig.1c and S6), which motivated us to split the decoding of $\mathrm{EV}_{\mathrm{back}}$ to three classifiers, each trained on a different label (see 'Classification procedure'). However, our approach of combining one-vs-rest training with oversampling and sample weights could not fully counteract these imbalances and a balanced accuracy did not surpass chance level (t-test against chance: $t_{(34)} = 0.96, p = .171$) and the probabilities each classifier assigned to its corresponding class ($\mathrm{P}^{2D}_{\mathrm{EV}_{\mathrm{back}}}$) were still biased by class imbalances. Specifically, the correlation of $\mathrm{P}^{2D}_{\mathrm{EV}_{\mathrm{back}}}$ and $\mathrm{EV}_{\mathrm{back}}$ was $\rho_\mu = .26, \rho_\sigma = .07$ across subjects, where '2D' indicates the classifier was directly trained on 2D trials, unlike with $\mathrm{P}_{\mathrm{EV}_{\mathrm{back}}}$ which comes from a classifier trained on EV in 1D trials. Since in this analysis we were

mainly interested in the neural representation of $\mathrm{EV}_{\mathrm{back}}$ regardless of whether $\mathrm{EV}_{\mathrm{back}}$ was 30, 50 or 70 in given trial, we solved this issue by using mixed effect models and setting a random intercept for each level of $\mathrm{EV}_{\mathrm{back}}$ (i.e. running the models nested within the levels of $\mathrm{EV}_{\mathrm{back}}$). Importantly, due to the symmetric nature of the RDM, this trial frequency bias is orthogonal to the main effect of $\mathrm{EV}_{\mathrm{back}}$ reported earlier (Fig.S6a-c).

Thus, when testing across the levels of $\mathrm{EV}_{\mathrm{back}}$, the model that best explained the data was:

$$P^k_{t,\mathrm{EV}} = \beta_0 + \gamma_{0k} + \beta_1 \mathrm{EV}_{\mathrm{back}_t} + \beta_2 P^k_{t,\mathrm{Context}} + \beta_3 P^{k,2D}_{t,\mathrm{EV}_{\mathrm{back}}} + \beta_4 P^k_{t,\mathrm{Context}} P^{k,2D}_{t,\mathrm{EV}_{\mathrm{back}}} + \zeta_{0_{k,\mathrm{EV}_{\mathrm{back}}}} \tag{8}$$

where similar to Eq. 7, $\mathrm{P}^k_{t,\mathrm{EV}}$, is the probability assigned to the EV class by the Value classifier for trial t and subject k, β_0 and γ_{0k} represent global and subject-specific intercepts and $\mathrm{P}^k_{t,\mathrm{Context}}$ is the logit-transformed probability assigned to Context class. $\mathrm{P}^{k,2D}_{t,\mathrm{EV}_{\mathrm{back}}}$ is the probability the $\mathrm{EV}_{\mathrm{back}}$ classifier assigned the correct class (in main text: $\mathrm{P}^{2D}_{\mathrm{EV}_{\mathrm{back}}}$, where 2D notes that this classifier was trained on 2D trials) and $\zeta_{0_{k,\mathrm{EV}_{\mathrm{back}}}}$ is $\mathrm{EV}_{\mathrm{back}}$ specific intercept nested within each within each subject level.

Linking MRI effects to behavior. When modelling the probability of $\mathrm{EV}_{\mathrm{back}}$ from the Value classifier ($\mathrm{P}_{\mathrm{EV}_{\mathrm{back}}}$, Fig. 6a.), we did not average across nuisance regressors. Our baseline model was: $P^k_{t,\mathrm{EV}_{\mathrm{back}}} = \beta_0 + \gamma_{0k} + \nu_1 side(t) + \nu_2 \mathrm{switch}(t) + \nu_3 \mathrm{Context}(t)$. Neither including a main effect nor interactions between EV, $\mathrm{EV}_{\mathrm{back}}$ and Congruency improved model fit. When including behaviorally wrong trials in the model, we used `drop1` in combination with χ^2-tests from `lmer4` package (Bates et al., 2015) to test which of the main effects or interactions improves the fit. This resulted in the following model as best explaining the data:

$$\begin{aligned} P^k_{t,\mathrm{EV}_{\mathrm{back}}} = \beta_0 + \gamma_{0k} + \beta_1 \mathrm{EV}_t \times \mathrm{EV}_{\mathrm{back}_t} + \beta_2 \mathrm{Congruency}_t \times \mathrm{Accuracy}_t \\ + \nu_1 t + \nu_2 \mathrm{side}_t + \nu_3 \mathrm{switch}_t + \nu_4 \mathrm{Context}_t \end{aligned} \tag{9}$$

where $\mathrm{P}^k_{t,\mathrm{EV}_{\mathrm{back}}}$ is the probability the Value classifier assigned to the $\mathrm{EV}_{\mathrm{back}}$ class (corresponding

to $\mathrm{EV}_{\mathrm{back}}$ of trial t) for subject k, β_0 and γ_{0k} represent global and subject-specific intercepts, EV is the maximum of the two relevant and $\mathrm{EV}_{\mathrm{back}}$ is the maximum of the two ignored values. Congruency reflects whether the actions chosen in the relevant vs. irrelevant context would be the same, and the Accuracy regressor has 1 if participants chose the highest relevant value and 0 otherwise. We note that the interaction EV $\times$ $\mathrm{EV}_{\mathrm{back}}$ ($\chi^2_{(1)} = 4.18$, $p = .041$) indicates higher in trials in which EV and $\mathrm{EV}_{\mathrm{back}}$ were more similar, the probability assigned to $\mathrm{EV}_{\mathrm{back}}$ was higher. However, we find this effect hard to interpret since this corresponds to the value similarity effect we previously reported.

In order to investigate the effect of vmPFC neural representations on behavioral accuracy, we used hierarchical model comparison to directly test the influence of neural representation of EV, $\mathrm{EV}_{\mathrm{back}}$ and Context on behavioral accuracy separately for congruent and incongruent trials (Fig. 6b-c.). First, we tested if adding $logit(P_{t,\mathrm{Context}})$, $mlogit(P_{t,\mathrm{EV}})$ or $mlogit(P_{t,EV_{back}})$ to Eq. 3, would help to explain the behavioral accuracy better. Because the analysis was split for congruent and incongruent trials, we excluded the terms involving a Congruency effect. For incongruent trials, only $logit(P_{t,\mathrm{Context}})$ improved the fit (LR-tests: $logit(P_{t,\mathrm{Context}})$: $\chi^2_{(1)} = 3.66$, $p = .055$, $mlogit(P_{t,\mathrm{EV}})$: $\chi^2_{(1)} = 0.28$, $p = .599$, $mlogit(P_{t,\mathrm{EV_{back}}})$: $\chi^2_{(1)} = 0.0$, $p = .957$). In a second step we then separately tested the interactions $logit(P_{t,\mathrm{Context}}) \times mlogit(P_{t,\mathrm{EV}})$ or $logit(P_{t,\mathrm{Context}}) \times mlogit(P_{t,\mathrm{EV_{back}}})$ and found that only the latter had improved the fit ($\chi^2_{(1)} = 1.78$, $p = .183$, $\chi^2_{(1)} = 6.33$, $p = .012$, respectively). For congruent trials, only $mlogit(P_{t,\mathrm{EV_{back}}})$ and marginally $mlogit(P_{t,\mathrm{EV}})$ improved the fit (LR-tests: $logit(P_{t,\mathrm{Context}})$: $\chi^2_{(1)} = 0.0$, $p = .922$, $mlogit(P_{t,\mathrm{EV}})$: $\chi^2_{(1)} = 3.5$, $p = .061$, $mlogit(P_{t,\mathrm{EV_{back}}})$: $\chi^2_{(1)} = 6.48$, $p = .011$). In a second step we tested separately the interactions $logit(P_{t,\mathrm{Context}}) \times mlogit(P_{t,\mathrm{EV}})$,$logit(P_{t,\mathrm{Context}}) \times mlogit(P_{t,\mathrm{EV_{back}}})$ or $mlogit(P_{t,\mathrm{EV_{back}}}) \times mlogit(P_{t,\mathrm{EV}})$ and found none of these improved model fit when adding them to a model that included both main effects from the previous step ($\chi^2_{(1)} = 0.34$, $p = .560$, $\chi^2_{(1)} = .278$, $p = .598$, $\chi^2_{(1)} = 2.49$, $p = .115$, respectively).

To investigate the effect of vmPFC neural representations on RT in behaviorally accurate trials, we asked whether subjects who had a stronger effect of Context representation ($\mathrm{P_{context}}$) on EV

representation ($\mathrm{P_{EV}}$) or a stronger Spearman rank correlation between $\mathrm{P_{EV}}$ and $\mathrm{P_{EV_{back}}}$ (taken from the Value classifier) also had a stronger effect of Congruency on their RT. Additionally, we asked whether subjects who had a stronger effect of $\mathrm{EV_{back}}$ on $\mathrm{P_{EV}}$ and or a stronger effect of $\mathrm{P}^{k,2D}_{\mathrm{EV_{back}}}$ on $\mathrm{P_{EV}}$ also had a stronger modulation of $\mathrm{EV_{back}}$ on the Congruency RT effect. To obtain subject specific effect of Congruency on RT we added $\gamma_{1k}\mathrm{Congruency}$ and $\gamma_{2k}\mathrm{Congruency}\mathrm{EV}_{\mathrm{back}_t}$ to the RT model (Eq. 2), representing subject-specific slopes of Congruency for subject k and for the interaction of Congruency and $\mathrm{EV_{back}}$, respectively. The subject-specific correlation of $\mathrm{P_{EV}}$ and $\mathrm{P_{EV_{back}}}$ was estimated by using only trials in which $\mathrm{EV} \neq \mathrm{EV_{back}}$. Probabilities were multinomial logit transformed and correlations were Fisher z-transformed (see above) before averaging across trials to achieve one correlation value per subject. In the main text and in Fig 5e-f we did not average the data to achieve maximum sensitivity to trial-wise variations. The results reported in the main text replicate when running the same procedure while averaging the data across nuisance regressors following the multinomial logit transformation ($R = .38$, $p = .023$). To extract subject-specific slopes for the effect of $\mathrm{EV_{back}}$ on $\mathrm{P_{EV}}$ we included a term for this effect ($\gamma_{1k}\mathrm{EV}_{\mathrm{back}_t}$) in Eq. 7, but due to convergence issues during model fitting, we had to drop the subject-specific intercept (γ_{0k}) in that model. Similarly, to extract subject-specific slopes for the effect of $\mathrm{P}^{2D}_{\mathrm{EV_{back}}}$ on $\mathrm{P_{EV}}$ we included a term for this effect ($\gamma_{1k}P^{k,2D}_{t,\mathrm{EV_{back}}}$) in Eq. 8.

Data availability statement

Behavioral data can be found in https://git.mpib-berlin.mpg.de/moneta/parallelrepresentation. All individual fMRI datasets can be found at https://gin.g-node.org/nirmoneta/SODIVA and are shared under Creative Commons Attribution-ShareAlike 4.0 International Public License (see LICENSE file in repository). We supply the fMRI data needed to reproduce the findings presented in the manuscript, i.e. conventionally preprocessed data (fmriprep, Esteban, Blair, et al., 2018; Esteban, Markiewicz, et al., 2018) from the functionally defined vmPFC ROI (smoothed at 4mm and 8mm, in MNI and native space). We additionally share data from various steps of the analyses: defaced T1 images, functionally defined ROIs in MNI and individual native

space, preprocessed data ready to be classified including individual classifier decoding results, individual RSAs (see README in https://git.mpib-berlin.mpg.de/moneta/parallelrepresentation for full details on the data folder structure). In case of interest in the whole brain raw data, please contact the corresponding authors. Source data are provided with this paper.

Code availability statement

Custom code for the task, behavioral analyses, preprocessing of fMRI data as well as fMRI analyses to reproduce the findings presented in the manuscript have been deposited in https://git.mpib-berlin.mpg.de/moneta/parallelrepresentation under Creative Commons Attribution-ShareAlike 4.0 International Public License (see LICENSE file in repository).

Author contributions

The following list of author contributions is based on the CRediT taxonomy (for details: Brand et al., 2015). NM, NWS contributed to conceptualization, formal analysis, funding acquisition, Methodology, project administration, software, validation, visualization and writing the original draft, reviewing and editing. NWS supervised the project and provided the Resources. NM contributed to Data curating and Investigation. MMG and HRH consulted at numerous steps of the planning, analysis and writing of the manuscript.

Competing interests

MMG is employee of Aya Technologies Ltd. The rest of the authors declare no competing interests.

Acknowledgments

NWS was funded by an Independent Max Planck Research Group grant awarded by the Max Planck Society (M.TN.A.BILD0004) and a Starting Grant from the European Union (ERC-2019-StG REPLAY-852669) and the Federal Ministry of Education and Research (BMBF) and the Free and Hanseatic City of Hamburg under the Excellence Strategy of the Federal Government and the Länder. NM was funded by and is grateful for a scholarship from the Ernst Ludwig Ehrlich Studienwerk (ELES) and Einstein Center for Neuroscience (ECN) Berlin

throughout this study. We thank Angela J. Langdon for comments on the manuscript. We thank Gregor Caregnato for help with participant recruitment, Anika Löwe, Lena Maria Krippner, Sonali Beckmann and Nadine Taube for help with data acquisition, all participants for their participation and the Neurocode lab for numerous contributions and help throughout this project.

Project 2: Reward morphs non-spatial cognitive maps in humans

Work with Luianta Verra, Charley M. Wu, Christian F. Doeller & Nicolas W. Schuck.

This work is not published and has not been peer reviewed. The project is ongoing and therefore results presented are preliminary.

Introduction

Understanding how humans and animals navigate and represent their spatial environment has been a central question in cognitive neuroscience (Doeller et al., 2010; O'Keefe, 1976). Following the observation that in pursuit of reward, rats make unexperienced shortcuts in complex mazes, Tolman coined the term 'cognitive map' as an internal representation of the world (Tolman et al., 1946; Tolman & Honzik, 1930b), for review see (Behrens et al., 2018). Decades of research have revealed a set of medial temporal and medial prefrontal spatial representations, such as hippocampal place cells which fire at specific locations of the map (Moser et al., 2008; O'Keefe & Dostrovsky, 1971) and entorhinal grid cells that fire at different locations, forming a hexagonal grid (Doeller et al., 2010; Gardner et al., 2022; Hafting et al., 2005). Alongside place and grid cells, more cell 'types' were discovered in the hippocampal formation, encoding goals (Sarel et al., 2017) and even locations of rewards (Gauthier & Tank, 2018). Together, these cells are seen as forming a coordinate system, incorporating knowledge from the environment for spatial navigation (Behrens et al., 2018; Jacobs et al., 2013; Sosa & Giocomo, 2021).

Recent work suggests that this system is not restricted to navigating in physical space and is also engaged when imagining navigation (Bellmund et al., 2016; Horner et al., 2016; Neupane et al., 2024) and also represents distances and trajectories between more abstract concepts and dimensions, such as visual features (Constantinescu et al., 2016; Theves et al., 2019, 2020), odors (Bao et al., 2019), sounds (Aronov et al., 2017), choice probabilities and magnitudes (Bongioanni et al., 2021) and even decisions along two-dimensional social hierarchies (Park et al., 2020, 2021). These findings suggest that representations of cognitive maps in this system might go beyond physical spatial navigation. Cognitive maps, in this broader sense, allow for the spatial and non-spatial organization of experiences and knowledge and play a crucial role in goal-directed behavior (Behrens et al., 2018; Bellmund et al., 2018; Epstein et al., 2017; O'Keefe, 1978; Schuck et al., 2016; Sharpe et al., 2019; Sosa & Giocomo, 2021; Stachenfeld et al., 2017; Tolman, 1948; Wilson et al., 2014).

To guide goal-directed behaviors in changing environments, cognitive maps must be dynamically

adaptive. Studies have shown that receptive fields of place and grid cells adapt to changes in the shape of the environment (Grieves et al., 2021; Krupic et al., 2015, 2018; O'Keefe & Burgess, 1996) and to salient locations within it (Boccara et al., 2019; Butler et al., 2019; Derdikman et al., 2009; Sanguinetti-Scheck & Brecht, 2020), even causing grid firing patterns to become irregular or 'distorted' (Ginosar et al., 2021; Grieves et al., 2021; Krupic et al., 2014; Sosa & Giocomo, 2021). Correspondingly, it has been suggested that the role of this system might not be to provide an objective metric of space, spatial or otherwise, but rather to reflect a more subjective, experience-driven representation (Behrens et al., 2018; Ginosar et al., 2023; Sosa & Giocomo, 2021; Stachenfeld et al., 2017). One main aspect of this subjective nature is that cognitive maps can over- or under-represent certain regions, for instance via changes in field density. These changes might help guide our perception of the environment and following behaviors (Aurelio et al., 2021; Bellmund et al., 2020; Ginosar et al., 2023; Theves et al., 2024). For example, areas of the map with an increase in field density might be represented in higher resolution, contributing to higher acuity, and making it easier to distinguish among the different locations there. Recent work suggests that objects in areas believed to have less spacing between grid fields (equivalent to the higher density of grid fields) are subjectively judged to be further apart relative to their true distance (Bellmund et al., 2020). This suggests that when field density increases, perceived distances increase with it. Having a more dynamic, experience-driven representation can thus help us achieve different goals more effectively, and changes in field density might be one potential account for such a mechanism.

The ability to predict reward locations is crucial for goal-directed behavior (Samuelson, 1947; Silver et al., 2021; Sutton & Barto, 1998). Understanding where and how value expectations are represented in the brain has been a long-standing goal of cognitive neuroscience research (Ballesta et al., 2020; Bartra et al., 2013; Clithero & Rangel, 2014; Moneta et al., 2023, 2024; O'Doherty et al., 2001). What role do values play in cognitive maps? One way to think of values is as another dimension along which information is organized (Moneta et al., 2024) such that value might simply be another task dimension (Bongioanni et al., 2021; Nitsch et al., 2024).

Indeed, areas in the medial temporal lobe, associated with representing cognitive maps, are also involved in representing rewards (LeGates et al., 2018; Wirth et al., 2009; Zeithamova et al., 2018). During navigation toward reward, cells in the hippocampus were found to increase their activity when approaching reward (Eichenbaum et al., 1987), with a firing pattern that matches subjectively anticipated outcomes, depending on recent experience (H. Lee et al., 2012). This suggests that during reward foraging, the hippocampus might represent a predictive value map, with an increased signal closer to the reward location. Values might also cause broader influences on the representation of the map. After exposure to reward, a proportion of place cells in the hippocampus re-allocate their receptive fields, a phenomenon known as remapping, to locate around the reward location (Hollup et al., 2001). This type of remapping was observed especially when the reward location is to be inferred and not when it was explicitly cued (Dupret et al., 2010). Similar changes in both firing rate and receptive field locations were observed in the entorhinal cortex (Boccara et al., 2019; Hollup et al., 2001; Sanguinetti-Scheck & Brecht, 2020; Sosa & Giocomo, 2021). Recent theoretical work suggests that these observed increases in representation around reward or other salient locations might indicate a more subjective, experience-driven representation in these regions, meant to provide a guide to future behaviors (Ginosar et al., 2021). Taken together, these findings imply that the representations of value and other elements of the cognitive map are interdependent, potentially influencing each other across varying temporal scales (Garvert et al., 2023; Moneta et al., 2023, 2024).

The goal of this study is therefore to investigate how reward causes long-lasting changes to the representation of an abstract cognitive map. By examining how reward affects perceptual discrimination between stimuli within the map and their neural representations, we hope to gain insights into the underlying neural mechanisms and their implications for behavior. Our first hypothesis was that exposure to reward will increase the perceptual discrimination between previously rewarding stimuli, accompanied by a systematic change such that a clustering of neural representational fields will form in the rewarded area.

One important question this hypothesis raises is where the increase in density comes from. Some

work has shown that clustering of representational fields in one area might come at the expense of other areas of the map (Schaffner et al., 2023; Thalmann et al., 2024). It remains unclear if this reallocation of fields is a result of a random allocation of resources or a systematic change to the representation of the environment. Based on previous theoretical work (Ginosar et al., 2023; Schaffner et al., 2023), we predicted that the increase in density around the reward would come at the expense of the immediately surrounding areas. More specifically, we expected that the increase in field density would be a result of a scaled pull of fields towards the reward, such that fields closer to the reward have a higher chance to be pulled towards the reward (henceforth: Pulling hypothesis).

We also considered alternative hypotheses such as partial remapping, i.e. re-distribution of a random selection of fields around the reward, or global remapping of all fields such that activity is scaled as a function of distance to reward. As discussed below, the latter is equivalent to a generalized reward predictive map. Such value representations are often found in mOFC (Moneta et al., 2024) and were also previously reported in the hippocampus during spatial navigation towards reward (e.g. H. Lee et al., 2012). Importantly, these different accounts are not mutually exclusive as some brain regions might be more focused on representing value maps whereas others might focus on representing other aspects of the map, such as (subjective) distances. We simulated all three hypotheses and found that although all three make the same prediction regarding the clustering of fields around the reward, they differ with respect to the influence the reward has on areas around it as well as generalizable changes to the rest of the map. Beyond the focus on rewarded areas of the map, our experimental design allowed us to focus on areas of the map that were never rewarded and to test these potential generalization effects. Preliminary univariate and multivariate fMRI analyses show support for the pulling hypothesis. Future analysis is planned to better understand the representational change and identify regions that might match these different hypotheses.

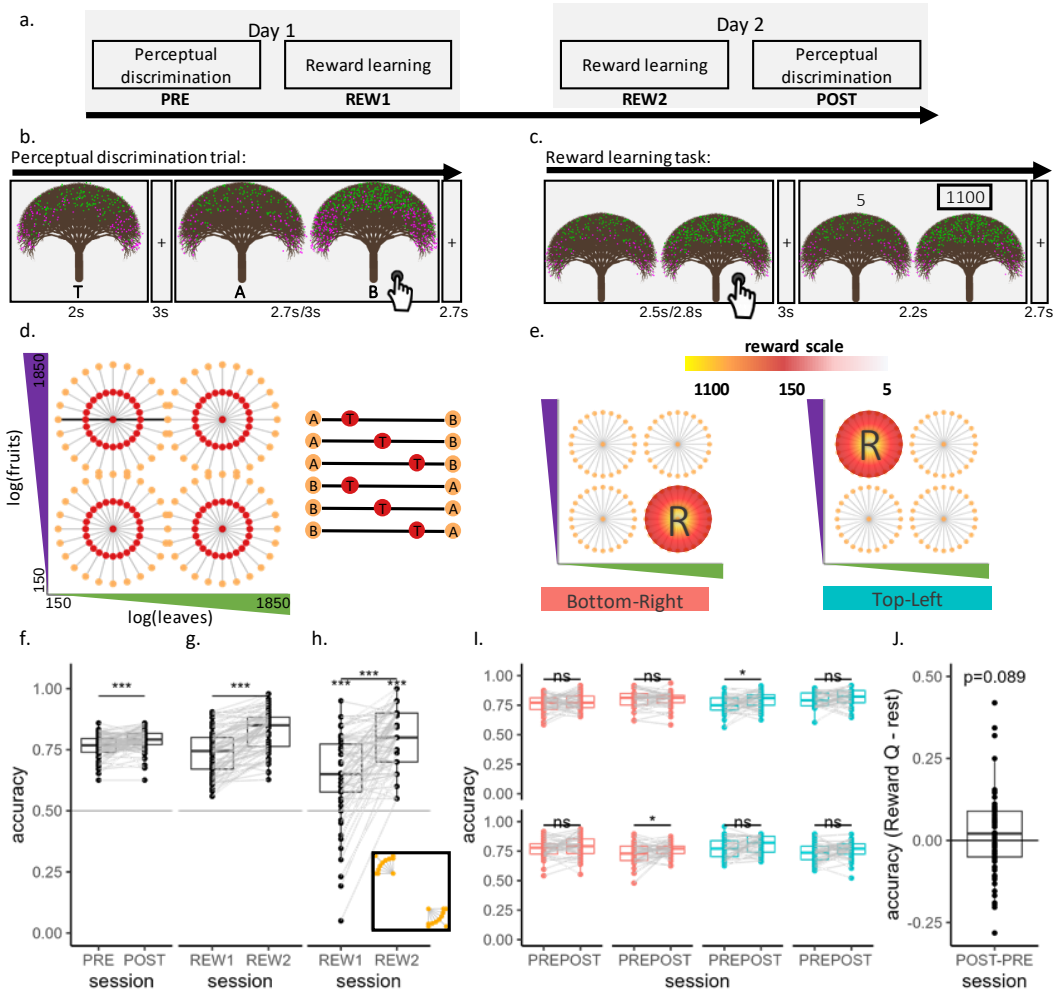


Figure 7

Task and Design a. On day 1, the perceptual discrimination task (PRE) was followed by a reward learning session (REW1). On day 2, the reward learning session (REW2) was followed by a repetition of the perceptual task (POST).**b.** In the perceptual task, a target tree (T, 2s) was followed by a fixation cross (μ=3s [1.8s-7s]) and two reference trees ('A' left, 'B' right, 2.7s/3s for behavioral/MRI). The task was to select a reference tree more similar to the target. Another fixation cross appeared before the next trial (μ=2.7s [1.8s-9s]). **c.**) In reward learning (REW1 & REW2), participants chose between two trees (2.7s/3s for behavioral/MRI), followed by a fixation (μ=3s [1.8s-7s]), then the same trees reappeared with outcomes (2.2s). A fixation cross appeared before the next trial (μ=2.7s [1.8s-9s]). Trees were the same size in both tasks. **d.** Trees can be mapped by the number of leaves (x-axis) and fruits (y-axis), sampled mainly in four quadrants, each split by 12 lines at 15-degree intervals (gray lines). Reference trees (A & B) were at line edges with target trees (T) between them. Each line represents 6 trials with A and B at the edges and T at the one-quarter, midpoint, or three-quarter point between them. **e.** In reward learning, participants were split by the rewarding tree's location: top left (Top-Left group, teal) or bottom right (Bottom-Right). The center tree of the rewarded quadrant gave 1100 points, surrounding trees 150, and others 5. Choices were mainly between the center and a surrounding tree (orange points) from four quadrants with a slightly smaller diameter (85%. see design section in methods). **f.** Participants performed above chance in both perceptual sessions and showed improvement from PRE to POST (all ps<.001). **g.** Participants performed above chance in both reward sessions and showed improvement from the first to the second (all ps<.001). **h.** Participants also performed above chance in each session and showed improvement when only comparing the most rewarding tree and trees more towards the corner (trials depicted in black square, all ps<.001). **i.** Only the rewarded quadrant showed a significant increase in accuracy from PRE to POST (Bottom-right group in bottom-right quadrant: p=.017, Top-left in top-left quadrant:p=.03, all the rest: p>.05) **j.** Participants showed marginal improvement in accuracy in the rewarded quadrant compared to all other quadrants ($p = .089$).

Participants, tasks, and design

A total of 74 participants (46 women, 1 diverse, $\mu_{age} = 26.1, \sigma_{age} = 4.69$, see participants section in methods for full exclusion criteria) completed four experimental sessions across two days (Fig.7a), 38 of which performed four sessions in an MR scanner (25 women, $\mu_{age} = 26.84, \sigma_{age} = 4.96$) and 36 participants performed the same task outside the scanner (21 women, 1 diverse $\mu_{age} = 25.38, \sigma_{age} = 4.33$). Across two consecutive days, participants first performed one session of a perceptual discrimination task at the beginning of the experiment (PRE), followed by a reward learning task that involved the same stimuli (REW1 & REW2), and another identical perceptual task (POST) (Fig.7b-c). Our main focus is on how the experience of reward changed behavior and neural stimulus representations while mainly focusing on the difference between PRE to POST sessions.

In the main **perceptual discrimination task** (Fig.7b), participants made perceptual similarity judgments by indicating which of two reference trees is more similar to a target tree shown immediately prior. Each tree was characterized by the number of leaves (green dots) and fruits (purple dots). These features varied independently and hence formed a 2-dimensional 'map' of a leaf/fruit space (Fig.7d). In anticipation of the non-linear perception of quantities, based on pilot data and as stipulated by the Weber-Fechner law, the trees were distributed in a logarithmic manner. Importantly, our models can account for any participant-specific variation in the logging of the space (see below and Fig.S14). Therefore for simplicity, we continue presenting the trees in logged space. Within this log space, trees were mainly sampled from four circles that covered the four quadrants of the map such that the target and reference trees were always along a straight line along the quadrant with the target tree either positioned at either the one-quarter, midpoint, or three-quarter point between the reference trees (see Fig.7d and Fig.S14). There was no location-based reward for this task. However, to motivate participants, they were told that if they were correct in over 70% of all trials in the 1st and 4th sessions they would receive an additional monetary bonus. Feedback was only given at the end of the block and only showed the mean accuracy of that specific block.

In the **reward learning task** (2nd and 3rd sessions, REW1 & REW2), participants performed a two-forced choice task and learned through trial and error to associate a specific set of trees from a circumscribed area within the 2-dimensional tree space with a reward (Fig.7c), i.e. some fruit & leaf proportions were rewarded while others were not, akin to introducing a reward to a location in a 2D cognitive map. Participants started the task with no knowledge of the reward location and were only told that similar trees yield similar rewards. To ensure that any change between the PRE- and POST sessions would be reward-specific, we split our participants into two groups which determined which quadrant would be rewarding: either the top-left or the bottom-right (henceforth: Top-Left, Bottom-Right groups, Fig.7e.). The reward was fixed for both reward sessions. Also here, we mainly sampled from four quadrants of the space such that the center tree of the rewarding quadrant was most rewarding (1100 pts. +-5), trees surrounding it were less rewarding (110 pts +-5) but still more rewarding than the the rest of the trees (5 pts +-5). Participants were told that points would be translated into monetary rewards. To further localize the effect of the reward, the sampled trees from each quadrant were more centered than in the perceptual task.

Importantly, throughout each session of each task, participants were exposed equally to the entire map, preventing over-exposure to any area. There were also never 'free-navigation' trials, i.e. participants could not control which trees would be shown. More details on the task and design can be found in Fig.16 in the methods section.

Results

behavioral results

General performance

Participants performed well above chance in both PRE and POST perceptual sessions in choosing the tree closest to the target tree (PRE: μ=.765,σ=.046, t-test against chance: $t_{(73)} = 48.9$, p<.001, POST: μ=.787,σ=.044, t-test against chance: $t_{(73)} = 56$, p<.001, Fig.7f, equidistant trees in logarithmic space were removed). Performance increased from PRE to POST session (paired t-test: $t_{(73)} = 3.9$, p<.001) Participants also learned with high accuracy to identify the most

rewarding tree in the reward task (REW1: μ=.736,σ=.092, t-test against chance: $t_{(73)} = 22$, p<.001, REW2: μ=.83,σ=.082, t-test against chance: $t_{(73)} = 34.6$, p<.001, Fig.7g) and significantly improved from the 2nd to the 3rd sessions (paired t-test: $t_{(73)} = 11.05$, p<.001, Fig.7g). To ensure participants learned the reward depended on a specific amount of fruits and leaves and not a generalized rule such as 'more leaves/fruits are more rewarding', we included trials in which participants decided between the most rewarding tree and trees with the maximum or minimum amount of fruits/leaves(Fig.7h, low). Participants performed above chance and showed improvement also in these trials, indicating they learned the reward is at a specific location of the 2D map (REW1: μ=.643,σ=.17, t-test against chance: $t_{(73)} = 7.27$, p<.001, REW2: μ=.8,σ=.11, t-test against chance: $t_{(73)} = 21.58$, p<.001, paired t-test between reward sessions, REW2-REW1: μ=.15,σ=.17, $t_{(73)} = 7.78$, p<.001).

Increased accuracy in the rewarded quadrant. We hypothesized that exposure to reward would increase perceptual acuity in the POST task in the rewarding region. To test this, we investigated the accuracy improvement from PRE to POST in each quadrant using repeated measures ANOVA with group x session x quadrant as factors and subject as a random effect. Post hoc tests found that for each group, accuracy improved significantly only in their rewarded quadrant, but not in any other quadrant (PRE-POST change in the bottom right quadrant in the 'bottom right' group: μ=.038, σ=.016, $t_{(288)}$=2.41,p=.017, Top left in the 'top left' group: μ=.035, σ=0.016, $t_{(288)}$=2.18,p=.03, Bonferroni corrected, p>.05 for all other quadrants for both groups (see behavioral analysis methods for details). A direct contrast of the improvement in the rewarded quadrant against the mean improvement in all other quadrants (across groups), indicated a marginal increase ($t_{(73)}$=1.36,p=0.089, Fig.7j).

Behavioral model

These changes in accuracy show that exposure to reward increased participants' perceptual discrimination ability even in a subsequent task without localized reward. To ask which perceptual changes could underlie these changes in performance, we set up a logistic regression model of participants' choices that included several factors that could model possible perceptual

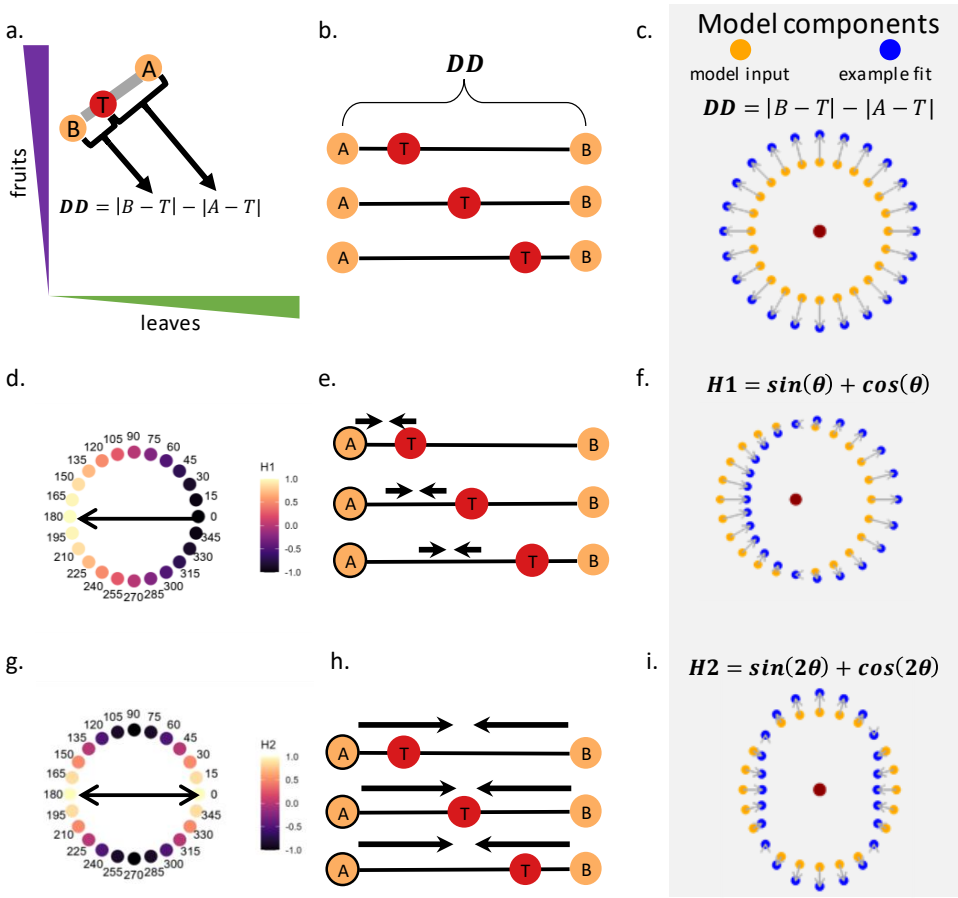


Figure 8

Computational model to capture choice bias in the cognitive map a. Distance of Distances (DD) is the difference in relative distance of each reference tree (A & B) to the target tree (T). When DD is positive, tree A is closer to the target, and when DD is negative, tree B is closer to the target. **b.** Depicted are three trials sharing the same reference trees and only differ in the location of the target. Scaling DD up or down does not change the correct choice (sign of DD), only how strong the evidence for it is. **c.** If when DD is high, choices become easier, then a potential interpretation is as if all trees are equally further apart from one another. Depicted is an example of model input (orange) and potential underlying perceived distances due to a fit with high DD (blue). **d.** Sinusoid wave with one peak (H_1). Adding this term to the model allows capturing angular-dependent choice bias. Depicted is an example where the bias is to choose trees pointing to the left. **e.** Same three trials from panel b. An effect of H_1 means a bias to choose the tree on the left (here, A), meaning the distance from A to the target is perceived as smaller, irrespective of whether the target is closer to it (first row), in the middle (second row) or closer to B (third row). **f.** An effect of H_1 indicates that distances between trees pointing in one direction to their respective targets are smaller compared to the trees pointing to the other side. One interpretation of such an effect is depicted similarly to panel d, showing an angle-dependent gradual change in distances equivalent to a positive effect of H_1 with a peak pointing left. **g.** Sinusoid wave with two peaks (H_2). This term allows us to capture axial choice bias. Depicted is an example where the bias is to choose trees pointing to the left or right over trees up and down. **h.** An effect of H_2 indicates that all the distances between the trees along one axis (here horizontal) are smaller compared to the orthogonal. **i.** One interpretation of H_2 effect is a compression of the respective quadrant along an axis, e.g. making choices along the horizontal axes harder compared to choices along the vertical axes. Note that in the model, both H_1 and H_2 had free phase and amplitude and could point to any direction.

biases and alterations. Specifically, we looked at the probability of choosing tree A over tree B, where tree A is always the tree presented on the left-hand side of the screen (including trials with a target presumed to be roughly in the middle between references, which were removed from the accuracy analysis above).

The goal of this analysis is to isolate different choice biases which could result from exposure to reward. We focused on two main components. The first was the relative distance of the distances of each reference tree to the target tree (distance of tree B to target minus distance of tree A to target, henceforth: DD, Fig.8a); When DD is positive, it means tree A is closer to the target and is the correct answer. When it is negative, it means tree B is correct. An increase in the general size of DD does not influence which is the correct response (set by its sign) but simply indicates a higher amount of evidence since the similarity of the trees to the target is more different (Fig.8b). A high beta coefficient for DD hence means that participants were more sensitive to even small differences in which tree was more similar to the target (akin to perceiving them as more 'distant', see example in Fig.8c.). Including the interaction of DD with session, group, and quadrant allows us to compare group-specific changes in acuity from PRE to POST in specific quadrants.

We reasoned that reward would not only cause changes in the rewarded region but also generalize to the whole map, introducing perceptual differences in other, non-rewarded quadrants. A main feature of our perceptual decision-making task was the target and two reference trees always lay on a line (see Fig.7d) that pointed in a particular direction. To capture whether participants had choice biases for particular trials as a function of their direction, we added a directional regressor for each quadrant that could rotate in any single direction.This was implemented through a harmonic addition of a sine and a cosine regressor that modeled the phase and amplitude of a single sine wave in directional space (harmonic addition theorem, see behavioral methods). This 1-peaked directional regressor hence allowed us to capture a bias to choose trees at a specific direction *irrespective if they are the correct ones* (henceforth: 1st harmonic, or H_1, Fig.8d., Fig.8e for details). One interpretation of such choice bias is that trees pointing in one direction, are perceived closer to the center of the quadrant compared to trees pointing at the opposite direction

(Fig.8f). We also included a two peaked direction regressor (henceforth: 2nd harmonic, or H_2, Fig.8g-h.) to capture possibly more complex directional modulations of perceptual similarity judgments. An interpretation of such choice bias is depicted in Fig.8i, where all trees on the left and right to the center are perceived closer to the middle, compared to the trees on the top and bottom of the quadrant. The interpretation of this regressor is somewhat challenging since in practice, choices were always between A and B which were on a straight line. We believe its main role is complimentary to H_1, modulating its effect in some direction, but plan to investigate this further in the future.

We tested for an effect of these components on all the data from the perceptual tasks, first ignoring any interaction of session and group (Fig.8j). We started with a baseline model with only nuisance regressors for cohort (behavioral only or MRI), group, session, quadrant, and interaction of session and quadrant and added each component hierarchically. Adding DD and its interaction with quadrant both improved the fit, indicating that participants had different acuity in the different quadrants (likelihood-ratio test comparing baseline to model with main effect DD, $\chi^2_{(1)} = 9340.19$, $p < .001$; comparing DD model with model including $\mathrm{DD} \times Q$ interaction: $\chi^2_{(3)} = 63.295$, $p < .001$). Adding a main effect of H_1, and its interaction with quadrant improved the fit further (LR test when adding main effect H_1, $\chi^2_{(2)} = 403.461$, $p < .001$, then further adding $\mathrm{H}_1 \times Q$: $\chi^2_{(6)} = 897.254$, $p < .001$). The fit improved even further we we also modeled H_2 (LR tests: H_2, $\chi^2_{(2)} = 15.090$, $p < .001$, with $\mathrm{H}_2 \times Q$: $\chi^2_{(6)} = 34.058$, $p < .001$). Note that we found no evidence that DD interacted with any of the sinusoid waves (adding $\mathrm{DD} \times \mathrm{H}_1$ to the best model: $\chi^2_{(2)} = 1.1235$, $p = .57$, adding $\mathrm{DD} \times \mathrm{H}_2$ to the best model: $\chi^2_{(2)} = 3.56$, $p = .169$). Since the DD mainly indicates which tree is the correct one, the lack of interaction indicates that the effect of the waves was independent of which tree is correct. Similar results were found when fitting PRE or POST sessions separately (see Fig.S14).

Next, focusing on the main question of this project, we tested for an interaction of the components with the session and group. Since model complexity increases exponentially (multiple 4-way interactions), we focused on the general improvement including all possible interactions, and later

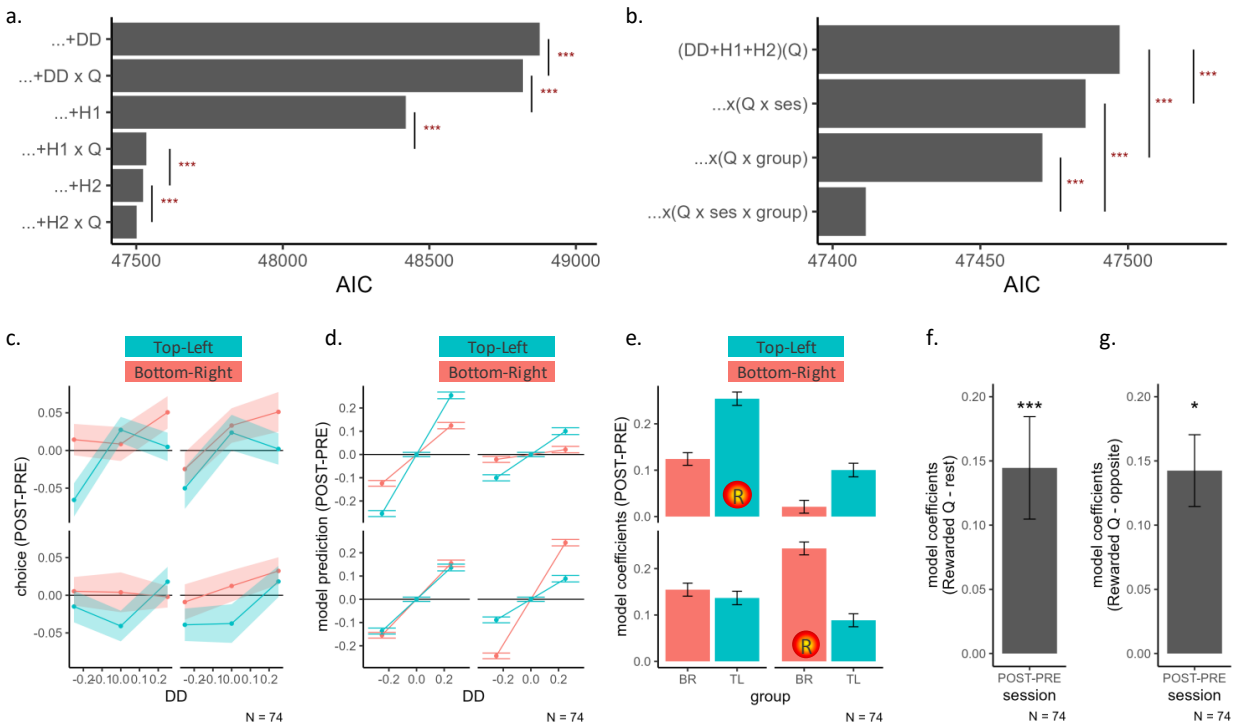


Figure 9

Reward-induced change in behavioral acuity a. Model comparison revealed that all the components had an effect on participants' choices and interacted with the quadrant, irrespective of session or group (all *p*s<.001). **b.** Including interaction with the session or group improved the fit, yet the best explaining model included all interactions with the session and group (all *p*s<.001). **c.** Averaged participants change in response to all trees in each quadrant from PRE to POST (POST-PRE, y-axis) with respect to DD (x-axis). Points and lines represent participants' mean responses and shades represent SEM. **d.** Model prediction equivalent to panel c. To show the marginal effect of DD, include only coefficients associated with DD. An increase in slope indicates an increase in acuity as even smaller DD would result in a higher choice probability. Each group (Top-Left, Bottom-Right) showed an increase in slope for the rewarded quadrant. **e.** The highest increase acuity, measured as an increase in coefficients associated with DD, was in the rewarded Quadrant. Bars indicate the mean effect of an increase in the rewarded quadrant of each group from PRE to POST compared to the mean effect of all the rest of the quadrants (*p*<.001) **f.** Following the same logic as panel g, contrasting specifically the rewarded quadrant with the opposite revealed a significant increase for both groups (Top left quadrant against bottom right quadrant for the Top-Left group, and the opposite for Bottom-right (*p*=.017).

turned to post-hoc tests for specific predictions generated from the experimental design. Adding an interaction of all components (DD, H_1, H_2) with either session or group improved the fit (LR-test: session $\times$ (DD + H_1 + H_2): $\chi^2_{(58)} = 51.532$, $p < .001$, group $\times$ (DD + H_1 + H_2): $\chi^2_{(2)} = 66.155$, $p < .001$). The interaction with the session hints at a potential general quadrant-specific change in choice biases from the PRE to POST session, independent of the reward location. The interaction with the group is less ideal, as it shows a potential session-independent group difference in perceiving the map. However, adding interaction with both group and session provided the best fit over only session interactions ($\chi^2_{(40)} = 154.4$, $p < .001$) or only group interactions ($\chi^2_{(40)} = 139.8$, $p < .001$). Together, these regressors help explain the differences between choices in the PRE vs POST session in a reward-group and quadrant-specific manner. Next, we will look at specific marginal effects to test our hypotheses of local and generalized changes in perceived distances due to reward exposure.

Exposure to reward increased local acuity. We first focused on the distance of distances (DD) which represents the difference in Euclidean distance between each reference tree and the target (for choosing tree A, Fig.8a-c). As mentioned above, an increase in the coefficient associated with DD does not influence which is the correct response, rather makes the trail easier (similar to a temperature parameter, Fig.9c-d.). We therefore interpret the effects associated with DD as proxy for all the perceived distances of each quadrant. Examining the relative change in DD's coefficients from PRE to POST revealed an increase in the respective rewarded quadrant of each group (Fig.9e). Contrast of the interaction effect using estimated marginal means (EMMs) with a z-test for significance revealed a specific higher increase in the rewarded quadrant from PRE to POST compared to all other quadrants (estimate $= 2.348$, $SE = 0.286$, p<0.001, Fig.9f). This effect holds even when specifically testing the rewarded quadrant against the quadrant opposite to it, with respect to the reward-group (estimate $= 0.485$, $SE = 0.204$, p=0.0176, Fig.9g). The latter shows an even clearer effect of our reward manipulation since for the top left group, the contrast takes the top-left quadrant minus the bottom right quadrant, and for the bottom right group exactly the opposite direction.

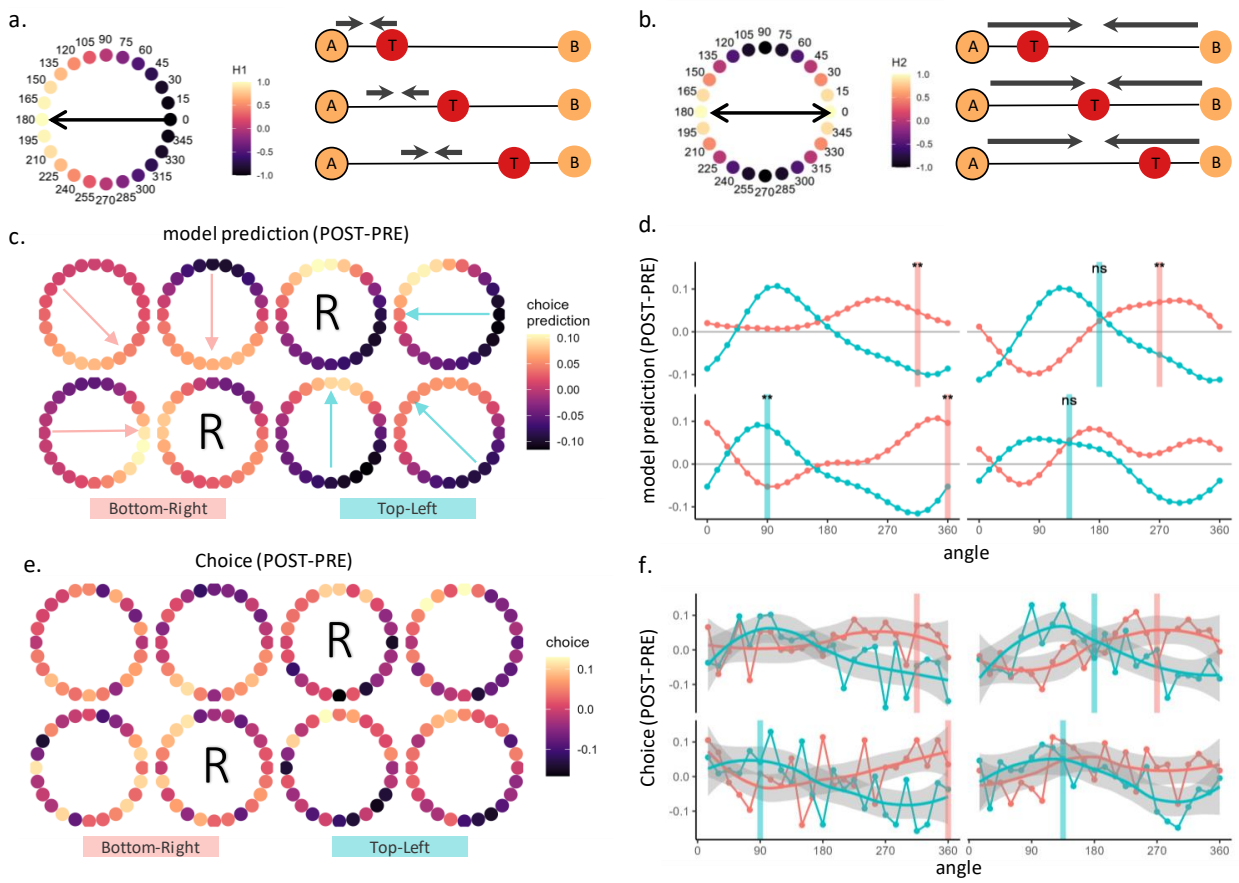


Figure 10

Generalized reward-induced change in behavioral acuity to non-rewarding areas a. Depiction of H_1 effect, taken from Fig.8. **b.** Depiction of H_2 effect, taken from Fig.8. **c.** Model prediction of change in choice bias from PRE to POST by group and quadrant. Color scheme represents model choice prediction change from PRE to POST. Black 'R' marks the quadrant that was rewarded between the sessions for each group and arrows in the other quadrants point at the angular direction for the reward. **d.** Model prediction of change in choice bias from PRE to POST by group and quadrant, sorted by angle. y-axis represents model choice prediction change from PRE to POST and x-axis the angle of the quadrant, whereas vertical lines in each quadrant show the angular direction pointing towards the reward of each quadrant (equivalent to arrows in panel c). Note that for Top-left group there is no vertical line in the Top-left quadrant and the same for Bottom-right group. Stars represents significance of a z-test for effect contrast between the two groups at particular angles using estimated marginal means (see stats in main text). **e.** True data of participants, same as (c). N=74. **f.** True data of participants, same as (d). N=74. Points represent average response of all participants. Smoothed line represent locally estimated scatter-plot smoothing (LOESS) across all individual participants' averaged responses, used for visualization without assuming a specific parametric form.

Generalized effect of reward to neighboring quadrants. Next, to test for the generalized influence of reward on areas that were not previously rewarded, we focused on the sinusoid waves (H_1 & H_2) to capture angle-dependent choice biases (Fig.10a-b). By examining the effect of these angular regressors at the perceptual indifference point (setting DD to 0), we can test for choice effects that would be invisible in any investigation of accuracy. This analysis revealed a choice bias directed towards the reward in the neighboring quadrants, i.e. in quadrants south of the formerly rewarded quadrant, participants tended to judge the northern trees to be more similar to the target, and in quadrants to the west, they were biased towards the eastern trees (Fig.10c-d). This occurred (1) even though none of the trees in question were ever directly rewarded in the REW session and (2) in a manner that was specific to the POST session and the relative location of the quadrant with respect to the reward location (see Fig.10d, vertical lines, for which angles point towards the reward in a group and quadrant specific manner). A post-hoc analysis of PRE-POST emergence of choice biases at reward-directed angles revealed a significant change between the groups in most of the quadrants, indicating a generalized choice bias due to reward exposure (see Table1 below). Lastly, we examined the averaged choices of participants and applied locally estimated scatter-plot smoothing (LOESS, without assuming a specific parametric form) across all participants' averaged responses and found qualitatively similar response patterns (Fig.10e-f). We note that in the right quadrants, the effects diverge in both model prediction and data, especially with respect to the top-left group. We plan to better understand this diversion in the future and hope the MRI data could shed some light on this.

Brought together, our behavioral models suggest reward increased acuity in the rewarded area. The influence of reward is generalized to neighboring quadrants, biasing choices towards it. Since during perceptual sessions participants' task was to indicate which of the reference trees had the least distance to the target, we reasoned that any bias in choice might be due to a generalized change in field density representing the map, accompanied by changes in perceived distances between the trees. An alternative interpretation is that participants carry over the reward task into the perceptual session such that they keep on following the rule of 'choose the most

Contrast	Quadrant	Angle (°)	Estimate	SE	z.ratio	p.value
Top-Left Quadrant	315	315	-0.569	0.198	-2.877	0.004
Top-Right Quadrant	180	180	0.076	0.205	0.371	0.710
Top-Right Quadrant	270	270	-0.557	0.208	-2.675	0.008
Bottom-Right Quadrant	135	135	-0.039	0.197	-0.196	0.845
Bottom-Left Quadrant	90	90	0.566	0.200	2.833	0.005
Bottom-Left Quadrant	360	360	-0.600	0.200	-2.998	0.003

Table 1
Post hoc tests of group difference in the change of choice bias from PRE to POST at specific angles pointing towards the reward of each group. For each quadrant, we tested the difference between the two groups in choice bias at angles pointing toward each of the two reward locations. Note that since the reward was centered in its respective quadrant, there is no reward-pointing angle for the Top Left group in the Top-Left quadrant and for the Bottom Right in the Bottom-Right quadrant. The reward-neighboring quadrant (Top-Right and Bottom-Left) was tested at two points: once pointing toward the reward location of each group.

rewarding tree'. However, this alternative can not explain the increase in performance in each group's rewarded quadrant. As discussed in more detail below, such a carry-over effect does not necessarily exclude a change in the representation of the cognitive map. In the next part, we ask what representational changes could accompany the observed choice biases.

Simulations of reward-induced representational change, univariate signals

To better interpret our behavioral results and generate hypotheses for the fMRI data, we first asked what would happen to a uniformly represented cognitive map after introducing a reward to it according to each of the three hypotheses described above: the pulling hypothesis, partial remapping, and global value-driven remapping. Each hypothesis provides a different mechanistic explanation for the clustering of fields following reward exposure and makes different predictions for the representational change in the map. We reasoned that a change in field density would correspond to a change in discriminability which would translate to a change in perceived distances (Bellmund et al., 2020), such that trees in areas of the map with a decrease in field density would be perceived more similar to one another.

We first simulated a simplified hypothetical representation for the cognitive map (both the x and y coordinates ranged from -1 to 1) by simulating 100,000 cells with 2D-Gaussian firing fields at random centers with random standard deviations ($\sigma \sim \mathcal{U}(0.01, 0.1)$), akin to hippocampus place cells showing firing patterns sensitive to specific locations in the cognitive map. We then tested

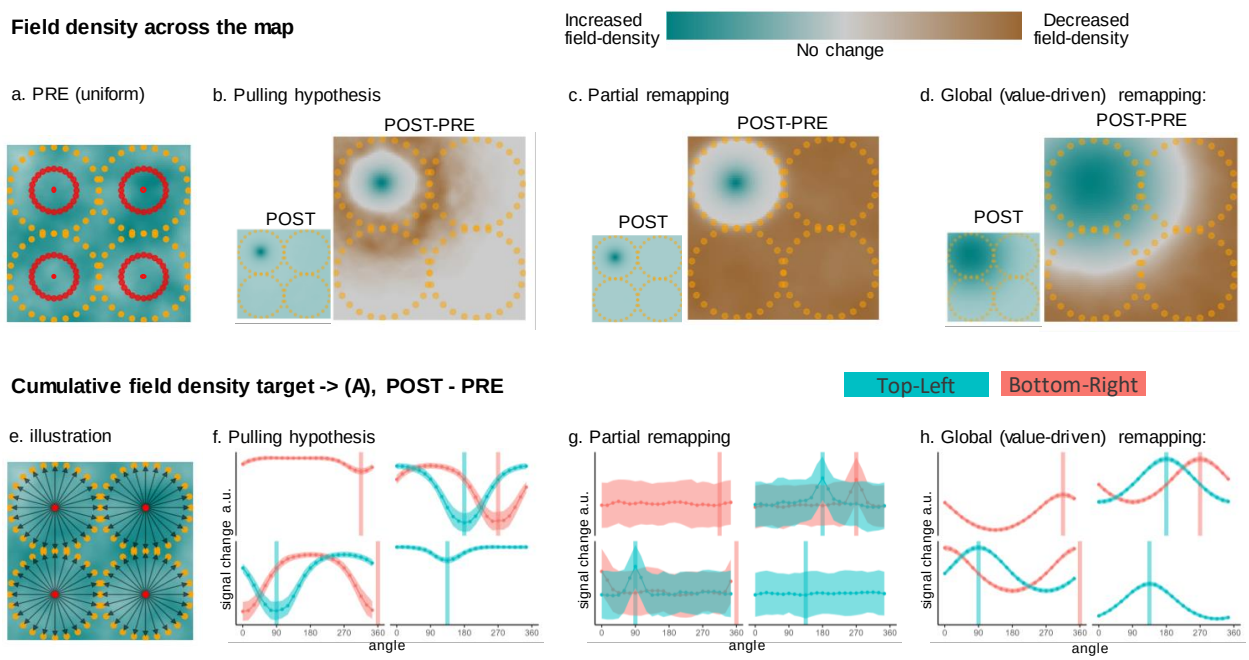


Figure 11

Simulation of univariate effects for place-cells-like representation of the cognitive map Legend on top depicts the color scheme throughout panels a-d, with teal indicating an increase in firing/field density, gray showing no change, and brown colors indicating a decrease in field density. Panels b-d show the predicted effect for the Top Left group only, symmetric effects were simulated for the Bottom-Right (maps not presented). **a.** Simulated random uniform activity map (PRE map). Orange points show the reference trees' location and red the targets of each quadrant. **b.** The pulling hypothesis predicts the pulling of fields towards the reward, scaled by their distance to the reward (POST map, bottom left). Contrasting the POST-PRE map shows that after scaled pulling of fields, the increase in field density in the rewarded area (teal) is accompanied by a decrease in neighboring areas (brown). **c.** Partial remapping of 20% of the cells to redistribute around the reward also predicts an increase in field density around the reward (teal) without prediction for change in other areas. **d.** Global remapping of all cells, scaled by the distance to the reward, also revealed an increase in field density around the reward location which stretches across the map, such that locations closer to the reward show an increase in field density (teal colors). The resulting map is equivalent to a generalized value-mapping where the firing rate scales with distance to reward. **e.** Panels f-h show cumulative activity from the center of each quadrant to a single reference tree, as indicated by black arrows on the depicted map. In all panels teal shows the top-left group and red shows the bottom-right. **f.** Pulling hypothesis predicted a decrease in field density (and univariate signal, y-axis) for trees at an angle pointing towards the reward (x-axis represents angles of quadrants presented in panel e). Each panel represents a quadrant and vertical lines the angles pointing towards the reward of each group. Shades represent +- standard deviation. **g.** Prediction per quadrant for the partial remapping, similar to panel f. No clear prediction for the non-rewarding quadrants. **h.** Prediction per quadrant for the global remapping, similar to panels f and g. Global remapping predicts an increase in signal pointing towards the reward in neighboring quadrants.

the sum of the activity of all cells at each location in the map, either without any change (PRE map) or after the predicted changes according to each hypothesis. We repeated this simulation 200 times. Our main motivation was to test which changes in field density could produce our behavioral results. As mentioned above, we expected that all three hypotheses would predict an increase in univariate signal at the rewarded location but differ in their prediction of the generalized effect on the rest of the map. Before any reward was introduced, the sum of activity was roughly uniform (Fig.11a), as expected.

Starting with the pulling hypothesis, we shifted the centers of all cells towards the reward, proportionally by their distance to the reward (similarly to Ginosar et al., 2023). The attraction force was calculated using an exponential decay function based on the distance from the reward location:

Attraction Force Formula: $F_{\text{attraction}}(r) = -\text{strength} \times \exp\left(-\frac{r}{\text{length}}\right)$, where r is the distance from the reward, strength is the force strength (set to 4), and length is a characteristic distance (set to 0.2).

We then contrasted the POST - PRE maps, showing the predicted change in field density across sessions. It is already visually apparent that under the pulling hypothesis, the increase in the rewarded quadrant comes at the cost of a decrease in field density in its surroundings (brown areas in Fig.11.b).

The first alternative hypothesis was that the increase in density was a result of randomly selected fields that were re-distributed to salient locations in the map (partial remapping). To simulate this, we randomly selected 20% of the cells and re-distributed them around the reward (new location variation: $\sigma = 0.15$). Since cells were randomly selected, this did not result in a visible change to any other quadrants (Fig.11c) and therefore predicts no reward-induced behavioral change in neighboring quadrants.

A second alternative is that all cells would remap, scaled by their distance to the reward (global remapping). In practice, this would mean that the firing rate in such a map would be higher for locations closer to the reward, akin to a generalized value mapping across the space. To simulate

this, we re-distributed all cells using a multivariate normal distribution centered at the reward location (new location variation: $\sigma = 0.5$). Contrasting POST-PRE maps revealed a gradual increase in field density towards the reward across the map (Fig.11d.).

By definition, all hypotheses predict an increase in signal in the rewarded quadrant. To generate behavioral predictions in the neighboring quadrants, we summed the activity of all cells along trials-trajectories from the center of each quadrant (red points, Fig.11e) to one of the references (orange points, Fig.11e), approximating the number of fields activated between the two. As mentioned above, a decrease in field density could correspond to a decrease in perceived distance. The pulling hypothesis predicts an angular-dependent *decrease* in field density pointing towards the reward in all non-rewarded quadrants (Fig.11f). The partial remapping did not make stable predictions for angle-dependent change in field density in the neighboring quadrants (Fig.11g). However, the global remapping (or value-mapping) made the exact opposite prediction compared to the pulling hypothesis. This alternative predicts a generalized effect of *an increase* in field density at non-rewarded quadrants pointing towards the reward (Fig.11h). We note that depending on allowed variation in the re-distribution of cell centers under partial remapping, a similar effect to Global remapping might appear (see tendency in Fig.11g), but an angular dependent decrease is very unlikely (see discussion for details).

Our behavioral results indicated that distances on the side pointing towards the reward were perceived as smaller (Fig.10). Although formal testing of the simulations is still to follow, this effect corresponds to the same angular decrease in field density according to the pulling hypothesis. Our behavioral results therefore show support for the pulling hypothesis as the representational change to the cognitive map due to reward. In the next section, we test our hypotheses regarding the change in univariate signal due to reward exposure. Our simulations of the pulling hypothesis suggest an increase in univariate signals in the rewarded quadrant accompanied by an angle-dependent decrease in signal in the neighboring quadrants such that trees pointing towards the reward would show the biggest decrease in univariate signal.

As mentioned before, behavior alone can not fully rule out that the effect is driven by participants

still performing the reward task during the perceptual task. In the case of such task confusion, we might expect co-activation of both a uniform perceptual map and a value-predictive map, i.e. a map where the signal is scaled by distance to reward, equivalent to the global (value-driven) remapping map in Fig.11d. Therefore we still tested for evidence for value-map representation in the next sections.

Preliminary fMRI univariate results

To test for the effect of reward on the neural representation of the cognitive map we set a GLM with one regressor for each unique target tree and one regressor for each unique reference tree (ignoring their side on the screen, see GLM section in methods). In the following analyses, we only focus on the betas of the target tree and their signal in three a priory defined regions of interest (ROIs, see ROI section in methods and Fig.12a.). Since our hypotheses focused on cognitive map representations, we looked into the Hippocampus (HP), medial Orbitofrontal cortex (mOFC), and Visual cortex (visual). We are currently focusing on these three pre-defined ROIs and plan more detailed whole-brain analyses in the future.

We first tested for an increase in signal in the rewarded quadrant compared to the opposite quadrant. Since the reward was located on opposite quadrants of the map for the two groups, the most controlled test would be to compare the rewarded quadrant to its opposite as those point at exact opposite quadrants for each group. We then set up a contrast to test the relative increase in signal in the rewarded quadrant from PRE to POST compared to the opposite quadrant ($Q_{r_{post}} - Q_{r_{pre}} > Q_{opposite_{post}} - Q_{opposite_{pre}}$). When averaging across all trials, we did not find a significant increase in the mean signal (12b.). Nevertheless, we wanted to test if an increase in mean signal in rewarded areas related to performance in the same location. Therefore, we correlated across subjects the increase of accuracy in the rewarded quadrant compared to its opposite with the increase in mean signal. We found a significant correlation in the Hippocampus and Visual cortex and not in mOFC (one-sided Spearman's rank correlation test: HP: $p = .027$, Visual: $\rho = .36$, $p = .017$, mOFC: $\rho = .32$, Fig.12c). Future work is planned to further investigate the potential mean signal change in the rewarded quadrant (also, see RSA below).

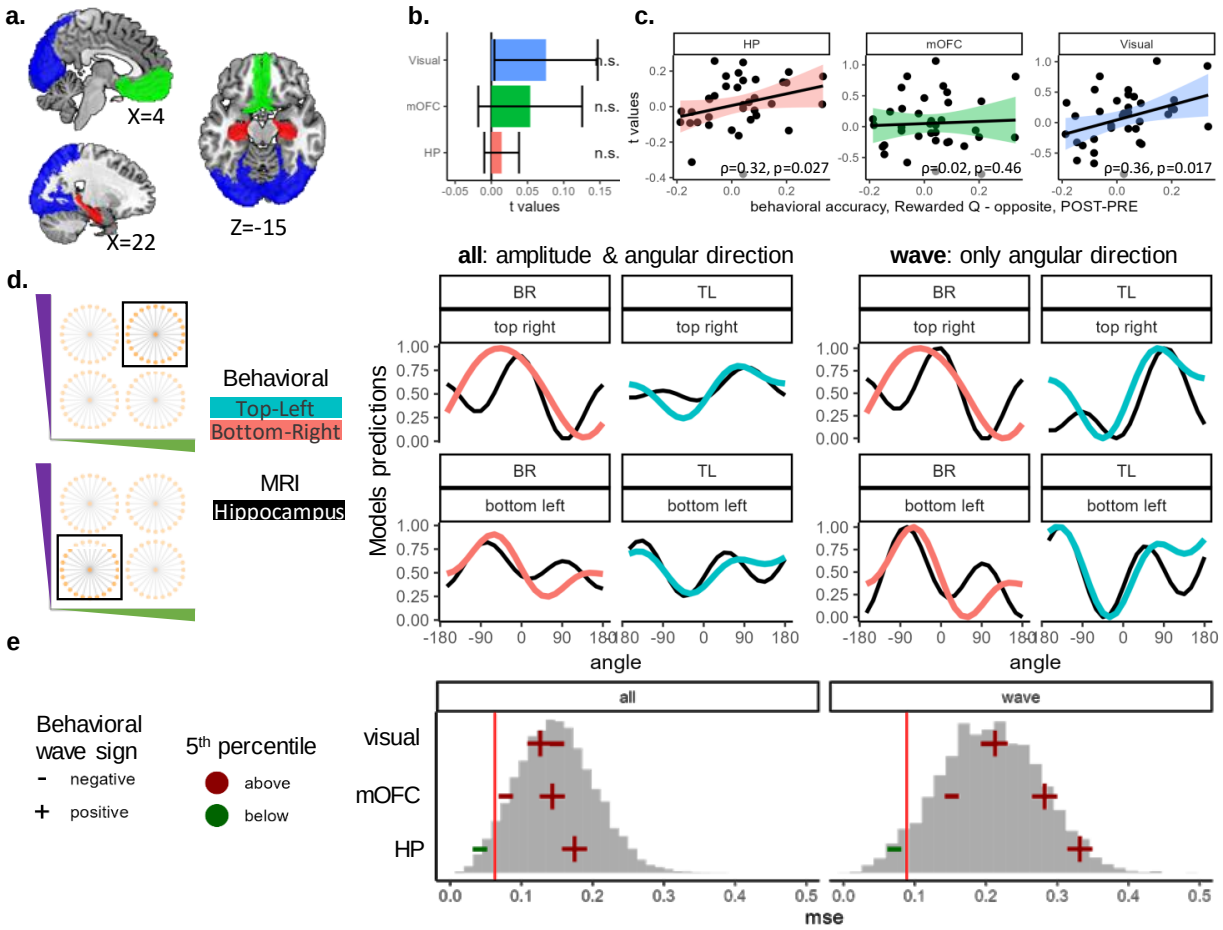


Figure 12

Preliminary univariate results support the pulling hypothesis a. Depiction of the different ROIs used: Visual cortex (blue), mOFC (green), and Hippocampus (red). Coordinates are in the MNI space. See the ROI section in Methods for the feature selection. **b.** We found no univariate increase of signal in the rewarded quadrant compared to the opposite from PRE to POST in any of the ROIs. **c.** Participants who showed a higher increase in accuracy from PRE to POST in the rewarded quadrant compared to the opposite also showed an increase in HP and Visual signal (p=0.027, p=0.017, respectively). No equivalent effects were found in mOFC. **d.** Model group prediction of angular dependent change in signal only in the neighboring quadrants (left), according to two normalization approaches. Middle panels show the normalization of all four waves together which keeps the amplitude different between the quadrants. The right panels show normalization by wave, ignoring amplitude and focusing only on the angular differences. **e.** Testing the Mean Square Error between the waves across normalization methods and ROIs showed the HP signal matches the negatively signed behavioral model, just as the pulling hypothesis predicted. No other tested region showed equivalent effects.

Next, we tested for generalizable univariate effects on neighboring quadrants. Our behavioral results suggest a choice bias pointing towards the reward and simulations showed a predicted drop in signal to accompany such a choice bias (pulling hypothesis). The opposite pattern was predicted by the global remapping (or carry-over value-mapping). This resulted in two potential MRI predictions based on the behavioral data, either an increase or a decrease in signal pointing to the same direction, only differed by their sign. Partial remapping made no generalizable predictions to neighboring quadrants. To test for these effects we took the mean activity during each target tree presentation from the above-mentioned GLM and fitted an equivalent model to the behavioral model for data coming from Hippocampus, medial OFC, or Visual cortex. This model included the same regressors for H_1 and H_2 for angular effects, without any effect for DD since the latter depends on the reference trees which were presented only after the target. We then took the cumulative coefficients of group, session, quadrant, and angle to generate a predicted mean signal for each ROI. Contrasting the PRE and POST resulted in a single wave of predicted signal change for each quadrant and group for each ROI. We compared this predicted signal to the predictions of the behavioral models, once positive and once negative, corresponding to an increase or decrease signal. We focused only on the neighboring quadrants (Fig.12d, left) since according to our simulations, those should have the strongest generalization and testable differences between the two groups. In Fig.12d we show the predicted waves for the hippocampus (black line) and the *negative* signed behavioral prediction, adhering to the prediction of the pulling hypothesis (colored lines). Since choice and MRI signal predictions are on different scales, we needed to bring them to the same range. We tested two approaches for such normalization: either normalizing *all* waves together, emphasizing potential between-quadrants differences in amplitude and phase (Fig.12d, left), or normalizing each *wave* separately, emphasizing the angular direction and ignoring between-quadrants amplitude differences (Fig.12d, right). To formally test for correspondence between the behavioral and fMRI predictions, we computed the Mean Square Error (MSE) between the MRI and the two opposite behavioral predictions (positive or negative angular change) and compared it to a permutation test. We found that regardless of

the normalization method, the MSE of the hippocampus signal and the *negative* behavioral signal was significantly below the 5th percentile of the permutation, indicating the hippocampus showed a drop of signal pointing towards the reward in neighboring quadrants, just as the pulling hypothesis predicts (Fig.12e). No equivalent results were found in the mOFC or the Visual cortex. Brought together, these show preliminary evidence for an increase in univariate signal in the rewarded quadrant for participants who showed a stronger increase in accuracy in the rewarded quadrant, suggesting a link between an increased signal in the hippocampus and an increase in accuracy. Furthermore, the signal in the hippocampus in reward-neighboring quadrants showed an angular decrease in signal pointing toward the reward, adhering to the predictions of the pulling hypothesis.

Simulations of reward-induced representational change, multivariate signals

Next, we tested for potential multivariate signal change as a result of systematic change in field density. The main goal of this simulation is to show what happens to multivariate dissimilarities when field density changes. To answer this question, we focused on one-dimensional representational space which simplifies visualization and helps avoid high computational demands. The underlying neural mechanisms that govern place cell activity are similar across different spatial dimensions (e.g. Sosa et al., 2024; Tessereau et al., 2025), suggesting that any systematic changes observed in one-dimensional simulations should be similar in two-dimensional maps. Intuitively, it is as if we took a single line from the 2D map and tested changes in dissimilarities between different positions on it (Fig.13a black dashed line).

We simulated 1000 cells with the same Gaussian activity and random standard deviations as in the univariate simulation with each cell center at equidistance along a line. Then, we either pulled the cell centers towards a single ’rewarding’ point, remapped a random 20% of the cell centers around the reward (partial remapping), or remapped all cell centers to normally distribute around the reward, causing the firing rate to be scaled by distance to reward location (global remapping or value mapping, Fig.13a), using the same parameters as before. We repeated this analysis 200 times to allow variations in sampling of the cells’ standard deviations and selection of remapped

cells.

We first looked at the change in univariate signal according to each hypothesis, similar to the 2D univariate simulation. This resulted in an increase of field density (and univariate signal) around the reward for all hypotheses, a decrease in the nearby surrounding of the reward for the pulling hypothesis, a constant decrease over distance to reward for the global remapping (or: value mapping) and only very local change for the partial remapping (Fig.13b).

We then looked at the multivariate dissimilarities of each pair of 100 points along the simulated line within each map and contrasted it with the dissimilarities from the PRE map, equivalent to testing changes of within-session dissimilarities from PRE to POST sessions (Fig.13c). We looked at both Pearson and Euclidean distances since they differ in their sensitivity to mean differences between the patterns (Pearson is mean-insensitive, e.g. Walther et al., 2016). Visually inspecting these dissimilarity matrices shows that all hypotheses predict that representations of all locations within each side of the reward would generally tend to get more dissimilar to one another (bright yellow colors in Fig.13c). The pulling hypothesis, however, makes a unique prediction of *decreased* dissimilarities across both sides of the reward, specifically for locations very close to it (dark blue shades in Fig.13c). To better understand and visualize the simulation results, we examined four specific locations and their change of dissimilarity to the rest of the map: the reward location, a location *close* to the reward, a mid-distance location chosen where we would expect a drop in univariate signal (and perceived distances) according to the pulling hypothesis and a fourth point located far from the reward (vertical lines in Fig.13b and panels in Fig.13d). As expected by the representation of a continuous map, all points showed a decrease in dissimilarity for points close to them (see 'dips' in dissimilarity in every subplot of panel Fig.13d).

There is still a lot to unpack from these simulations as well as formally test their predictions. A systematic test across different potential parameters is still lacking (see discussion for details). For now, we focus on one finding regarding dissimilarities in the immediate proximity of the reward. The pulling hypothesis, unlike the global remapping, makes a unique prediction of a decrease in dissimilarities for locations on opposite sides of the reward (Fig.13c-d). Such a decrease might be

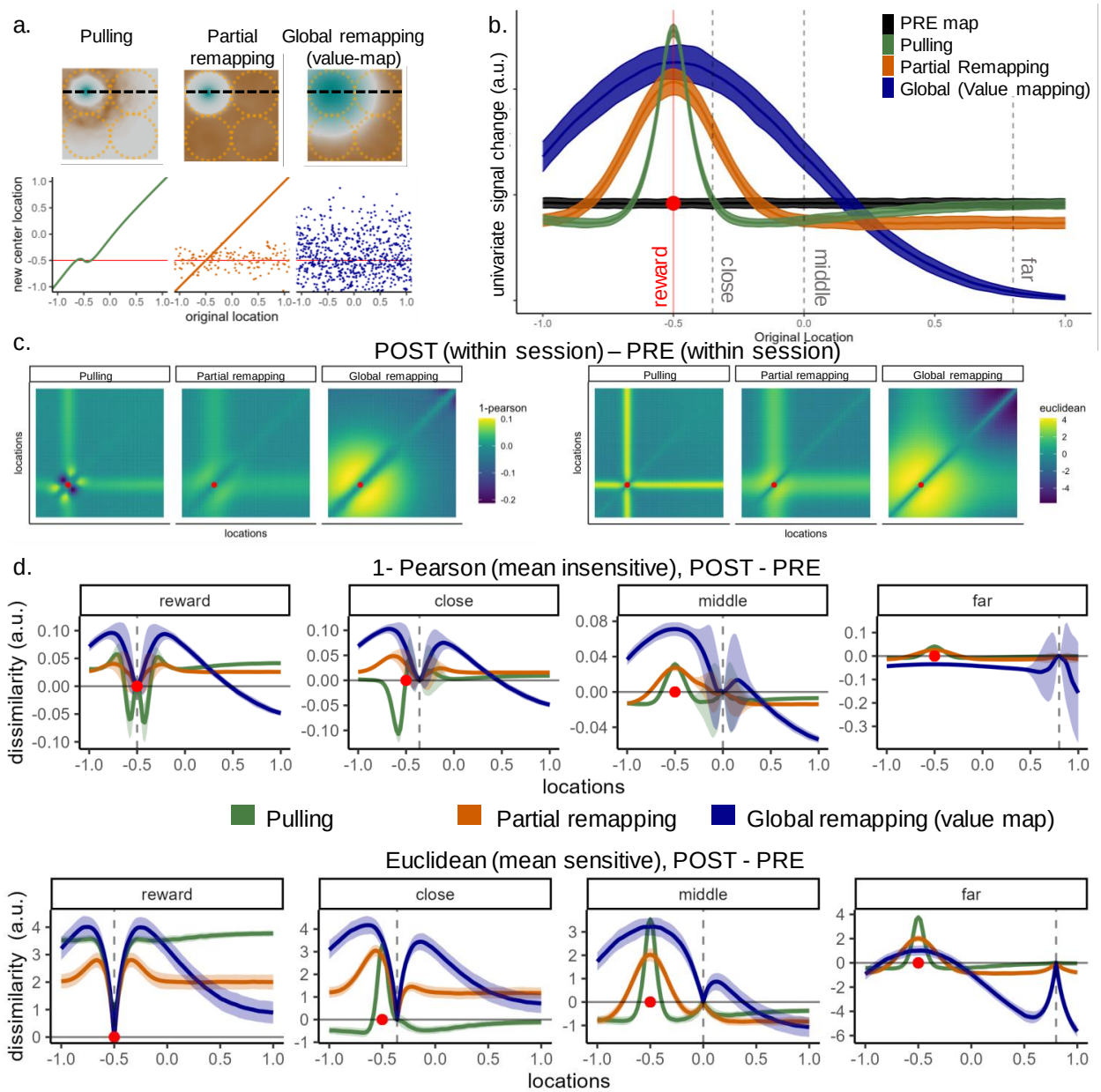


Figure 13

simulation of multivariate effects for place-cells-like representations a. The three hypotheses for representational change due to reward. The first row shows the previously presented 2D prediction and the second depicts equivalent 1D predictions according to each hypothesis. The pulling hypothesis (green) shows cells on each side of the reward became gradually closer to it, scaled by their distance. Partial remapping (orange) shows a random set of 20% cells allocated around the reward and global remapping (value-mapping, blue) shows a full remap of all cells normally distributed around the reward. Ribbons represent standard deviation across simulations. **b.** Mean *univariate* signal across all simulations at each location shows equivalent results as the 2D simulation. Vertical lines represent the selected locations depicted in panel d. **c.** Full dissimilarity matrix of all location pairs. Left: 1-Pearson, mean-insensitive. Each panel shows the prediction of a different hypothesis contrasted to the unchanged map (equivalent to the POST-PRE test). Mean-sensitive Euclidean is depicted on the right. The red point marks the 'reward' location. **d.** Mean dissimilarity of selected locations (panel b) and all other locations on the map. Ribbons represent standard deviation across simulations. 1-Pearson (first row) shows the strongest effect of decreased dissimilarity for locations on opposite sides of the reward.

the result of the strong pull of fields around this area, causing fields to be pulled closely together. If these cells have large receptive fields (i.e. simulated standard deviation), the multivariate neural signal might become less distinct since the overlap is too big there. Most importantly, the ’close’ locations are *not* the same locations where there is the biggest decrease in field density, i.e. univariate signal becomes negative (’middle’ point in Fig.13b.). Rather these are locations much more adjacent to the reward (’close’ point in Fig.13b.). Behaviorally, we saw the effect of a decrease in field density in the neighboring quadrants. This suggests that the observed effect in the simulation here might be more localized, potentially mainly within the rewarded quadrant and not in the neighboring ones. The decrease in dissimilarities effect across the reward location appears in both 1-Pearson (mean-sensitive) as well as Euclidean (mean-insensitive) measures. Brought together, the multivariate simulation predicts that under the pulling hypothesis, trees located in the immediate surroundings of the reward would become more similar to one another. The global remapping (value map) makes the opposite prediction, namely an increase of dissimilarity across the sides of the reward region. Next, we turned to the MRI data to look for representational changes that could match the different hypothesized changes, currently only focusing on changes within the rewarded quadrant. We note that the variation in re-distribution under partial remapping could cause a similar effect to the pulling hypothesis within the rewarded quadrant. Therefore, we continue, for now, by comparing the predictions based on the pulling hypothesis against the global remapping (see discussion for details and plans for future work).

Preliminary multivariate fMRI results

Next, to test our multivariate predictions, we took the same betas from the above-mentioned GLM which corresponded to the signal during the presentation of every target tree and computed the multivariate dissimilarities of all trees within each quadrant, split by session and group (Fig.14a, see RSA section in methods for details). The above simulation predicted that a scaled pulling of fields towards the reward would result in a decrease in dissimilarities across the sides of the reward, but only in its immediate surroundings. Global remapping (or: value mapping) predicted an increase in dissimilarity for these comparisons (Fig.14b). Therefore we focus on dissimilarities

within the rewarded quadrant, which corresponds to trees *very* close to the peak reward location. Currently, we only test average dissimilarities and plan more detailed modeling work for the future. All tests below are two-sided to allow for an opposite effect due to the remapping hypotheses, although our hypothesis favored the pulling hypothesis.

We first focused on the predefined hippocampus ROI and, similar to the DD analyses, we averaged the dissimilarities across participants by session, group, and quadrant. Contrasting the rewarded quadrant against its opposite revealed a significant *decrease* in 1-Pearson dissimilarity, which matches the pulling hypothesis (two-sided t-test against 0: $t_{(35)} = -2.38$, p=.023, Fig.14c). No equivalent result was found for Euclidean distances ($t_{(35)} = .91$, p=.367, not presented). Since our simulations predicted that this effect should be quite local to the reward location, we asked if it relates to how accurate participants were in identifying the exact reward location. To measure this, we took two behavioral indicators from the *previous* reward learning sessions. We first took the general improvement in accuracy from the first to second reward session as an indirect measure of participants' learning curves across the days. We found that participants who showed a higher learning curve between the reward learning sessions also showed a stronger decrease in dissimilarity (spearman rank correlation test: ρ=$-.52$ p=.001, Fig.14d). The second indicator we took was the performance only on the reward session on the second day, as a proxy for more immediate, recent reward experience. Here, we found no relation to hippocampus representational change (spearman rank correlation test: ρ=$-.01$, p=.95, 14e).

Next, we performed the same analyses on the signal from the mOFC. Similar to the hippocampus, we found an indication for a marginal effect showing a *decrease* in dissimilarity from PRE to POST in the rewarded quadrant compared to its opposite, but only for Euclidean (mean sensitive multivariate measure, two-sided t-test against 0: $t_{(35)} = -1.72$, p=.09, Fig.14f) and not for 1-Pearson (mean-insensitive, $t_{(35)} = -1.48$, p=.145, not presented). We then compared this change in signal to the across-day learning curve and same-day immediate experience with reward. No relationship was found with the across-day learning curve (ρ=.11, p=0.52, Fig.14g). However, participants that showed a stronger decrease in Euclidean dissimilarities in the mOFC,

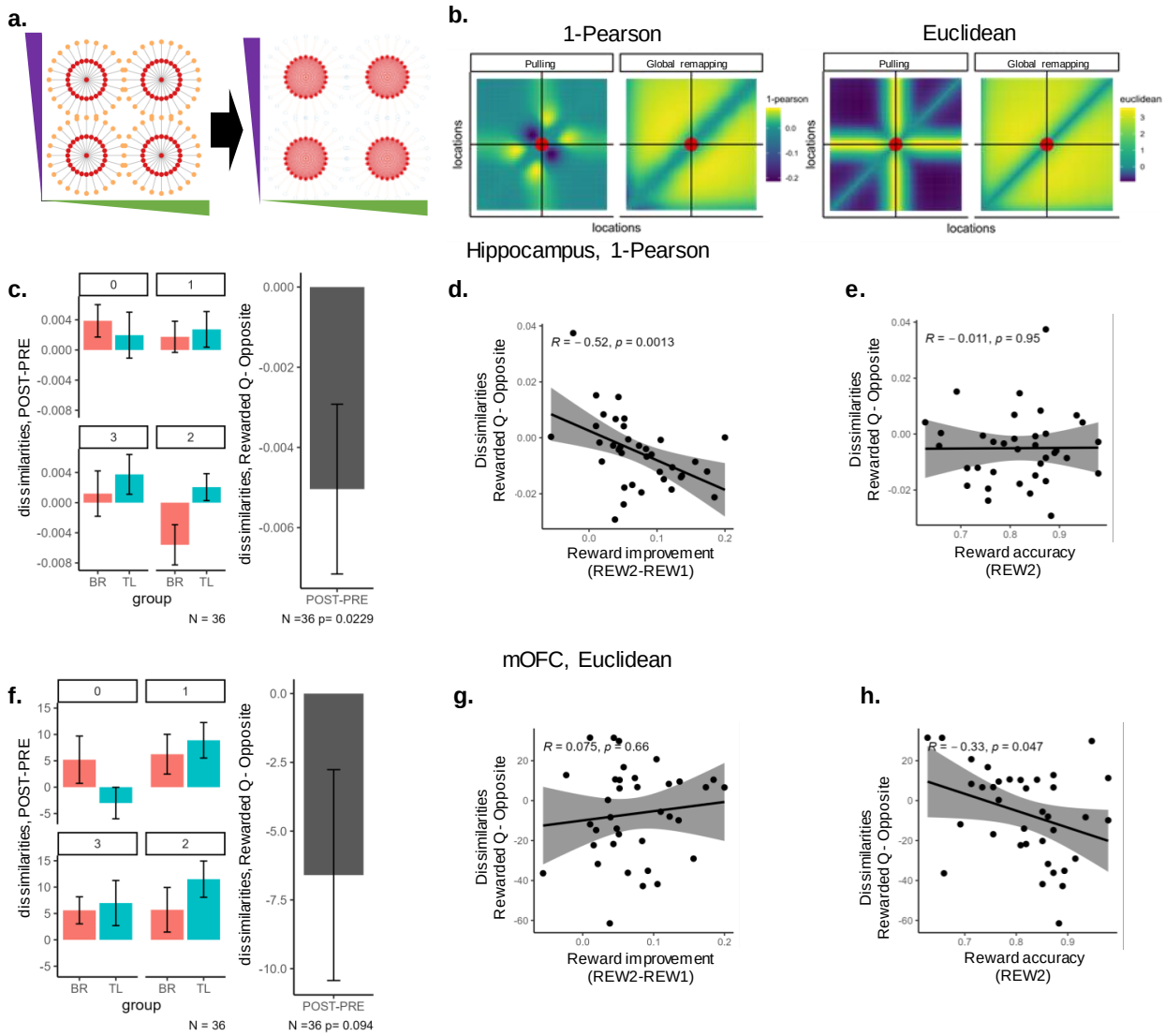


Figure 14

Preliminary multivariate results support the pulling hypothesis **a** We looked at the multivariate dissimilarity of all trees within each quadrant. The red lines on the right panel show all the pairs of trees used for this analysis. **b** Closely examining the predicted similarity in the *immediate* surroundings of the reward revealed a predicted decrease in dissimilarity on both sides of the reward. Such a decrease is only predicted by the pulling hypothesis. The red point represents the reward location. **c** Change in 1-Pearson dissimilarities (mean-insensitive) in the hippocampus from PRE to POST in all quadrants by group (left). Directly contrasting the rewarded quadrant against its opposite with respect to the group revealed a decrease in dissimilarities, just as the pulling hypothesis predicted (p=.043). Error bars represent the standard error of the mean. **d** Participants that showed a higher increase in accuracy from the first to second reward session (x-axis), also showed a stronger decrease in hippocampus dissimilarities (y-axis, p=.001). **e** No relationship was found between change in hippocampus dissimilarities (y-axis) and participants' performance on the second reward session (x-axis, p=.95). **f** Change in Euclidean dissimilarities (mean-sensitive) in the mOFC from PRE to POST in all quadrants by group (left). Directly contrasting the rewarded quadrant against its opposite with respect to the group revealed a decrease in dissimilarities, just as the pulling hypothesis predicted, p=.043. **g** Mo relation was found between participants' increase in accuracy from the first to second reward session (x-axis) and mOFC dissimilarities change (y-axis, p=.52). **h** Participants that performed better on the second reward session (x-axis) showed a higher decrease in mOFC dissimilarities (y-axis, p=.95). We note that due to technical reasons, we had to exclude two subjects from all analyses presented here (both in the BR group). We hope to solve this issue in the near future.

also performed better on the same day (ρ=.33, p=0.049, Fig.14h) hinting at a link between mOFC representational change and more recent experiences with reward.

Visual cortex showed no significant change in the rewarded quadrant compared to its opposite (1-Pearson: $t_{(35)} = -0.36$, p=0.97, Euclidean: $t_{(35)} = 1,03$, p=0.308), and no relation to learning curve over the days (1-Pearson: ρ=-0.02, p=0.93, Euclidean: ρ=-0.039, p=0.82). The only effect seen in the visual cortex is related to reward performance on the second day (similar to the mOFC), and only for 1-Pearson measures (ρ=-0.42, p=0.011) and not Euclidean (ρ=-0.17, p=0.31, similar to the hippocampus). Future work is needed to disentangle these results.

Brought together, these results further support the pulling hypothesis as an explanation for the representational change of a cognitive map as a result of reward learning. Initial evidence suggests that hippocampus representational change was sensitive to across-day improvement in reward performance whereas mOFC might be sensitive to more immediate recent experiences with reward. Future analyses are planned to dive deeper into these changes and look both at other regions as well as multivariate generalizable effects into neighboring quadrants.

Discussion

In this project, we asked which representational changes could be caused by reward exposure in an abstract cognitive map. Participants improved discriminability among trees in a previously rewarded area and showed a generalized choice bias toward the reward in areas that were never rewarded before. By splitting our participants into two groups with opposing reward locations, we ensured the observed changes were not due to general training in the task. Based on previous literature, we tested three potential representational changes that could account for the observed choice bias: a systematic pull of fields towards the reward, partial remapping of random cells around the reward, or global remapping where all cells remap such that the firing rate is scaled by distance to reward across the map, equivalent to value-mapping. Participants' behavior in the neighboring quadrants matched angular-dependent decrease in perceived similarities similar to the pulling hypothesis predictions. Preliminary findings from fMRI analyses provide further support for the pulling hypothesis as the guiding principle for reward-driven representational

changes in the hippocampus. First, although we could not find significant group-level evidence for an increase in mean signal, closer examination of between-subjects differences revealed that participants who showed a stronger increase in signal in the rewarded quadrant also showed better behavioral performance in the rewarded quadrant (Fig.15a). When looking at generalization to other quadrants, we found an angle-dependent decrease in signal in the reward-neighboring quadrants which was related to a behaviorally observed decrease in perceived distances (Fig.15b). Such angular decrease was uniquely predicted by the pulling hypothesis. Multivariate analyses revealed a decrease in neural dissimilarities in the hippocampus in the rewarded immediate surroundings on the group level, an effect which, on the participant level, related to across-days reward learning improvement (Fig.15a). All of these findings were predicted by univariate and multivariate simulations of representational changes according to the pulling hypothesis. mOFC signal during the target presentation did not show univariate increase, but a similar decrease in multivariate dissimilarities in the rewarded quadrant. Unlike the hippocampus, this effect was related to reward performance on the second reward session and not to cross-day learning. Brought together, behavioral and preliminary fMRI results show support of hippocampus representational fields pulled towards the reward, with some evidence suggesting a similar pattern in mOFC.

One alternative explanation to the behavioral findings is that participants ’carry over’ the reward session task itself, i.e. continuing to also answer ’which tree is closest to reward’. We can not fully rule out this explanation merely based on the behavior, but find it very unlikely for a few reasons. First, such rule confusion can not explain the increase in accuracy and discriminatability in the rewarded quadrant (accuracy effect Fig.7i-j, discriminatability Fig.9f-g). Second, there was a 40+ minute break between the last reward session and the perceptual task. Moreover, participants were told that if they were accurate in more than 70% of the trials in the perceptual task in both PRE and POST sessions, they would get an additional monetary bonus. From their perspective, if they kept on performing the reward task, it would harm their chances to gain this bonus. The core question of this project is to try to characterize the carry-over effect between the

tasks. If participants were following a rule of choosing the most (hypothetical) rewarding tree, then we might see evidence of value mapping which should be the optimal representation for such a task. We found no evidence that supports value mapping in the Hippocampus or the mOFC, at least during target presentation, although both have been previously found to represent predicted value signals in other tasks (e.g. H. Lee et al., 2012; Moneta et al., 2023). In the future, we hope to identify evidence for such value mapping during the reward session and test for evidence for the same mapping during the POST session.

Our simulations of both multivariate and univariate signals made clear opposite predictions for global remapping according to the value and pulling hypothesis. As mentioned above, the variance of the distribution of the new locations for the cells plays a crucial role in the partial remapping predictions. Univariately, increasing the variance would cause the map to be more similar to the global remapping. However, since cells are randomly chosen, there can not be an angular decrease in signal in the neighboring quadrants. This means that the univariate findings show specific support for the pulling hypothesis. Multivariately, however, a decrease in variance around the new location could potentially cause the same decrease in dissimilarities around the reward observed under the pulling hypothesis. This means that currently, the multivariate results do not clearly distinguish between the pulling hypothesis and the partial remapping. However, all results show evidence that it is unlikely that the behavior is driven by task confusion, as none of which show support of a co-activation of a value-map that would support such interpretation. In the future, we hope to find better markers for the different hypotheses. Specifically, similar to the univariate simulation, closer examination of the multivariate signal in the neighboring quadrants might help better distinguish the hypotheses. Furthermore, testing directly the similarity across the sessions (PRE -> POST), and not only the difference of similarities within sessions as presented here, might be the ideal test to show the pulling of fields. All of these analyses are planned for the near future.

One interesting and somewhat counter-intuitive finding is a decrease in neural dissimilarity in the hippocampus, in a similar area of the map where we saw an increase in behavioral performance.

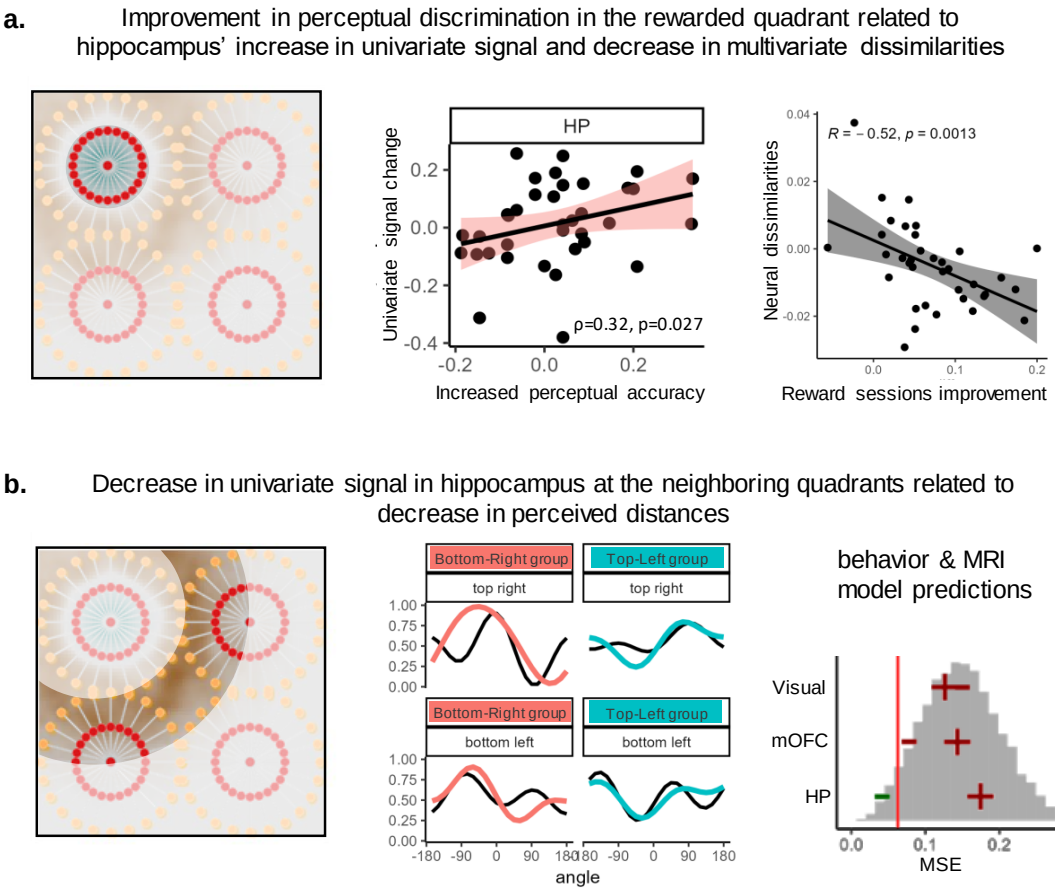


Figure 15

Summary of representational changes in the hippocampus. **Summary of representational changes in the hippocampus a.** The pulling hypothesis predicts an increase in field density in the rewarded quadrant (left). Focused on the change of neural signal during target tree presentation in the rewarded quadrant from PRE to POST (highlighted on the left) compared to the opposite quadrant, participants' increase in behavioral performance in the perceptual task (x-axis, middle), was related to increase in univariate increase in signal (y-axis, middle). Participants who showed a steeper learning curve between the reward sessions (x-axis, right), showed a stronger decrease in multivariate dissimilarities (y-axis, right). **b.** The pulling hypothesis predicted a decrease in field density (and univariate signal) in reward-neighboring quadrants (left). The decrease in univariate neural signal in the hippocampus (black lines, middle) was related to behavioral choice bias corresponding to decreased perceived distances (colored lines, middle). Lines are model predictions. On the right is the permutation test after normalizing all model predictions to be on the same scale.

If trees in the rewarded quadrant became more neurally similar to one another, how come there is an increase in accuracy and discriminability? One potential explanation could be the variation in localization of the reward effect. According to our simulation, it is areas *very close* to the reward that increased their similarities. One speculation would be, that if the effect of reward is very localized, then trees around the middle would become more similar, yet trees just slightly further away would become more distinct. In the RSA analyses, we took only the target trees, which form an 'inner' circle of the quadrant. Participants' choices, however, were along a much longer line across the 'outer' circle of the quadrant (Fig.15). This interpretation is supported by the fact that comparisons made in the rewarded session were between trees closer to the center of the quadrate than the reference trees (roughly 85%, see design in methods). It could be that the loss of discriminability was very localized, among the target trees, which participants never needed to choose among. Although this does not provide a full account of this discrepancy, we plan to investigate this complexity more thoroughly.

Related to this, the lack of a clear univariate signal change in the rewarded quadrant in the hippocampus is puzzling. One potential explanation is that in the transition from cells to voxels, the distribution of cells that moved did not translate to a strong enough univariate signal. Another possibility is that such an effect is present more at the early stages of the POST session and washes off over time. Examining the accuracy of the behavioral data also only revealed marginal effects. In fact, it was crucial to account for the full set of different perceptual biases (DD,H_1, and H_2) to find a clear and strong behavioral effect. In the future, models fitted to the fMRI data might better identify such effects. Lastly, as mentioned before, the univariate and multivariate analyses pointed at a different 'flip' point in the sign of the predicted signal, with multivariate predicting a flip closer to the reward. The location of such a flip might depend on the size of the receptive fields of the cells as well as on the slope of the attraction force applied according to the pulling hypothesis. Between-subject differences in these parameters might make a group effect harder to identify. We plan to address these challenges in the future.

We also did not find an increase in univariate signal in mOFC for trees closer to the reward,

although this region is often associated with representing univariate value signals (e.g. Moneta et al., 2024). There could be a few potential explanations. First, we only looked at the presentation of the target tree. There is some limited evidence that mOFC only represents values when they are actively relevant for choice, not when imagining potential outcomes (Castegnetti et al., 2021). Since there was no choice during the target presentation, value signals might have simply not been there. Another issue could be that our ROI is not sufficient, especially as it does not include the adjacent and overlapping vmPFC region. Lastly, it could also be that, at least in this task, mOFC does not represent value and simply shows a similar effect to the hippocampus. One big conundrum we plan to face in the near future is the finding that the hippocampus showed an effect in 1-Pearson measures whereas the mOFC showed a marginal effect in Euclidean. This difference between mean sensitive and insensitive measures might relate to the presence of value signals but requires more detailed analyses. Future modeling work will look into these explanations in closer detail.

Brought together, behavioral and preliminary fMRI results support the notion that reward increased perceptual discrimination around it, in a manner equivalent to pulling representational fields, scaled by their distance to it. Reward thus causes systematic and generalizable changes to an abstract cognitive map. This effect of reward in behavior as well as hippocampus was generalized both in time, to a following task where the reward was not present anymore, as well as in space, to areas of the map that were never rewarded before. Future analyses, including modeling of uni- and multivariate signals during target presentation as well as grid-like coding and potential reward-induced influences during choice are planned soon.

Methods

The study complies with all relevant ethical regulations and was approved by the ethics board of Deutsche Gesellschaft für Psychologie e.V. (DGPs) (Ref. Number: SchuckNicolas2022-01-14WV).

Participants

92 young neurotypical adults took part in the experiment (58 women, 1 diverse, $\mu_{age} = 26.3, \sigma_{age} = 4.54$, minimum age: 18, maximum age: 36) in exchange for monetary reimbursement. Out of which, 45 performed the task inside the MR scanner (MRI cohort: 30 women, $\mu_{age} = 27.1, \sigma_{age} = 4.8$), and 47 performed all sessions outside the scanner (28 women, 1 diverse $\mu_{age} = 25.5, \sigma_{age} = 4.2$). Participants were recruited using the participant database of the Max-Planck-Institute for Human Development.

For the MRI cohort, Beyond common MRI-safety-related exclusion criteria (e.g. piercings, pregnancy, large or circular tattoos, etc.), we also excluded participants if they had a head circumference larger than 58 cm (due to the limited size of the 32-channel head-coil). Across both cohorts, we also did not admit participants to the study if they reported any history of neurological disorders, tendency for back pain, or color perception deficiencies. The gender of participants was self-reported (note that the study was conducted in the German language where there is no clear distinction between sex and gender). We had no reason to suspect any gender differences in the task and therefore did not include this information in the analyses.

Task-performance exclusion criteria: we excluded participants based on several behavioral criteria:

First, 7 participants in the behavioral cohort and 2 participants in the MRI cohort did not perform in the last two reward learning blocks with accuracy above 60%. This blocks were crucial to make sure they learned the exact reward location. Second, one participant in each cohort had 30 or more no answer trials in both reward sessions as well as in both perceptual sessions. Participants were instructed that no answer would count as wrong answer for calculation of the rewards. Third, in the perceptual task, we interleaved special one-dimensional trials where the

target tree showed both leaves and fruits, but the following reference trees showed only leaves or only fruits. These were control trials to make sure participants fully processed both features. This reasoning was explicitly told by participants and they were trained before the main task specifically on these types of trials. Nevertheless, three participants in the behavioral cohort performed below 60% in those trials across perceptual sessions (in addition to one of the subjects with increased no-answer trials).

Altogether, we excluded 11 participants from the behavioral cohort and 3 participants from the MRI cohort. We note that the relatively large number of excluded participants in the behavioral cohort was mainly driven by a lack of reward learning. This motivated us to include an additional 'easy' block in the first reward session for the MRI cohort (see below) which seems to have improved performance. However, This increase could be that without the MRI total payment was lower (see below), lowering general motivation. Another potential reason is that due to technical reasons, we had limited control over the size of the screen inside the MRI scanner, resulting in potentially larger stimuli for this cohort.

MRI-based exclusion criteria: We excluded an additional 6 participants from the MRI cohort due to MRI-based criteria: Three participants exhibited excessive motion throughout at least one of the perceptual sessions (one of which also had a total of 130 no-answer trials across all sessions). Whereas on average, the rest of the group had 1.65 TRs across all sessions (out of 8522, $\sigma = 3.8$) where framewise displacement was above 2mm (full voxel) to any direction, these three subjects had 150, 231, and 247 TRs. Framewise displacement was measured by fmriprep, see below for details.

Although we tried to optimize the tilt angle of the field of view to minimize signal drop at both MTL and OFC, two additional subjects exhibited significant signal drop out at the entorhinal cortex and were therefore excluded from the analyses (one of which also did not pass the reward-learning threshold). To set the exclusion threshold, we took the bilateral ROI as defined according to the freesurfer-based output of fmriprep and excluded from it voxels that exhibited less than 50% of the mean signal of the whole time series (per block). We then took for each

subject and ROI the maximum drop in voxels for each session and averaged across sessions. We then excluded two participants for having more than 50% signal loss in the entorhinal cortex in native space. One additional participant was excluded since the MR scanner constantly crashed, resulting in 4 repetitions of the first block of the POST session in addition to too much data loss. The final sample of participants resulted in 74 participants (46 women, 1 diverse, $\mu_{age} = 26.1, \sigma_{age} = 4.69$) with 38 in the MRI cohort (25 women, $\mu_{age} = 26.84, \sigma_{age} = 4.96$) and 36 in the behavioral cohort (21 women, 1 diverse $\mu_{age} = 25.38, \sigma_{age} = 4.33$).

Experimental procedures

Stimuli. Tree generation was inspired by (Flesch et al., 2022). The two colors for leaves and fruits, as well as all colors used for the tree base (trunk and branches, shades of brown) had the same luminance to make sure more leaves and/or fruits coverage does not also increase luminance. The same tree base was used for all trees and it was symmetric on the left and right such that the shape of the tree should not influence or interact with the amount of leaves and fruits. We note that due to technical reasons beyond our control, the screen inside the MRI had a strange color gradient from one side to the other, making half the screen slightly darker. By design, the trees presented on each side of the screen were balanced across the session (see below), but this might hinder future analyses if they plan a very detailed examination of the visual cortex responses.

For all participants, fruits were distributed on the lower half of the branches (roughly mid-third of the tree) and leaves were on the top part. This separation was done to encourage a perception of both amounts separately. However, to prevent a too harsh separation, ca. 10% of the leaves/fruits in each tree were distributed on the opposite side (leaves below fruits above). Trees were generated at the beginning of each session. The same base tree was used for all participants. The total amount of leaves and the total amount of fruits were built of a subset of images overlayed on each other. The images used were a distribution of each prime number from 1 to the total sum needed. Location across images did not overlap. To make sure there is a clear difference between the amounts and a more holistic perception of the amounts (especially for example for the most

rewarding tree), we always added a small noise to the number of items on the tree (+-5 items). This addition, in combination with using prime-number templates, resulted in almost complete re-distribution of the items on the tree, not only between trees but also for repetition of the same number, e.g. tree with 510 leaves did not look exactly like the previous time 510 appeared, but also not exactly like 500 plus 10 items, rather every time is complete random locations.

Experimental design

After a general description of the experimental design, we will explain each part in detail (Fig.16a). The experimental sessions spanned over 2 consecutive days. On the first day, participants first performed a perceptual training session for the perceptual discrimination task. This session had mainly full one-feature trials, meaning the target tree and references had either only leaves or only fruits. This was done outside the MR scanner for all cohorts. Afterward, they performed the first session of the perceptual discrimination task, followed by the first session of reward learning. These two sessions took place in the MRI for the MRI cohort. On the second day participants first performed another reward learning task and then the second perceptual discrimination session, also inside the MRI for the MRI cohort. After the last session, participants performed a memory placement task to test their memory of the reward location (data currently not presented). Each of the sessions started with a short training and instructions which included explicit questions to confirm participants followed the instructions. Training also took place on the second day, to further ensure no task confusion. In all sessions and tasks, stimuli remained on screen for a fixed amount of time, irrespective of choice (i.e. did not disappear after choice). To give participants an indication that their choice was received, whenever a choice was required (second part of each trial), two small circles appeared on the left and right ends of the screen and disappeared once a choice was made. Between the two main sessions, there was always a break of 40-60 minutes in the MRI cohort and at least 20-30 minutes in the behavioral cohort.

General training session. For the perceptual training task (mainly 1D), participants completed 178 trials per context similar in structure to the main perceptual task, only that all targets and references had either only leaves or only fruits. The design for each context consisted of 58

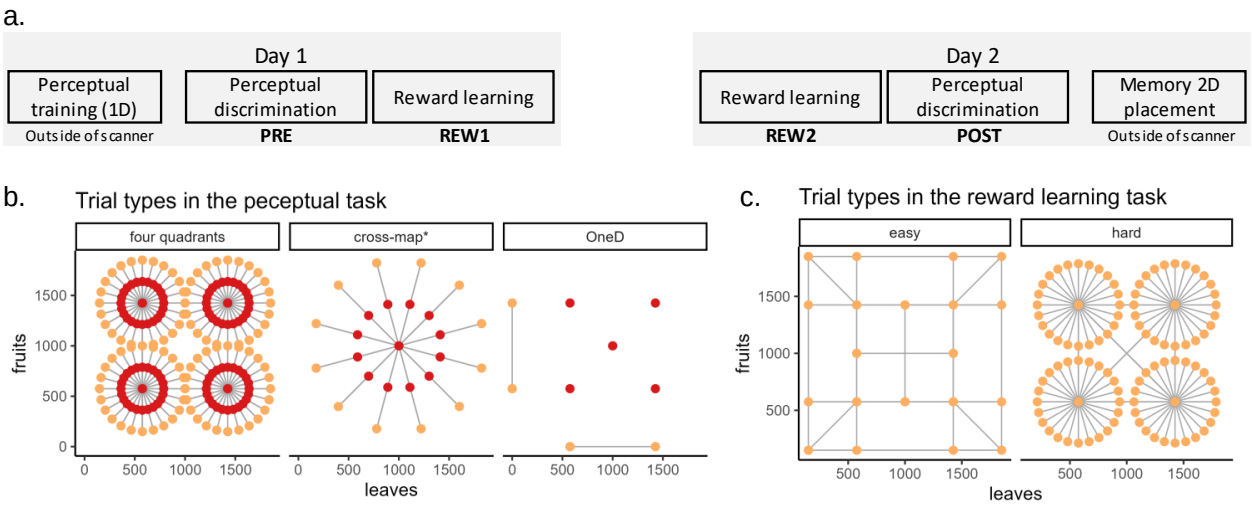


Figure 16

Trial types in the different tasks: a. Full experimental design. On the first day, participants first performed a perceptual training session outside the MR scanner. Then, they performed two consecutive sessions in the MR scanner: the perceptual discrimination task followed by reward learning. On the second day participants first performed another reward learning task and then the second perceptual session. After the last session, participants performed a memory placement task to test their memory of the reward location. The behavioral cohort did all sessions outside the scanner. **b.** In the perceptual discrimination task there were three trial types: trials drawn from the 4 quadrants (used for all main analyses), trials drawn from a bigger area across the map (cross-map), and OneD trials where the target tree had both leaves and fruits and the reference trees had either leaves or fruits. We note that all trials sampled from the big circle where the target was not in the middle (2/3rd) to 1D trials (dropped leaves or fruits from the reference trees, see methods for details). **c.** In the reward learning task, there were two types of blocks. Easy blocks (left) had trials with mainly large distances and mainly 90 degrees making it easier to learn the reward location. Hard blocks (right) had mainly trials where participants needed to compare the center of each quadrant to one of the trees in its surroundings, similar to the location of the reference trees, slightly closer to the center. In all blocks, we systematically sampled across the space to ensure equal exposure to the whole space.

standard trials, which were then flipped in position to control for motor responses. The 58 trials were consistent with 16 ”short distance” trials with items positioned close together, and 8 catch trials where the target was identical to one of the references. We also added 7 trials with very short distances where targets were presented in 2D, this was meant to start exposing participants to the integration of the axes together. Trials for each context were split into two blocks, resulting in four blocks in total. The starting context was pseudo-randomized across participants, balanced within each reward-location group. The trials were pseudo-randomized and balanced to make sure no hidden structure could bias participants’ choices. We made sure there was no more than one repetition of trials where the tree with the highest amount of items was on the same side, no

more than two consecutive trials choosing the same motor response, and no more than three trials with the target in the middle position (very hard trials), no more than two repetitions of the same decision difficulty level (e.g. short), and no more than three consecutive trials requiring selection of the highest (or lowest) value. Trials featuring 2D targets were spaced no more than 11 trials apart to maintain regular exposure.

Perceptual discrimination task (PRE & POST sessions). The main perceptual task (mainly 2D) included 360 trials divided into 6 blocks of 60 trials each. The spatial structure was geometrically organized with 4 peripheral ”circles” (or quadrants) containing stimuli arranged at 24 possible angle positions (15-degree increments) and one central circle with stimuli at 12 angle positions (30-degree increments, Fig.16b). Each trial presented three points forming a straight line with 5 possible positions along each line with the external two points always the references and the centered three acting as target trees. Trials were sampled in chunks of 9 trials: twice, one trial of each small quadrant in random order, followed by one trial from the bigger circle. This was done to encourage participants to orientate to the full map. Trials were carefully controlled to prevent repetition of the same angle, limit consecutive presentations of angles within the same 60-degree bin to no more than two, and restrict consecutive identical motor responses to no more than three. The design incorporated an additional 24 trials (4 per block) featuring 1D reference trees to make sure participants paid attention to both contexts separately while encoding the target tree. These trials were strategically positioned at least 20 trials apart. Lastly, we transferred all trials sampled from the big circle where the target was not in the middle (2/3rd) to 1D trials (changed only the reference). We made sure that each half of the experiment (three blocks) sampled all possible angles from the small quadrants, however balancing the motor (which side of the screen A or B are placed) was only possible across both halves. Currently, only trials from the 4 small quadrants are used in the analyses. In each trial, the duration of the target tree was 2s, and the reference trees for 2.7s for the behavioral cohort. Pilot data suggested reaction time is slightly slower in the scanner, which is why we increased the duration of the reference trees to 3s. The fixation between the target and reference trees was drawn from a truncated exponential

distribution with a mean of 3s and ranged between 1.8-7s. Similarly, the fixation after choice had a mean of 2.7s and ranged from 1.8-9s.).

Reward learning task (REW1 & REW2 sessions). The reward learning task had two components with different difficulty levels - ”easy” and ”hard” trials - balanced across 5-6 blocks (Fig.16c). The easy component (59 trials per block) featured grid-arranged stimuli with points at corners and centers, while the hard component (60 trials per block) used angle-based arrangements similar to the 2D perception task but with modified spacing from the center to centralize the reward effect (perceptual task had a radius of 426 in logged space and reward task was 362 which is 85%). Reward values were distributed according to a multivariate normal pattern centered at either the bottom-right or top-left quadrant center (counterbalanced between subjects), with reward values of 1100 (center of quadrant), 150 (surrounding trees in the same quadrant) or 5 (rest of the map). The design included several controls: no more than one repetition of non-zero reward differences, no more than three consecutive trials with zero reward difference, and no more than two consecutive trials with maximum reward on the same side. For the behavioral cohort (and the first two MRI participants) the second and third sessions each started with an easy block followed by 4 hard blocks. After observing that some participants struggled with learning the reward location, we introduced an additional easy block to the second reward session (easy-hard-hard-easy-hard-hard). Sampling of trails ensured that each half of the session (first and last pair of hard blocks) sampled the full space of potential angles for each quadrant. In each trial, the duration of the choice period tree was 2.5s for the behavioral pilot. After noticing some participants struggled with learning the reward location, we increased this timing to 2.8s for the MRI cohort. The outcome presentation duration was 2.2s. The fixation before the outcome and after it were the same as in the perceptual task (mean of 3s range of 1.8-7s and mean of 2.7s, range 1.8-9s, respectively)

Behavioral analyses

accuracy tests: For all accuracy tests in the PRE and POST sessions, we excluded trials where the target was presumed to be roughly in the midpoint between the references (in log space). Note

however that since the models below account for subjective distances, these included these trials and found equivalent and even stronger effects (see below and main text). For the reward sessions, we only included trials where the value difference between the trees was different than 0, in easy and hard blocks. We used repeated measures ANOVA to assess accuracy improvements from PRE to POST sessions, with group, session, and quadrant as factors, and subject as a random effect. Post hoc tests included estimated marginal means which were computed using the ‘emmeans‘ package (Lenth, 2025). The full result of the pairwise comparisons on session differences within each quadrant and group with Bonferroni correction to adjust for multiple comparisons revealed significant improvements only in the rewarded quadrants for each group:

Quadrant	Group	Contrast	Estimate	SE	df	t.ratio	p.value
TL	BottomRight	POST - PRE	0.01776	0.0157	288	1.129	0.2596
BL	BottomRight	POST - PRE	0.02066	0.0157	288	1.313	0.1901
TR	BottomRight	POST - PRE	-0.00365	0.0157	288	-0.232	0.8169
BR	BottomRight	POST - PRE	0.03790	0.0157	288	2.410	0.0166
TL	TopLeft	POST - PRE	0.03517	0.0162	288	2.176	0.0304
BL	TopLeft	POST - PRE	0.02879	0.0162	288	1.781	0.0759
TR	TopLeft	POST - PRE	0.01654	0.0162	288	1.023	0.3070
BR	TopLeft	POST - PRE	0.02610	0.0162	288	1.615	0.1073

Table 2
Pairwise comparison results for accuracy in each quadrant and group with POST-PRE contrast

behavioral models: We defined the distance of distances, DD, as the relative distance of the target tree from reference tree B minus its distance to A:

$$\mathrm{DD_t} = \sqrt{(T_{x_t} - B_{x_t})^2 - (T_{y_t} - B_{y_t})^2} - \sqrt{(T_{x_t} - A_{x_t})^2 - (T_{y_t} - A_{y_t})^2} \tag{10}$$

where T, A and B are the target tree are the two references for trial t. x represents the number of leaves and y the number of fruits (in logged space).

To model angular-dependent choice biases, we implemented the harmonic addition theorem which states that a linear combination of sine and cosine functions of the same frequency can be expressed as a single sine function with a different amplitude and phase. Specifically, if we have: $A\sin(\theta) + B\cos(\theta)$, this can be rewritten as: $R\sin(\theta + \phi)$, where R is the amplitude and ϕ is the

phase shift. The values of R and ϕ can be given by: $R = \sqrt{A^2 + B^2}$,$\phi = \tan^{-1}\left(\frac{B}{A}\right)$.

The best fitting model to capture choice biases, irrespective of group or session in the perceptual task was:

$$\begin{aligned} \log \frac{P(A)}{P(B)}^{t}_{k} = \beta_0 + \gamma_{0k} + \\ \beta_1 \mathrm{DD} + \beta_2 q \mathrm{DD} + \\ \beta_3(\sin\theta + \cos\theta) + \beta_4 q(\sin\theta + \cos\theta) + \\ \beta_5(\sin 2\theta + \cos 2\theta) + \beta_6 q(\sin 2\theta + \cos 2\theta) + \\ \nu_1 \mathrm{c} + \nu_2 \mathrm{g} + \nu_3 \mathrm{s} + \nu_4 \mathrm{q} + \nu_5 \mathrm{qs} \end{aligned} \tag{11}$$

where $\log \frac{P(A)}{P(B)}^{t}_{k}$ is the probability to choose A over B for subject k at trial t and β_0 and γ_{0k} represent global and subject-specific intercepts. DD is the distance of distances, $\sin(\theta)$ and $\cos(\theta)$ and ($\sin(2\theta)$ and $\cos(2\theta)$) represent the 1st and 2nd harmonic components respectively whereas θ is the angle between the reference trees of the specific trial. As mentioned above, adding $\sin(\theta)$ and $\cos(\theta)$ is equivalent to adding a single $\sin$ wave allowing free phase and amplitude. c the cohort (MRI or behavioral only), g the group, s the session and q quadrant. All the νs represent nuisance regressors of no interest.

The best fitting model to capture choice bias including sessions and group interaction with the main components was:

$$\begin{aligned} \log \frac{P(A)}{P(B)}^{t}_{k} = \beta_0 + \gamma_{0k} + \\ (\beta_1 + \beta_2 q + \beta_3 s + \beta_4 g + \beta_5 qs + \beta_6 gs + \beta_7 qg + \beta_8 qgs)\mathrm{DD} + \\ (\beta_9 + .. + \beta_1 6 qgs)(\sin\theta + \cos\theta) + \\ (\beta_1 7 + .. + \beta_3 2 qgs)(\sin 2\theta + \cos 2\theta) + \\ \nu_1 \mathrm{c} + \nu_2 \mathrm{g} + \nu_3 \mathrm{s} + \nu_4 \mathrm{q} + \nu_5 \mathrm{qs} \end{aligned} \tag{12}$$

Where all the parameters are the same as in eq.11, only now βs 1-8 represent the main effect of DD, interactions with group, session, and quadrant, all possible 3-way interactions and the 4-way

interaction. βs 9-16 show the same for the first harmonic (H_1) and βs 17-32 for the second harmonic (H_2).

Post hoc tests were done using Estimated marginal means with a z-test for significance using (Lenth, 2025) toolbox in R.

fMRI data

fMRI data acquisition. MRI data was acquired using a 32-channel head coil on a research-dedicated 3-Tesla Siemens Magnetom TrioTim MRI scanner (Siemens, Erlangen, Germany) located at the Max Planck Institute for Human Development in Berlin, Germany.

Each MRI session consisted of 5-6 functional runs (see above) with anatomical scan and field maps (see below) placed right after the 3rd functional run. We placed the anatomical scan in the middle of the session to provide a mid-session short rest and made sure the design is roughly balanced between the two halves (see above).

The structural images were acquired using a T1-weighted magnetization-prepared rapid gradient-echo (MPRAGE) sequence (sequence specification: 192 slices; TR = 1900 ms; TE = 2.52 ms; FA = 9 degrees; inversion time (TI) = 900 ms; matrix size = 192 x 256; voxel size = 1 x 1 x 1 mm). The scan with parallel imaging (GRAPPA acceleration factor = 2) and isotropic resolution allow for precise anatomical measurements.

To correct for susceptibility-induced distortions, two spin-echo echo-planar imaging (EPI) field maps were acquired immediately after the structural scan with reversed phase-encoding directions (sequence specifications: TR = 1.75 s; TE = 28.8 ms; FA = 71 degrees; multiband acceleration factor = 4; partial Fourier = 0.75; matrix size = 112 × 112; voxel size = 1.75 × 1.75 × 1.75 mm; 80 slices). The field maps were collected with identical geometric distortions as the functional scans but with phase-encoding directions reversed: one in the anterior-to-posterior direction (AP) and one in the posterior-to-anterior direction (PA). These field maps were used for distortion correction of the functional data during preprocessing (see below).

The functional images were acquired using a multiband echo-planar imaging (EPI) sequence (sequence specification: 80 slices; TR = 1.75 s; TE = 28.8 ms; FA = 71 degrees; multiband

acceleration factor = 4; partial Fourier = 0.75; phase encoding direction = j-; matrix size = 112 × 112; voxel size = 1.75 × 1.75 × 1.75 mm). The scan was performed with effective echo spacing of 0.77 ms, bandwidth per pixel phase encode of 11.596 Hz/px, and a total readout time of 0.0855s. A total of 360 volumes were acquired during each functional run. Scans were manually aligned to the corpus callosum (i.e. no tilt angle) based on pilot testing to maximize signal from both the MTL and the orbitofrontal cortex (OFC, see Weiskopf et al., 2006). Unfortunately, after collecting the full sample, it seems that the entorhinal cortex still suffered from signal drop (see exclusions). For each functional run, the task began after 10 seconds to avoid partial saturation effects and allow for scanner equilibrium. Each run was about 10 minutes in length. We measured respiration and pulse during each scanning session using pulse oximetry and a pneumatic respiration belt part of the Siemens Physiological Measurement Unit.

BIDS conversion and defacing. Data was arranged according to the brain imaging data structure (BIDS) specification (K. J. Gorgolewski et al., 2016) using the HeuDiConv tool (version 0.6.0.dev1; freely available from https://github.com/nipy/heudiconv). Dicoms were converted to the NIfTI-1 format using dcm2niix [version 1.0.20190410 GCC6.3.0; (X. Li et al., 2016)].

fMRI preprocessing (fmriprep). Results included in this manuscript come from preprocessing performed using *fMRIPrep* 23.1.4 (Esteban, Markiewicz, et al., 2018; Esteban, Blair, et al., 2018; RRID:SCR 016216), which is based on *Nipype* 1.8.6 (K. Gorgolewski et al., 2011; K. J. Gorgolewski et al., 2018; RRID:SCR 002502).

Preprocessing of B_0 inhomogeneity mappings A total of 4 field-maps were found available within the input BIDS structure for this particular subject. A B_0-nonuniformity map (or *fieldmap*) was estimated based on two (or more) echo-planar imaging (EPI) references with 'top-up' (**topup**; FSL None).

Anatomical data preprocessing A total of 4 T1-weighted (T1w) images were found within the input BIDS dataset. All of them were corrected for intensity non-uniformity (INU) with 'N4BiasFieldCorrection' (Tustison et al., 2010), distributed with ANTs (version unknown) (**RRID:SCR_004757**; Avants et al., 2008). The T1w-reference was then skull-stripped with a

Nipype implementation of the 'antsBrainExtraction' workflow (from ANTs), using OASIS30ANTs as the target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white matter (WM), and gray matter (GM) was performed on the brain-extracted T1w using 'fast' [FSL (version unknown), RRID:SCR 002823, (Y. Zhang et al., 2001)]. An anatomical T1w-reference map was computed after registration of 4 T1w images (after INU-correction) using *mri_robust_template* [FreeSurfer 7.3.2, (Reuter et al., 2010)]. Brain surfaces were reconstructed using 'recon-all' [FreeSurfer 7.3.2, RRID:SCR 001847, (Dale et al., 1999)], and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle [RRID:SCR 002438, (Klein et al., 2017)]. Volume-based spatial normalization to two standard spaces (MNI152Lin, MNI152NLin2009cAsym) was performed through nonlinear registration with 'antsRegistration'(ANTs (version unknown)), using brain-extracted versions of both T1w reference and the T1w template. The following templates were selected for spatial normalization and accessed with *TemplateFlow* [23.0.0, (**templateflow**)]: *ICBM 152 Nonlinear Asymmetrical template version 2009c* (**mni152nlin2009casym**) RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym.

Functional data preprocessing (fmriprep). For each of the 21 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using 'mcflirt' [FSL <ver>, (Jenkinson et al., 2002)]. The estimated *fieldmap* was then aligned with rigid-registration to the target EPI (echo-planar imaging) reference run. The field coefficients were mapped onto the reference EPI using the transform. BOLD runs were slice-time corrected to 0.834s (0.5 of slice acquisition range 0s-1.67s) using '3dTshift' from AFNI (**RRID:SCR_005927**; Cox & Hyde, 1997). The BOLD reference was then co-registered to the T1w reference using 'bbregister' (FreeSurfer) which implements boundary-based registration

(Greve & Fischl, 2009). Co-registration was configured with six degrees of freedom.

Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, Power et al., 2014) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al., 2002). FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* [following the definitions by (Power et al., 2014)]. The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction [*CompCor*, (Behzadi et al., 2007)]. Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM, and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, a mask of pixels that likely contain a volume fraction of GM is subtracted from the aCompCor masks. This mask is obtained by dilating a GM mask extracted from the FreeSurfer's *aseg* segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the *k* components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (**confounds_satterthwaite_2013**). Frames that

exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. Additional nuisance time-series are calculated by means of principal components analysis of the signal found within a thin band (*crown*) of voxels around the edge of the brain, as proposed by (**patriat_improved_2017**). The BOLD time-series were resampled into several standard spaces, correspondingly generating the following *spatially-normalized, preprocessed BOLD runs*: MNI152Lin, MNI152NLin2009cAsym. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. The BOLD time-series was resampled onto the following surfaces. (FreeSurfer reconstruction nomenclature): *fsnative*, *fsaverage*. All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using 'antsApplyTransforms' (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using 'mri_vol2surf' (FreeSurfer). Many internal operations of *fMRIPrep* use *Nilearn* 0.10.1 (**RRID:SCR_001362**; Abraham et al., 2014), mostly within the functional processing workflow. For more details of the pipeline, see [the section corresponding to workflows in *fMRIPrep*'s documentation](https://fmriprep.readthedocs.io/en/latest/workflows.html "FMRIPrep's documentation").

Additional 18 physiological parameters were calculated to be used as confound regressors (8 respiratory, 6 heart rate, and 4 of their interaction). These were derived from collected physiological data. For these 18 physiological parameters, correction for physiological noise was performed via RETROICOR (Glover et al., 2000; Hutton et al., 2011) using Fourier expansions of different order for the estimated phases of cardiac pulsation (3rd order), respiration (4th order) and cardio-respiratory interactions (1st order) (A. K. Harvey et al., 2008): The corresponding confound regressors were created using the Matlab PhysIO Toolbox (Kasper et al., 2017, open source code available as part of the TAPAS software collection (Version 3.2.0):

https://www.translationalneuromodeling.org/tapas. For more details of the pipeline, and details on other confounds generated but not used in our analyses, see the section corresponding to workflows in fMRIPrep's documentation.

Regions of Interests (ROIs) and feature selection. We constructed several Regions of Interest (ROIs) based on the output of fmriprep and freesurfer. For each ROI we took specific labels from the freesurfer output, across both hemispheres (Entorhinal cortex: 1006, 2006; OFC: 1012, 1014, 2012, 2014; and Visual Cortex: 1005, 1011, 1013, 1021, 1029, 2005, 2011, 2013, 2021, 2029). Then, similar to the procedure done in SPM for implicit masking (Penny et al., 2011), we excluded voxels that showed a mean signal lower than 50% of the global mean signal across all voxels and TRs, for each functional run (unsmoothed but preprocessed). Then we took the intersect across runs and sessions to result in the subject-specific ROI used in the analyses. To compute the % loss used for exclusion, we divided the number of voxels that survived by the number of voxels from the freesurfer map (i.e. if no mean signal was computed).

GLM analyses. We conducted a block-wise General Linear Model (GLM) analysis on our fMRI data to investigate neural responses during the perceptual discrimination task. The fMRI data were preprocessed using fmriprep, and physiological parameters were included as regressors of no interest. The analysis was performed in MNI space with a smoothing kernel of 4mm.

For each block, we specified a GLM with the following components: Reference trees were split by their location in the map, such that one regressor was set for each target tree. All target trees repeated twice within a session, except when the target was in the middle of the quadrant which repeated more. For the reference trees, we split the regressors by the angle of the line (0 to 180 degrees in jumps of 15). Separate regressors were used for OneD trials and or the Big quadrant trials. All regressors had a duration of 0. A total of 17 regressors from fmriprep and 18 from physiological parameters were included as regressors of no interest. The GLM was estimated using classical restricted maximum likelihood. The model included a high-pass filter and an AR(1) model to account for serial correlations.

RSA analyses. RDM was conducted using betas taken from the above mentioned GLM, currently only taking the betas associated with the target tree. We then performed multivariate noise normalization (normalize each voxel by its residuals from the GLM, Walther et al., 2016). Lastly, we computed either the 1-Pearson or the Euclidean distance between each pair of patterns across runs (to avoid temporal auto-correlations) using Nilearn (Abraham et al., 2014). Note that noise-normalized Euclidean distance is equivalent to the Mahalanobis distance (Walther et al., 2016). To control for temporal autocorrelation, we only included trials across runs, and excluded the the diagonal, we excluded any correlation within a run across conditions (where the diagonal would be 1). This resulted in an RDM for each subject and each block comparison between all possible target trees. For the currently presented analyses we only included trials from the four small quadrants.

Project 3: Representational spaces in orbitofrontal and ventromedial prefrontal cortex: Task states, values, and beyond

work with Shany Grossman & Nicolas W. Schuck.

Published at: Trends in Neuroscience (2024) DOI: https://doi.org/10.1016/j.tins.2024.10.005

Highlights

- Orbitofrontal and ventromedial-prefrontal areas encode a rich diversity of variables that intertwine value, task-state, and outcome properties, using mixed selectivity
- A core benefit of this representational space is that it can encode contextual variables that are not directly observable but are required for predicting outcomes
- Task-state representations emerge in deep reinforcement learning networks alongside value-like signals, offering insights into why the brain multiplexes value with other task-related variables
- The complex activity patterns in OFC/vmPFC can be interpreted as an internal representation that maps internal states, sensory observations and past knowledge onto values and choice preferences

Main Text

Computational and neural underpinnings of value-based decision making

Humans and other mammals are versatile decision-makers, skilled at quickly learning how to achieve their goals in diverse environments. To do so, we learn to anticipate the outcomes of our choices. But while learning outcome expectations is straightforward in simple tasks, optimizing real-world behavior is more complex. It requires generalizing expectations across events, and understanding how changing goals affect outcome desirability.

One prominent notion is that the goal of decision making is to maximize the so-called expected value of a decision (Silver et al., 2021; Sutton & Barto, 1998), which is defined as the (time-discounted) sum of all expected future rewards after a choice is made. The basic idea of value maximisation goes back centuries to Expected Utility Theory (Peasgood, 2014), which states that decisions aim to maximize the expected value of a utility function that represents our subjective preference. Expected Utility theory has had a marked impact on psychological theory ever since it was observed that rational decision makers will behave as if they are maximizing

expected utility (Samuelson, 1947), although psychological literature has pointed out many important additional perspectives on what drives human choices (Gigerenzer & Gaissmaier, 2011; Kahneman & Tversky, 1979).

In parallel to these discussions, several neuroscientific studies have found that ventro-medial prefrontal (vmPFC) and adjacent orbitofrontal (OFC) areas are implicated in value processing in humans, non-human primates and rodents [e.g., (O'Doherty et al., 2001; Padoa-Schioppa & Assad, 2006), for reviews see (Bartra et al., 2013; Clithero & Rangel, 2014) as well as Fig. 17A], and interact with the wider corticolimbic dopaminergic reward system (Averbeck & O'Doherty, 2022). One influential study showed that when monkeys choose between different quantities of flavored juice or water, single neurons in the OFC reflect the animal's subjective value of the different outcomes (Padoa-Schioppa & Assad, 2006). The existence of value signals throughout the vmPFC and OFC has since been confirmed in humans, monkeys and rodents [e.g., (Ballesta et al., 2020); see (Bartra et al., 2013; O'Doherty et al., 2001) for reviews] and is broadly supported by lesion studies (Ballesta et al., 2020; Fellows, 2007; Hogeveen et al., 2017; Vaidya & Fellows, 2020). Note that we will use the term expected value hereafter to denote the subjective belief of the subject about the expected outcome of a decision. While in many cases the objective and subjective values align, some experimental paradigms can dissociate the two types of value.

In this review, we argue that the role of OFC/vmPFC goes beyond providing an (subjective) value signal, and suggest instead that they have a broader function focused on integrating information in the service of *learning* to predict outcomes in rich and partially observable environments. We first summarize findings from human, primate and rodent studies that relate hidden **task-state** (see Glossary) representations to these regions, and show that value and task-state codes are intertwined (Fig. 17B). We then show that a similar intertwining occurs in value-maximizing neural networks capable of performing complex tasks. Finally, we argue that these **deep reinforcement learning models** indicate that value maximizing computations do not necessitate the dominance of value representations as envisaged in neuroscientific research, and might serve as a useful model of OFC/vmPFC function that emphasizes the integration of predictive and

possibly unobservable task states with expected values.

While our focus is on learning, our conception of OFC/vmPFC function includes a deep interaction with memory processes which can reinstantiate pre-existing value or policy knowledge when needed (Abitbol et al., 2015; Frömer et al., 2019; A. H. Harvey et al., 2010; Lebreton et al., 2009; Lopez-Persem et al., 2020; Plassmann et al., 2007; Suzuki et al., 2017). This process is also crucial when old knowledge needs to be recombined in the service of inference (e.g. Barron et al., 2020), and during continual learning processes that involve ongoing refinement. We propose that the interaction between OFC/vmPFC and the hippocampus is critical for such a reinstatement.

We also acknowledge that VMPFC and adjacent OFC are anatomically diverse regions with many subdivisions (Cavada et al., 2000). Although many studies hint at differences between sub-regions (M. Z. Wang et al., 2022), anatomical differences between species and lack of terminological agreement make integration of evidence at a finer anatomical scale difficult. Our focus on the "OFC/vmPFC" region reflects the most pronounced distinction between medial and lateral areas (e.g. Izquierdo, 2017), in line with previous work (Levy & Glimcher, 2012), and broadly corresponds to the "orbital and medial prefrontal cortex (OMPFC)" region as defined by Ongur and Price (Öngür & Price, 2000).

All you need is value?

Inspired by early economic theory [e.g., (Samuelson, 1947)], some researchers have proposed that value signals reflect a "common currency" that acts as a stable cardinal desirability scale guiding individuals' decisions (Fehr & Rangel, 2011; Levy & Glimcher, 2012). According to this conceptualization, one of OMPFC's main functions is to map incommensurable options onto a unidimensional, cardinal scale. Early observations that OFC value signals were independent of sensory features, motor aspects, or other choice options (Padoa-Schioppa & Assad, 2006, 2008; Tremblay & Schultz, 1999) support this idea, leading to the assumption that OMPFC signals are tailored to generalize across aspects that are represented in other brain areas. This is supported by

the finding that vmPFC signals can be decoded across tasks with different goals (Castegnetti et al., 2021; Frömer et al., 2019; Gross et al., 2014; Howard et al., 2015; Kobayashi & Hsu, 2019; McNamee et al., 2013; Westbrook et al., 2019; Yao et al., 2023; Z. Zhang et al., 2017), and even when cognitive effort (Westbrook et al., 2019), or acquisition of knowledge (Kobayashi & Hsu, 2019) drive valuation or choice.

Other lines of research, however, challenged this idea. Choice preferences, for instance, are affected by irrelevant alternatives, and the range of outcomes (Bavard et al., 2018; Vlaev et al., 2011), counter to the predictions of common currency accounts. Contextual information is also important for value encoding, as illustrated in several studies [e.g., (Winston et al., 2014); for a review see (Juechems & Summerfield, 2019)]. Internal states, such as tiredness, modulate choices and also affect values in the brain [(Pastor-Bernier et al., 2021; Yoshimoto et al., 2022)], in line with the observation that desirability and neural value signals are goal-dependent – for instance, a hammer is better than a spoon if you want to drive a nail into the wall, while the opposite is true if you want to eat soup (Castegnetti et al., 2021; Elliott Wimmer & Büchel, 2019; Frömer et al., 2019; Moneta et al., 2023; Zhou et al., 2019). Some evidence also suggests that OMPFC is involved in optimizing other objective functions. Decision confidence, for instance, affects OMPFC firing, suggesting that this region might support maximizing confidence (Barron et al., 2015; De Martino et al., 2013; Gherman & Philiastides, 2018; Lebreton et al., 2015), even when it is orthogonal to expected values (Shapiro & Grafton, 2020).

Context matters: how tasks shape choice and neural value signals

Much work has highlighted context-dependency of decisions, further underlining the forementioned challenges to common currency ideas (see e.g. Hayden & Niv, 2021; Juechems & Summerfield, 2019; Knudsen & Wallis, 2022; Miller et al., 2019; Palminteri & Lebreton, 2021). Decisions made across contexts, for instance, can systematically violate the principle of value maximization (Bavard & Palminteri, 2023; Bavard et al., 2018; Molinaro & Collins, 2023a;

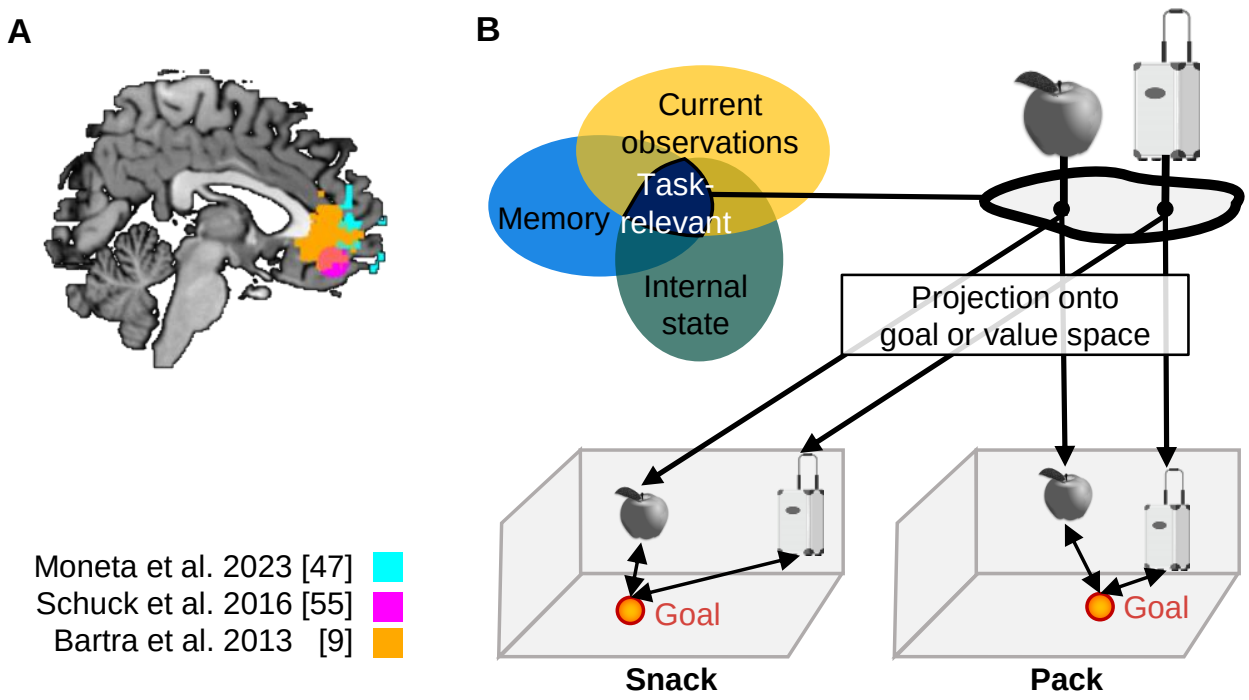


Figure 17

Values and task states in orbitofrontal and ventromedial prefrontal cortex (A) A midsagittal section of the human brain (MNI template) overlaid with medial prefrontal regions identified as encoding value in a meta analysis of BOLD fMRI experiments (Bartra et al., 2013) in orange. This region overlaps with the OFC area reported by (Schuck et al., 2016) where representations of (partially-observable) states were found (pink), and the medial prefrontal cortex region reported in (Moneta et al., 2023) where value and state representations co-exist and interact (cyan). **(B)** Illustration of how task states influence option values. Task states reflect the combination of sensory and non-sensory variables that are predictive of future outcomes. The computation of task states therefore requires input from several other areas which supply sensory processing, memory function, and access to internal affective and arousal states, amongst others. This information serves to map options onto the values they have for a given goal, thereby allowing the same options to have different values in different contexts. Images of suitcase and apple were adapted from vectorportal (Licensed under CC BY 4.0) and Wikimedia Commons (Licensed under CC BY 1.0), respectively.

Palminteri et al., 2015). In one study, participants were trained to decide between outcomes ranging either from 14 to 50 points or from 14 to 86 points (Bavard & Palminteri, 2023). Asked to pick options across sets, participants chose based on the within-set relative rather than the absolute values, making seemingly irrational decisions. This suggests that values are normalized within each context. Similarly, single cell recordings in macaque OFC found that value signals are normalized by the range of the current context (Conen & Padoa-Schioppa, 2019; Padoa-Schioppa, 2009), in line with human fMRI findings (Nelli et al., 2023), and modeling work (Zimmermann et al., 2018). Interestingly, value range adaptation does not seem to appear in OFC during forced choices, suggesting that this form of context-sensitivity is itself context-dependent

(Yamada et al., 2018).

One of our recent studies has provided additional insights into the relation between context and value signals in vmPFC (Moneta et al., 2023). Participants first learned discrete values of four colors and four movement directions while undergoing fMRI. They were then asked to make a choice between two moving and colored stimuli, based only on either the color or motion-direction, but not both (feature relevance was explicitly cued, and changed every 4-7 trials; see Fig. 18a.). Standard analyses confirmed that value was decodable from vmPFC. Nonetheless, using only value-responsive voxels, current task context was decodable too – although contexts were matched in value (Fig. 18B). Two keys observations were made: first, value and context were related, i.e. stronger context signals correlated with a stronger value signal within participants (Fig. 18C), as well as with the degree to which behavior was influenced by the irrelevant context (Fig. 18d). Moreover, these two effects were related: a strong connection between vmPFC context and value signals was linked to less influence of "irrelevant" context on behavior, Fig.18E. Context thus seemed to co-exist with and enhance value representations, and determined which values influenced behavior. Second, behavior and vmPFC signals were influenced by the irrelevant feature with the highest value, which sometimes was not the chosen option. This implies a hypothetical calculation of the maximal possible value, assuming the alternative context, and possibly another choice (Fig.19A-B). Results hence suggested that vmPFC calculated the values of each context, which then competed for representation. Strikingly, context signals modulated which value signal dominated vmPFC – the true value or the above-mentioned hypothetical value (Fig.19C). Hence, task-context signals, not options or actions, organized value representations and choices (see Fig.19 and Box 1), providing a way for vmPFC value signals to reflect possible future goals in addition to ongoing tasks.

That vmPFC representations of contexts and values interact with each other, is broadly in line with evidence for context signals alongside values in OMPFC (Baram et al., 2021; Cromwell et al., 2018; Farovik et al., 2015; Z. Zhang et al., 2017), but also extends previous work by demonstrating an intricate interplay of value and context signals.

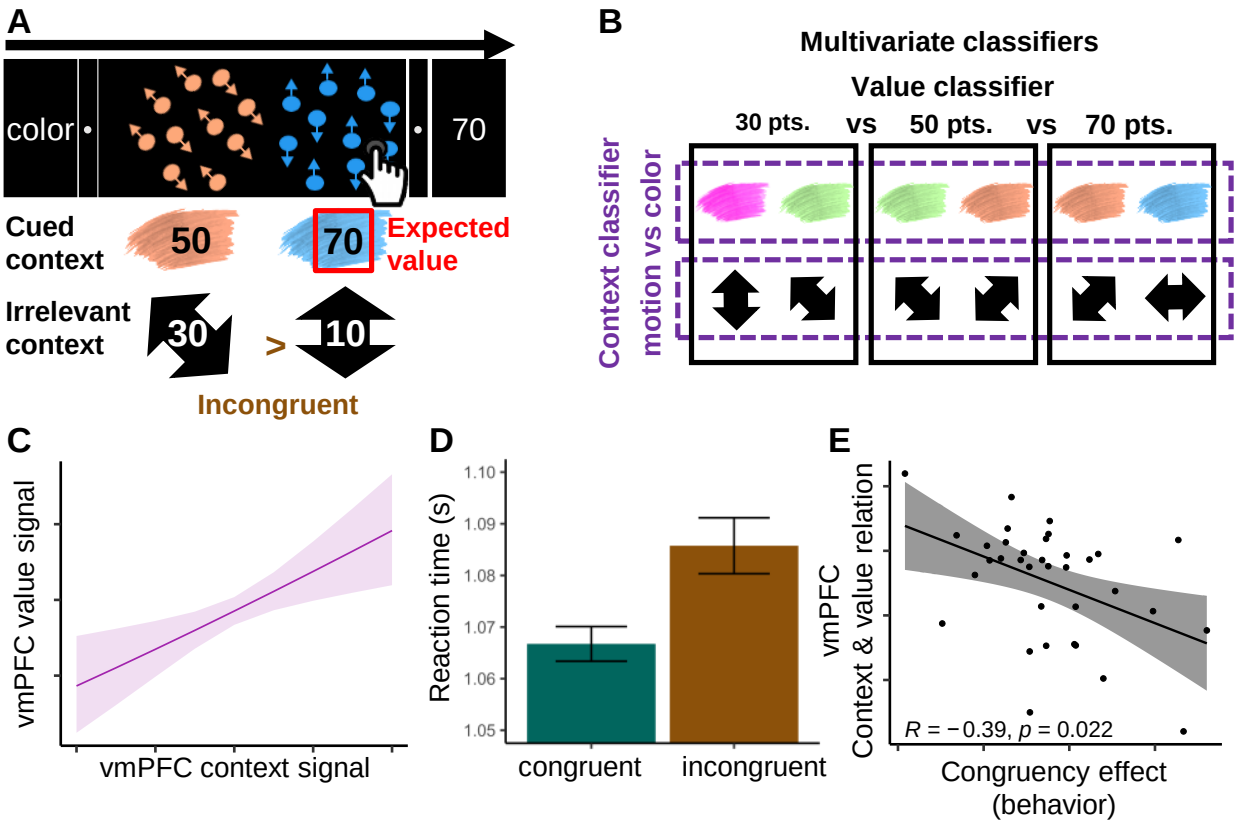


Figure 18

Interlinked vmPFC representations of task-state and expected value (A) Schematic illustration of the experimental paradigm in (Moneta et al., 2023). After learning to associate rewards with a set of colors and motion directions, participants made choices between two color-motion stimuli. Before each decision, a context cue indicated whether rewards were dependent on color or direction (here: color). The expected value of a trial was the maximum reward of the cued features (here: 70). On congruent trials, choosing the maximally rewarding cued feature also selected the most rewarding uncued feature; on incongruent trials, the reverse was true (see example). Outcome presented after each choice only depended on the features of the cued context. **(B)** Pattern classifiers were trained on vmPFC data to either distinguish between different values (irrespective of context) or trial context (irrespective of value). The ROI is indicated in Fig.17A. **(C)** Expected value and context could both be decoded from the same vmPFC area, which was defined based on value only (main effect not shown). Moreover, these decoding strengths were related to one another: a stronger context signal (x-axis) accompanied a stronger expected value signal (y-axis). Shown are mixed effects models testing the association between expected value and context decoding. **(D)** Participants were slower on incongruent compared to congruent trials, i.e. when the contexts didn't agree which decision was best, showing that alternative context influenced behavior. **(E)** Participants who showed a weaker relationship between context and value representations in vmPFC (y-axis, panel c) also showed a stronger behavioral influence of the irrelevant context (congruency effect, x-axis, panel d). Plot shows the correlation of the betas from an expected value decoding model (y-axis) with the congruency effect in reaction times (x-axis). Panels modified from (Moneta et al., 2023).

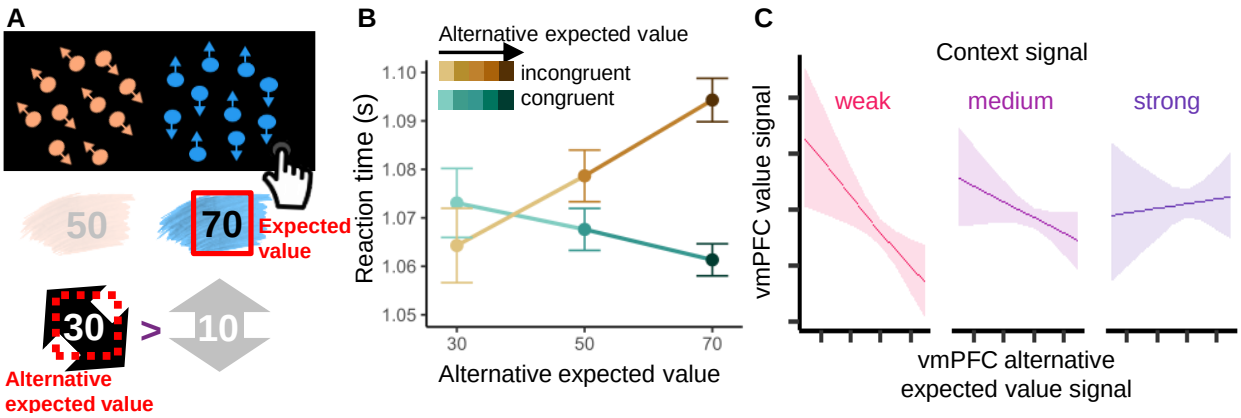


Figure 19

Context encoding modulates irrelevant value signals (A) Schematic illustration of the experimental paradigm in (Moneta et al., 2023). Participants made choices between two color-motion stimuli, cued to focus only on color or only on motion. Reward was only predicted by the cued context (here, color). In the trial presented, the best choice according to the cued dimension is on the right (denoted expected value), while the maximum rewarding feature of the cued-to-ignore context is on the left, making this trial incongruent. Thus, the 'alternative' expected value reflects the maximum number of points that would have been obtained in the alternative context (i.e. if the irrelevant features were the relevant ones). **(B)** Alternative expected value influenced behavior, but only in relation to the congruency of contexts. Larger irrelevant values led to faster reaction times in congruent trials (green), and slower reaction times on incongruent trials (brown). Note that on incongruent trials, the irrelevant expected value reflected a different, hypothetical choice (c.f. panel a). **(C)** Relationship between neural signals for expected value (y-axis) and the strength of irrelevant expected value (x-axis), separately for trials in which the context signal was weak, medium or strong. The neural representation of both expected values (stemming from the relevant and irrelevant contexts) was negatively related. However, this negative relation was modulated by the context signal. When the context signal was strong, the influence of the irrelevant values on the vmPFC signal was reduced, akin to an arbitrating effect between competing value signals. Shown relationships reflects mixed effects model testing the association between expected value decoding and alternative expected value decoding.

From Context to Task States: how cognitive maps influence values

The context-sensitivity of value signals in the brain might not be surprising given that adaptive behavior needs to reflect how goals and contexts influence outcome desirability. But how exactly should 'context' be defined? One perspective, albeit not the only one, comes from reinforcement learning (RL) theory (Sutton & Barto, 1998), which formalizes how agents can learn reward maximizing behavior from trial-by-trial feedback. The simplest RL algorithms receive handcrafted information about the current "state", or context, of the environment, which does not have to be directly observable but can for instance be defined by past events or internal needs. If RL models activate the wrong state they will also retrieve the wrong value, which means that reward learning is always contingent on current state knowledge (Fig. 17B). While this

perspective agrees with other theories that values are abstract in nature and enable comparison of incommensurable options, it suggests that relevant task details must exist in the same region – a level of specificity that has been de-emphasized in particular by common currency approaches. The aforementioned findings of context signals that reside alongside value in OMPFC (Baram et al., 2021; Cromwell et al., 2018; Farovik et al., 2015; Schuck et al., 2016; Z. Zhang et al., 2017) and their close connection to value signals in the same area (Moneta et al., 2023) support this perspective.

What are states, specifically? First, RL accounts emphasize that states must (exclusively) reflect information that is needed to predict future reward. This can be sensory information (whether the sun is shining or not), but it can also be something that cannot be observed directly, such as how much time has passed. The second major aspect is that, in some RL models, states are part of a cognitive map that specifies transitions between them. This emphasis on predicting future states is core to RL approaches that provide additional flexibility (e.g. model-based RL Sutton, 1991; successor representations Stachenfeld et al., 2017 or replay approaches Schuck & Niv, 2019, see below). In sum, the RL perspective therefore emphasizes the role of reward and state predictiveness as defining features of context. Both aspects are in line with research on state representations in OMPFC(Bradfield & Hart, 2020; Niv, 2019; Schuck et al., 2016; Wilson et al., 2014) as well as on OFC role in generalization and inference (Baram et al., 2021; Boorman et al., 2021; Shi et al., 2023). More broadly, it can also explain the dominance of goal-aligned value signals in OMPFC in cases where values depend on goal context (Castegnetti et al., 2021; Frömer et al., 2019; Grueschow et al., 2015; Hare et al., 2009; Moneta et al., 2023, see below).

Adopting this perspective, one study showed that OFC lesions in rats affect reward-related dopamine firing in line with predictions from an account that assumes OFC is needed to signal, latent task states that are independent from sensory input (known as “partially observable”) (Wilson et al., 2014). A fMRI study in humans also found that partially observable states can be decoded from medial OFC (Schuck et al., 2016), in line with a number of other studies (Bradfield & Hart, 2020; Chan et al., 2016; Costa et al., 2023; Elliott Wimmer & Büchel, 2019; Farovik

et al., 2015; Moneta et al., 2023; Muhle-Karbe et al., 2023; Zhou et al., 2019). Hence, OMPFC may infer latent states that are needed to retrieve context-sensitive values, which is crucial when the same choices lead to different outcomes given partially observable states. Some studies suggest that OFC represents task structures and rules even without any explicit value (Lipton et al., 1999; Schuck et al., 2016; Zhou et al., 2021), akin to "schemas" that seem to reside in OMPFC too (Bein & Niv, 2023; Gilboa & Marlatte, 2017). OFC representations also appear similar when the same task is done with and without rewards, presumably reflecting stimulus-stimulus associations (Sadacca et al., 2018), akin to latent learning ideas (Tolman & Honzik, 1930a) that emphasize how stimulus-stimulus learning done in the absence of rewards can be used for later reward tasks. This suggests that OMPFC might serve more broadly as a cognitive map that guides decisions (Behrens et al., 2018; Schuck et al., 2018; Tolman, 1948), a function that is likely to occur in close connection with the hippocampus (see below for further discussion, and c.f. Garvert et al., 2023; Kaplan et al., 2017; Wikenheiser & Schoenbaum, 2016; Wikenheiser et al., 2017).

In sum, a state representation perspective envisages a dynamic process in which choice options can be flexibly projected onto different expected values, depending on goals and past history that influences how desirable an option is. States also support efficient learning by forming a cognitive map that facilitates generalization and planning in information rich environments with complex temporal structure (see e.g. Eppinger et al., 2023; Niv, 2019; Schuck et al., 2018).

While the evidence discussed above generally supports this perspective, several questions remain. One notable deviation is that OMPFC systematically represents task-irrelevant information and even "irrelevant" values – a finding that we discuss in Box 1. A second issue concerns cases in which non-value quantities are optimized, such as distance to a non-value goal (see below), where the predictiveness of OMPFC signals will not refer to future rewards, but closeness to goal. Finally, we acknowledge that while some have presented direct computational or empirical evidence that context and cognitive maps can be understood as states that arise in RL machinery (e.g. Stachenfeld et al., 2017; Whittington et al., 2022), further evidence is needed.

Box 1: Representation and compression of task-irrelevant values

A major role of state representations is to define which information is needed to predict outcomes in a given context. State representations have been linked to OMPFC (Schuck et al., 2016; Wilson et al., 2014), where lesions hinder the ability to ignore irrelevant choice-options (Noonan et al., 2017). Moreover, OMPFC activity compresses inputs to focus on goal-relevant information (Ebitz et al., 2020; Mack et al., 2020; Muhle-Karbe et al., 2023), and guides hippocampus in forming reward-predictive relational maps (Garvert et al., 2023). Compression also occurs in deep neural networks, although contingent on factors such as activation functions (Saxe et al., 2018) or weight initialization (Flesch et al., 2022). This raises the question how complete such compression is, i.e. if OMPFC still maintains some representation of (1) irrelevant task/stimulus information, and (2) irrelevant values.

In a study addressing the first question, participants were initially instructed to focus on one of two stimulus features, but later, unbeknownst to participants, the previously irrelevant feature suddenly became task-relevant (Schuck et al., 2015). Some participants did notice the changed relevance – even when this was never needed to complete the task (Gaschler et al., 2019) – and MRI results showed irrelevant information processing arose in mPFC before participants abruptly changed choices to tap into changed relevance. Neural network simulations of the same task demonstrate that regularized gating can lead to preserved latent knowledge of irrelevant aspects which can be accessed rapidly if needed and leads to similarly abrupt and spontaneous behavioral switches as those observed in humans (Loewe et al., 2024; Löwe et al., 2024). Broadly in line with this idea, some studies have shown that task-irrelevant features can be decoded from the frontal eye fields in monkeys and motor cortex in humans (Mante et al., 2013; Takagi et al., 2021).

If some representation of task-irrelevant information is maintained, what happens to task-irrelevant *values*? Many fMRI studies reported task-relevant values within vmPFC, but no univariate evidence of task-irrelevant values has been found [(Castegnetti et al., 2021; Frömer et al., 2019; Grueschow et al., 2015; Hare et al., 2009); but see (Abitbol et al., 2015; Frömer et al., 2019; A. H. Harvey et al., 2010; Lebreton et al., 2009; Levy & Glimcher, 2011) for task-independent value-like signals]. However, using multivariate methods, recent work from our group showed that such task-irrelevant values do exist in vmPFC (Moneta et al., 2023), interact with other value and non-value representations, and influence behavior (Fig. 18-19). This raises the possibility that multivariate fMRI methods are better suited to uncover compressed representations.

A complex and versatile code in OMPFC

The evidence reviewed so far suggests that values are part of a complex activation manifold with multiple dimensions related to choice values, (partially observable) task states, and alternative

values (Conen & Padoa-Schioppa, 2019; Moneta et al., 2023; Padoa-Schioppa, 2009; Schuck et al., 2016; Wilson et al., 2014). Electrophysiological studies support this idea and indicate diverse information encoding in OFC (Farovik et al., 2015; Lopatina et al., 2015, 2017). For instance, neurons in OMPFC encode summary statistics of the current task such as previous offers and outcomes, or the location of the currently attended offer (Mehta et al., 2019). Recording studies have also shown that the same neurons in OFC often encode multiple variables at once, a phenomenon known as mixed selectivity (Rigotti et al., 2013). One study recorded neurons from monkeys performing a choice between options characterized by the flavor and probability of a juice reward. The authors reported that most neurons in OFC showed mixed selectivity for probability and flavor (Stoll & Rudebeck, 2024). Another study showed that the same neurons in macaque OFC can represent both spatial and reward information, even when those are unrelated (Yoo et al., 2018). Although different variables can be encoded by the same neurons (Ebitz et al., 2020) or voxels (Moneta et al., 2023), merely representing different variables does not mean that they are integrated. Perhaps the most direct evidence for such an integration comes from (Moneta et al., 2023, see above), where context signal strength covaried with value signal strength and behavioral markers or context adaptive behavior.

This line of evidence raises a major conundrum: How can the findings that OMPFC activity multiplexes many task variables with reward expectations be reconciled with the reports of generalizable, content-independent value representations discussed above? One explanation could be that, on a population level, neurons with mixed selectivity can still form a high dimensional representation in which mostly orthogonal planes reflect different variables (Fusi et al., 2016). This means that downstream neurons can easily ready out independent codes for each variable. Indeed, in (Stoll & Rudebeck, 2024) the subspaces of population activity within OFC reflected probability and flavor and were minimally dependent (i.e. nearly orthogonal). We note that mixed selectivity is not unique to OMPFC and prevalent throughout the frontal cortex (Rigotti et al., 2013; Tye et al., 2024). This suggests that mixed selectivity has a very broad function in high-order and flexible cognition that goes beyond the specific computations in OMPFC.

Another, not exclusive, possibility is that goal-independent representations emerge during a late computational stage when state-dependent values transform into specific action-selection signals. In support of this idea, it has been shown that while the same OFC neuron population can be involved in evaluation and selection during value-guided choice, activations during these different phases lie on almost orthogonal subspaces (Yoo & Hayden, 2020). Others found expected values in OMPFC arise only when the task requires a selection [(Castegnetti et al., 2021), but see (Lebreton et al., 2009)], without encoding motor signals (Kennerley et al., 2009; Knudsen & Wallis, 2022; Moneta et al., 2023; Padoa-Schioppa & Assad, 2006)].

A neural network perspective on value and state representations

As reviewed above, OMPFC does not appear to be exclusively committed to signaling only value or only task-states. But can an exclusive focus on univariate codes for either value or task states even be expected from a complex computational system like the brain? One avenue for addressing this question is to study deep RL models, which reflect an integration of RL with deep neural networks (Botvinick et al., 2020). In addition to being powerful AI tools that can master games (Mnih et al., 2015; Silver et al., 2018), or control self-driving cars (Kiran et al., 2022), deep RL models can be useful for neuroscientific research of learning and decision-making (see Box 2).

Deep RL models' advantage over classic RL is their ability to master reward learning in complex environments by learning task-specific representations in their late layers, positioned close to the output. Classic RL models directly receive information about a small number of hand-crafted, discrete states, such as their position in an artificially discretized spatial environment. Deep RL models, by contrast, can learn directly from a high-dimensional, continuous, and noisy sensory description (e.g. perceived distance to the wall) by relying on their representation learning power to form task-appropriate lower dimensional representations of the input (Bengio et al., 2014) (see Fig. 20). Critically, these emergent representations arise without explicit guidance other than the network's objective function, which for deep RL is typically the optimal reward (e.g. Mnih et al.,

2015) or policy function (e.g. Heess et al., 2017; Mehta et al., 2019; Silver et al., 2016), and often end up having many features of (partially-observable) task-states we reviewed above. A reward-focused objective function therefore does not only lead to value representations, but also to task-appropriate abstractions of the sensory input, as we discuss below.

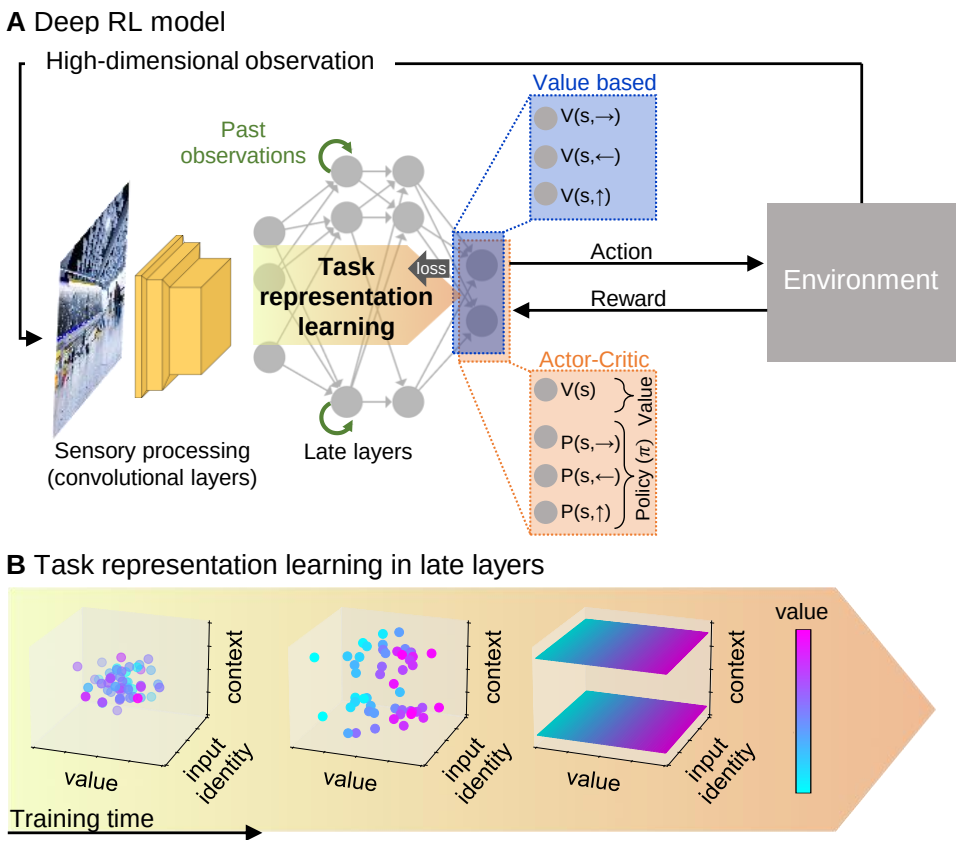


Figure 20

Principles of deep RL models and emergent representations in late hidden layers (A) General scheme of a deep RL model. High-dimensional inputs (e.g., thousands of pixels) are first processed through stacked convolutional layers (akin to sensory processing), usually followed by non-convolutional fully-connected layers ("late layers", marked here by the yellow-orange arrow). Adding recurrency to the network (green arrows) allows to incorporate past events with present observations for representing partially-observable states. In the case of purely value based models, such as DQNs, the output nodes are trained to approximate the expected action values; for more policy oriented models, output nodes are trained to reflect action probabilities, sometimes in addition to value estimates (such as in **Actor-Critic**). Once an action is chosen and rewarded, the error is used to update the weights of the model, incrementally forming hidden representations. While the learning process is focused on achieving maximally accurate value or action probability estimates, it also shapes the representations in the late layers such as to form distributed representations of task and value relevant variables that represent a compact world model, or 'cognitive map', of the task. **(B)** A schematic exemplifying the formation of hidden task representations in late layers of the deep RL network, while multiplexing value with non-value task variables. The axes reflect reduced dimensions of the population code (for example through principal component analysis or other dimensionality reduction methods). Note that context is not explicitly signaled in the input and the network needs to infer contexts, often based on observations that go beyond those currently observed (as is the case in partially-observed tasks).

Because of their densely connected multi-layer architecture, deep RL models also learn differently. Standard models only update the currently activated state when receiving input. But in deep RL models input and output are connected via many intermediate hidden layers that often feature mixed-selectivity (Dabney et al., 2020; Song et al., 2017; Wierda et al., 2023; Z. Zhang et al., 2018), similar to what has been observed in the OMPFC (see above). A single weight-update will therefore affect many representations, and learning is never confined to just one state (Fig. 20b).

Value and state representations in deep RL models

Using the principles outlined above, deep RL models solve complex tasks by learning to extract multiple layers of representations, with an increasing level of abstraction (Kozma et al., 2018). A major question is what characterizes late layer representations which are only a few computations away from the decision output, and whether their features correspond to what we know about OMPFC. One notable paper has shown that a recurrent deep RL model can capture several core aspects of OFC function, and might reconcile value and state accounts on this brain area (Pessiglione & Daunizeau, 2021). Other evidence from neural network studies also suggests late layer representations are not merely sorted by the value of the input they correspond to. In one study, (Mnih et al., 2015) a deep Q network (DQN; a form of deep RL model) was trained to play various Atari games. Visually investigating the DQN's last layer representations showed that the different input frames were not uniformly sorted by the value that the network predicted for them. Rather, value led to some clustering, but other factors such as perceptual or strategic similarity were reflected too. A subsequent study (Cross et al., 2021) found that the geometry of representations in late layers of the same DQN (i.e. their pairwise similarities) correlated with a hand-crafted geometry which retains abstract information about input features (e.g. the ball position, or the position of the two paddles in the game pong). Such abstract signals are reminiscent of the task-states discussed above. Counter to our argument, however, in (Cross et al., 2021), no link was found between these representations and human participants' OMPFC activity. Other work has used encoding models to ask whether hidden units reflect human-generated

concepts. One study (McGrath et al., 2022) demonstrated this approach on a deep RL model for chess, AlphaZero (Silver et al., 2017), and found late hidden layers come to represent many concepts other than the expected value of the current board, such as whether the player is in check, or whether the opponent can capture the queen. The selectivity profile of single units of deep RL models portrays a similar picture. In a deep RL model trained to solve a spatial reward task, a recent study (Suhaimi et al., 2022) found that good performance was related to the emergence of value selective units. But these units made up only 10% to 50% of the population, and units not related to value were also highly correlated with the performance of the model (although this analysis was performed on units taken from all 4 model layers, so its result is not exclusive to the late layers).

Similar observations have been made for different deep RL architectures. Hidden units in a decision network of an actor-critic-based recurrent RL model (see Box 2) show mixed selectivity to combinations of task conditions, such as context and stimulus coherence levels (Song et al., 2017). Other work on recurrent RL models has shown that hidden representations capture task structure by retaining information about recent choices and rewards (Hattori et al., 2023; J. X. Wang et al., 2018; Z. Zhang et al., 2018). Another line of work has demonstrated the importance of non-value representations more broadly by showing that adding other constraints on hidden representations than a reward maximization objective helps performance (de Bruin et al., 2018; Lesort et al., 2018).Unsupervised pretraining of neural networks, for instance, can speed up later training with a specific **objective function** (Hinton & Salakhutdinov, 2006). In addition, maximizing mutual information between hidden representations of inputs that are adjacent in time and space can enable better abstraction and generalization in Atari games (Anand et al., 2019). Further, a deep RL architecture can benefit from being endowed with grid-like representations prior to learning (Banino et al., 2018) or self-supervised learning objectives (Fang & Stachenfeld, 2023).

In sum, we argue that the late layers of deep RL networks offer a useful model to understand the computational role of OMPFC. This role consists in using (reward) feedback signals to shape a

mixed selectivity code in a way that emphasizes outcome predictive state and value representations. This process depends critically on input from many other regions, which for instance provide appropriately processed sensory information, or access to working memory. Perhaps the biggest challenge to this idea is that studies that directly compare late layers in deep RL models and OMPFC signals are largely lacking, and available evidence is inconclusive. Future experiments should test the extent to which deep RL models truly align with observed OMPFC signals.

Beyond standard deep RL: flexibility through long-term memory, meta learning and model-based RL

If OMPFC signals reflect computations akin to late layers in deep RL models, how can this account for value or liking signals in OMPFC that occur when no learning or action are needed? A potential explanation is that once value information has been established (e.g. value of known food items), the information can be reinstantiated in the network, for instance when conditioned stimuli are presented. This reinstatement is also critical for inference or when past goals are revisited, and even can occur spontaneously in the absence of a choice task.

The importance of flexible access to long term memory has often been overlooked because most laboratory tasks capture "isolated" learning processes that start from a blank slate and are completed after a few hours of experience. Yet, the perhaps most remarkable aspect of animal and human learning is the flexibility with which subjects apply previously gained insights to new problems (Sandbrink & Summerfield, 2024), and learn over long time spans to extract commonalities between learning problems (Lake et al., 2017). A notable recent advance that captures this idea, and can make deep RL models more flexible, is deep "meta-RL". Deep meta-RL uses a meta-learning approach in which slow weight changes come to encode fast learning in the activity dynamics of a recurrent network (Duan et al., 2016). The resulting models capture instances of accelerating learning over a set of new but related problems (J. X. Wang et al., 2018). Notably, recent findings have shown that plasticity within OFC is necessary for such a process (Hattori et al., 2023). Hence, the combination of access to previously established

knowledge with the aforementioned learning powers could give OMPFC a unique power to meta-learn and integrate fast with slow learning processes. One aspect of neural processing of potential relevance for this notion is the interaction between OMPFC and hippocampus, given the important role of the hippocampus in long-term memory and memory reactivation, the functional similarities between both regions, and their close connectivity (Öngür & Price, 2000). It should be noted that this idea suggests a more complex deep RL architecture with separate long-term value storage systems that interact with OMPFC.

A second approach within reinforcement learning frameworks to support flexibility when adapting to new problems is model-based RL, which learns a model of state transitions separately from values. The combination of values and a state transition model can then be used to make on-the-fly value calculations. Yet, most deep RL models we discussed so far are in fact model-free, i.e. they do not incorporate structural knowledge and consequently tend to be inflexible. When it comes to the brain, one possibility is that transition knowledge is stored outside of the OMPFC, but can still influence OMPFC computations via offline updating (Sharpe et al., 2019). In line with this idea, previous work has found that replay in the hippocampus – a putative mechanism used by the brain to sample from a model of the task during rest (Wittkuhn et al., 2021) – is linked to state representations in the OFC (Schuck & Niv, 2019), suggesting a role of hippocampal-OFC interactions in the service of flexibility.

A final consideration concerns the availability of task-irrelevant signals in deep RL models that is in line with findings about irrelevant signals in the OMPFC discussed above (Abitbol et al., 2015; A. H. Harvey et al., 2010; Moneta et al., 2023; Schuck et al., 2015). An intriguing open question is whether such irrelevant signals are intentionally retained to accommodate for a dynamic environment with constantly changing contexts ("a feature"), or whether the computational machinery is limited in suppressing them fully ("a bug"). Further studying such cross-task signals in deep RL models trained on several tasks might help elucidate the origins of their neural counterparts.

Box 2: Using deep RL models for the neuroscientific study of decision-making and learning
The core idea of deep RL models is to train a deep neural network through trial-and-error reward feedback, rather than through supervised training. These models usually receive sensory observations as inputs, such as image pixels, and are trained to output expected values and/or actions that maximize reward. Most deep RL models process visual inputs in early **convolutional layers**, on top of which **fully-connected layers** are stacked. A popular type are **deep Q-Networks (DQNs)**, which approximate the expected values of a set of discrete actions, given the input. Important additions to this standard architecture are **recurrent layers** that provide the network with memory, and replay buffers that allow offline sampling (Botvinick et al., 2020; Wittkuhn et al., 2021).
Deep RL models are regarded as a useful tool in neuroscience because they share some basic properties with the brain. They process information through layers of connected and distributed nodes in a stage-like fashion, and learn by adjusting the connection strength between nodes as a function of feedback. These broad principles are reminiscent of the distributed information processing and synaptic plasticity found in real neurons. Although these similarities are relatively superficial - substantial differences exist for instance in how synaptic weight updates are propagated throughout the networks - the main promise of deep RL networks is to offer a useful level of abstraction for studying algorithmic aspects of cognition. Because deep RL models can perform complex cognitive tasks on par with humans (Mnih et al., 2015; Silver et al., 2017), they seem to retain at least some of the necessary ingredients for complex cognitive skills. A burgeoning field now uses deep RL as models for behavior (Kuperwajs et al., 2023). Some work has made progress by deriving analytical solutions of learning dynamics in simplified neural networks which yield precise explanations for observed learning trajectories (Saxe et al., 2019). Others have used them to derive testable neurobiological predictions about context-dependent learning (Flesch et al., 2022) or to provide explanations about why certain computational ingredients are essential for achieving human-like learning (Flesch et al., 2018; Nelli et al., 2023). Finally, deep RL models are useful because they allow studying the interaction of learning algorithms, behavior, and representations, providing for instance ideas about which representations can be expected in value maximizing networks. One example that showcases this strength comes from the area of distributional RL models (Bellemare et al., 2017), which suggests benefits of computing many diversely tuned reward prediction errors (RPEs), rather than only the single RPE assumed in standard RL. While the single RPE signal has famously been found in the firing rate of dopamine neurons (Schultz, 1998), the notion of distributional RL in midbrain circuits has recently gained empirical support (Dabney et al., 2020; Muller et al., 2024). In a similar vein, deep RL models might help refine understanding of OMPFC codes in the brain.

Concluding remarks and future perspectives

We provided an overview of information encoded in OMPFC during decision making tasks. OMPFC representations are multifaceted, shaped not only by immediate and expected rewards, but also by sensory and non-sensory information required for optimizing behavior in current

tasks. As discussed in previous sections, such representational richness aligns with the concept of task-states in reinforcement learning, and with late hidden layer activations that arise in deep RL models that learn to perform complex tasks. Ultimately, this suggests that value-oriented computations do not necessarily lead to simple representations of expected value in the form of a universal currency for decision-making. Instead, we propose a perspective in which the OMPFC provides an integration of value and task states in the service of decision-making in complex environments. We also highlighted the important observation from neurophysiological as well as simulation work that single neurons are characterized by mixed selectivity to linear and nonlinear mixtures of value, outcome, task state and other variables. Notably, while information is mixed on the single neuron level, it is still possible for different variables to be read out independently on the population level (example visualized in Fig. 20B)). This implies that complex neural codes which feature information integration on the single neuron level do not contradict the existence of more abstract, independent representations on the population level (Tye et al., 2024)

While deep RL can offer useful insights for OMPFC function, we believe important aspects need to be considered. Of particular relevance is the need to complement on-task learning powers of standard models with access to long term memory in a way that enables learning across tasks over larger horizons. Some promising first results have indicated a link between OMPFC and meta-learning and memory replay, but more work will be need, in particular concerning the role of hippocampus-OMPFC interactions in this regard. We also argue that re-instatement of established (value) knowledge could explain the documented role OMPFC plays for flexible generalization, as well as in tasks that neither require learning nor choices. Another important observation that requires deeper investigation is that value signals can reflect not only current but also future or hypothetical tasks (see Moneta et al., 2023), suggesting OMPFC decision making function reflects not only past tasks, but also future ones.

While neural network-based computational models might inspire new concepts and predictions concerning representations in OMPFC, another remaining challenge is the lack of a clear correspondence between network components or computations and specific brain regions. Most

observations about similarities of deep RL models and OMPFC remain qualitative, and a previous study (Cross et al., 2021) failed to find any direct relation between the model representations and fMRI activity in OMPFC. Additional in-depth investigations are therefore critical (see Outstanding Questions).

Finally, we believe that it is time to reconceptualize value as a multidimensional signal that tracks distance to the current task-goal, rather than accumulated reward (Frömer et al., 2019; Juechems & Summerfield, 2019; Martino & Cortese, 2023). This approach could open the door for frameworks that integrate goal and value signals (Molinaro & Collins, 2023b), integrate confidence into the decision (Barron et al., 2015; De Martino et al., 2013; Gherman & Philiastides, 2018; Lebreton et al., 2015), and even ones which assume no explicit computation of value at all (Hayden & Niv, 2021). Together, we believe these shifts in focus will help gain better understanding of the full complexity of OFC/vmPFCs function.

Outstanding Questions

- Do late layers of deep RL models offer a model for representations in OMPFC? If so, which network architectures and layers therein best match OMPFC regions?
- Can recurrent neural network architectures reveal previously unidentified links between deep RL models and OMPFC, given their role in partially-observable tasks and meta learning?
- What influences encoding of non-value signals and information irrelevant for the current task in OMPFC? Can this be linked to objective functions, weight initialization and activation functions in neural networks?
- What controls the amount of compression and (dis)entanglement in OMPFC representations?
- Learning is a dynamic process that evolves over time. Is OMPFC's role in decision making most prominent during early learning stages?
- How do the representation learning dynamics of deep RL on a trial-by-trial level, and across episodes, compare with those in OMPFC?
- Does OMPFC guide value-free decision-making processes in the brain? How can one disentangle value-based vs value-free learning and choice?
- How do ventromedial and orbital areas interact with the hippocampus during offline replay and during on-task periods to guide decision making?

Glossary

- **Actor-Critic:** A policy-based RL algorithm that learns the reward-maximizing probability of choosing among possible actions in a given state of the task (the actor). The model also learns an estimate of the state's value, independent of which action will be chosen, and uses it as a learning signal to optimize the policy (the critic).

- **Convolutional layer:** The building block of convolutional neural networks (CNNs). Each node receives input from a small set of spatially confined nodes (receptive field). With network training, the restricted connectivity leads to nodes acting as filters which detect a specific input feature within their receptive field. Convolutions are typically applied over successive layers, allowing the network to form more complex filters.

- **Deep reinforcement learning (RL) models:** Deep neural network models trained with reward signals, instead of supervised teaching signals. This fusion integrates the representation learning abilities of deep learning with the decision-making abilities of RL, and allows powerful machine-learning solutions to real life tasks such as autonomous driving.

- **DQN (Deep Q Network):** A value-based deep RL model which receives inputs and maps them to the values of possible actions, with each action being an output node of the network.

- **Objective function:** The function the model is trained to minimize, usually expressed as the difference between a model prediction, and a target, i.e. what the model should have predicted. For example, a deep RL model can be trained using a Q-loss function, in which the output nodes are trained to match the current and future reward resulting from each action in a given state. In supervised neural networks, the most common objective function is cross entropy.

- **Fully-connected layer:** Usually contrasted with **convolutional layer**, a fully connected layer is composed of artificial units which are all connected to each other by adjusted weights.

- **Recurrent layer:** A recurrent layer holds previous observations in its memory and allows them to shape its responses to current inputs. This can be especially useful for solving partially observable tasks, where the estimation of the current state of the world cannot be fully determined by the current sensory inputs. Recurrent layers are usually contrasted with feed-forward layers which process each observation independently of previous ones.

- **Task-State:** Collection of observable and non-observable information necessary to predict decision outcomes.The transitions between task-states constitute a Markov Decision Process that allows RL algorithms to solve the reward maximization problem.

Acknowledgements

We thank Chris Summerfield, Ondrej Zika and Alexa Ruel for valuable comments on this manuscript. N.W.S. is funded by a Starting Grant from the European Union (ERC-2019-17 StG REPLAY-852669) and the Federal Ministry of Education and Research (BMBF) and the Free and Hanseatic City of Hamburg under the Excellence Strategy of the Federal Government and the Länder. S.G is funded by the Zuckerman-CHE female postdoctoral program.

General Discussion

This thesis examined the bidirectional effects of task structural knowledge and value representations and their joint influence on guiding behavior. Throughout three projects, it explores how task and state representations guide value and behavior, how value triggers representational change of cognitive maps and states within, and how all these factors might integrate into representational spaces to guide behavior.

Summary of the different projects

The first project asked how contextually irrelevant values influence behavior and neural activity in vmPFC. When participants were faced with decisions between choices where only one context predicted the outcome, they still reacted slower when the irrelevant context favored a different choice and faster when it favored the same. This congruency effect increased with increasing reward associated with the hypothetical choice in the irrelevant context. This shows that participants fully processed both contexts and simulated the potential outcome, even though one context was explicitly cued to ignore. In the fMRI analyses, we isolated vmPFC voxels which were scaled by the relevant expected value, and found that the same voxels were also sensitive to the irrelevant context and its expected value; Moreover, higher irrelevant expected values, or a stronger neural representation of them, impaired the relevant expected value signal. The neural representation in vmPFC during choice was conflicted between the potential values, akin to a representational conflict between the two contexts and their associated values. We also found a clear signal representing the task state (context: motion or color, irrespective of value) in the same value-sensitive voxels. We then showed that a stronger context signal not only related to a stronger task-relevant expected value signal but also diminished the previously observed competition between the relevant and irrelevant expected values. All the neural findings in vmPFC, namely the relation between expected value and context and evidence for value competition, were linked to participants' behavior in a manner generally consistent with the idea that the representations of the alternative/irrelevant context and its associated value were present within vmPFC and guided behavior. The strength of these 'irrelevant' representations within

vmPFC was related to slower and less accurate choices when the different contexts implied different actions, and faster and more accurate choices when they agreed on the action to be made. Brought together, the first project shows that task-irrelevant values are not filtered out during context-dependent decision-making and that state representations are likely the governing framework to determine which values should guide behavior.

In the second project, we asked how exposure to reward influences broader task representations. Participants were tasked with making similarity judgments between sets of trees that differed by the amount of leaves and/or fruits. Then, in a separate task using the same stimuli, they learned that specific trees were more rewarding than others, akin to a reward located in a two-dimensional spatial map. Following reward learning, participants repeated the same similarity judgment task. Participants increased perceptual discrimination among previously rewarded trees, as predicted, shown by accuracy and behavioral models. Based on previous theoretical work, we hypothesized that this increase in acuity would be a result of an increase of representational fields around the reward that would come at the expense of other areas of the map, such that the reward pulls representational fields towards it, akin to a gravitational pull. Behavioral analyses showed the effect of reward on participants' choices generalized to areas of the map that were never directly rewarded, aligning with the pulling hypothesis. Specifically, participants judged trees in areas believed to have a decrease in field density as more similar. At the same time, we observed a decrease in hippocampus univariate signals in the same locations participants judged as more similar, further supporting a decrease in field density in reward-adjacent areas. Multivariate analyses revealed an increase in similarity among trees very close to the reward location in both the hippocampus and mOFC, a representational change also predicted by the pulling hypothesis, according to simulations. Representational changes in the hippocampus were related to participants' general learning of reward across multiple days, whereas changes in mOFC were linked to more recent, immediate experiences with reward on the same day of the task. Taken together, these findings suggest that reward not only causes localized changes to an abstract cognitive map but rather induces a systematic change by pulling representational fields across the

map toward it.

In the third project, we took a broader perspective and reviewed recent literature to explore how value expectations guide humans and animals to achieve goals in complex environments. Our focus was on the vmPFC and OFC, crucial regions for value processing and integrating information for learning. We show support for a reinforcement learning perspective, emphasizing the importance of context as the state of the environment for predicting rewards, strengthening the idea of OFC/vmPFC as supporting cognitive maps in guiding decisions. This emphasis challenges earlier notions of a task-independent ”common currency” as the main role of these regions. We showed that vmPFC and OFC encode multifaceted information, integrating value and task states for decision-making, and suggested that mixed selectivity potentially allows for versatile coding, enabling abstract representations.

Brought together, the first project showed how task structure guides value representation and the influence its neural signature, state representation, had on value signals. The second project showed that value generalization influenced the entire cognitive map and its neural representation by causing systematic changes to other task-relevant dimensions. In the last project, we reviewed recent literature to try and reconcile how values could be represented in the decision-making representational space and what guides this representation in OFC/vmPFC.

Task structure and states govern value representations

The main finding of the first project was that the expected value of each context influenced vmPFC representation and choice behavior. An alternative could have been that the irrelevant feature of the same object would have guided behavior but it was not the case, neither behaviorally nor in vmPFC. It is as if while choosing between two apples for snacking, focusing on the texture, the value associated with the best apple for baking interfered with the decision, even if that was the apple that was not chosen. This finding further reinforces that participants fully processed and simulated the potential alternative choice, holding representations of both in mind. Such simulation is strongly dependent on the task structure, suggesting it is the context, or states, that govern which values would guide participants’ choices.

One assumption we made in the first project was that knowledge of values would remain stable during the main task. This assumption was supported by the fact that participants learned the value-feature mapping before the main task in multiple separate sessions and performed near-ceiling in identifying the most rewarding choice. Therefore we didn't model any 'learning' or changes to subjective values across time during the main task. This was also the motivation to focus on choice deliberation in the fMRI analysis, i.e. when participants needed to sort out the perceptual input and make a choice. However, if irrelevant features are not fully filtered out during choice deliberation, then they might also interact with the processing of the subsequent outcome. In reinforcement learning, identifying the right features to associate with an outcome is known as the credit assignment and is crucial in determining the correct prediction error signal to update cashed values (e.g. Asaad et al., 2017; Sutton & Barto, 1998). Imagine a decision between yogurts that differ based on two features: shape and taste. One is a strawberry round-packaged yogurt and the other is a squared-packaged pineapple yogurt. As a strawberry enthusiast, your choice to maximize potential outcomes would likely be the strawberry. However, what would happen to the value of the pineapple flavor after *not* choosing it? Humans do maintain value information of unchosen options (Boorman et al., 2011; Hayden et al., 2009; Tobia, Guo, Schwarze, Boehmer, et al., 2014) and even update those options based on the obtained reward (Biderman & Shohamy, 2021). This suggests that the positive experience with strawberry flavor might not only positively reinforce choosing strawberry in the future, but also decrease the likelihood of choosing the pineapple option, even if strawberry is not available anymore. Our findings raise the question if these reported spill-over effects also generalize to contextually irrelevant features. For example, after a good experience with a strawberry-flavored square-shaped yogurt, we could find ourselves on future visits to the supermarket gravitating towards squared products, mistakenly reinforcing the irrelevant feature (shape) along with the relevant one (taste). We are currently working on a follow-up project re-analyzing the same behavioral data to investigate if also task-irrelevant features generalize over time to future decisions. One potential mechanism through which such interference could occur might be by

causing false value updating via contextually irrelevant prediction error signals. This line of work could hopefully shed light on how different value functions interact within the task representational space not only when needing to identify the right one, but also when these interferences generalize over time based on recent experiences.

Are neural representations sub-optimal at separating different value functions?

One important question the first project raises is why different value functions interact if one is not predicting the current outcome. Is it a bug in the system or is such contextual spill-over beneficial? In our experiment, we (artificially) imposed full independence of the two contexts such that the irrelevant context never influenced the outcome. Not only have we told that to participants explicitly, but we also made sure outcomes are deterministic (e.g. choosing blue when relevant always gave 70 points). We reasoned that lack of stochasticity would prevent potential inference of hidden rules (e.g. accidentally inferring from experience that relevant blue feature, when paired with diagonal motion, provides a higher outcome). However, in daily life, different contexts rarely exhibit such independence, where two feature dimensions are not only orthogonal, but one is completely relevant in one context and completely irrelevant in another. Even the example of apples previously presented is not ideal. In this example, we proposed that texture is a crucial feature for snacking, but the color might also have an objective effect on the expected outcome of the same experience of snacking. Moreover, in many cases, the dimensions themselves are not fully orthogonal. For example, maybe out of the potential apples available in your supermarket, green apples are often the crispiest. This suggests that reward-predicting feature dimensions might be more related in real life, as opposed to the way we set up the design. Although the lack of ecological validity might be considered a limitation of our study, this perspective suggests that the observed interaction of different value maps might be a more ecologically valid explanation of the findings. Our results can then be interpreted as incorporating into the representational space dependencies, and accompanied uncertainty, that is learned through daily experiences.

Another related aspect of our task was that decisions were partially-observable, i.e. not all the

information needed for the decision was present at the time of choice. In the third project, we discuss the importance of accounting for partial observability in the representation of decisions in more detail. Partial observability inherently introduces uncertainty about the state. To try and mitigate this uncertainty, we designed the experiment to have a full separation of state reward predictability, i.e. only one context-predicted reward at each trial. In real life, however, we often want to optimize for multiple goals at the same time. We might want to buy the best apple for baking a pie, but if we don't have the time to bake, we might end up snacking it. This suggests that even when one is certain in the current state, taking other potential states might prove to be beneficial. In our task, it was beneficial to have a complete representational separation between the context and ignore the 'irrelevant' context. However, in a dynamic and complex world, it might, more often than not, be beneficial to consider all potential states while making a decision. This suggests that even if features are indeed 'irrelevant', taking them into account while making a decision might not have been a 'bug' in the system, as our experimental design suggests, but rather a 'feature'.

Supporting the interpretation of decisions as an integration of all potential contexts, we found that the irrelevant context did not only make participants slower when the trial was incongruent but also made them faster when trials were congruent. This increase in speed went beyond the reaction time of single-feature trials, where the irrelevant context was absent. This hints at the possibility that the two contexts integrated into a single choice, rather than one only interfering with the other. One limitation of our study was that we excluded trials in which the irrelevant context was present, yet did not indicate a preferred choice, i.e. when both irrelevant features were the same. Comparing this condition to single-feature trials could have further shed light on the representation of the irrelevant context in the guiding representations in the lack of competing action signals. If the irrelevant context is 'only' interfering and not integrating, then we should not have seen a difference between single-feature trials and these trials. Concurrently, comparing the representation of the irrelevant context in the lack of competing actions might have influenced the level to which the irrelevant context is actively suppressed as it does not pose a competition to

guide behavior (see below for a detailed discussion on suppression of alternative maps). Brought together, the findings of the first project bridge the gap between state and value representations found in the vmPFC and OFC. We suggest that values are not only state-dependent but that their representations are intertwined such that, when facing multiple potential value-predicting features, the state signal organizes and governs which would guide behavior. At the same time, state separation is not fully intact, as multiple potential states are co-activated and the strength of each state relates to which would guide behavior. Such imperfect separation, however, might be beneficial in more ecologically valid daily situations.

Value generalization triggers representational changes to cognitive maps

The second project took a different angle on the interaction of rewards, states, and perceptual features and asked how values influence task representations and perceptual processing. As mentioned before, reward is often referred to as a tangible experience, such as the pleasure derived from eating chocolate. Value expectations, however, often encompass a broader interpretation, integrating different elements such as personal preference or long-term goals. One way to think of expected values is as a generalization of related reward experiences over time. Whereas in the first project, we demonstrated how learned value functions influence one another and are integrated given the states in the task structure, in the second project, we investigated whether generalizing reward experiences integrate into the task structure and the underlying representations of a cognitive map.

In the pursuit of understanding how tasks are represented in the brain, disentangling value and task structure was the goal of many studies, including, to some extent, the first project. As debated in the third project, classic neuroeconomic studies usually emphasize value signals as task-independent ’common currency’ whereas studies grounded in reinforcement learning frameworks highlight the importance of task structure and its governing power over values. Two prominent frameworks in reinforcement learning that grew out of this peruse are model-based decision-making and successor representations (Daw et al., 2011; Dayan, 1993; Momennejad et al., 2017). In a way, both approaches have a focus on decoupling environment dynamics from

reward structures. The advantage of these approaches is that they enable dynamic switching between value functions without requiring a complete restructuring of environmental knowledge. This is especially beneficial in environments where rewards are stochastic. For example, they allow us to generalize structural knowledge from one supermarket to another, such that when we arrive at a new supermarket we have never seen before, we can navigate between the aisles effectively even if we never saw where products are placed. In the second project, however, we found that exposure to reward altered the representation of the cognitive map and generalized to tasks where reward was not present anymore. This suggests that, at least in some cases, rewards and task structures exert bidirectional influences and dependencies and might not be as easy to disentangle from one another. While dynamically adapting the representation according to the experience of rewards could be expected from the OFC (as detailed in the 3rd project), the hippocampal formation is often believed to hold a more stable representation over time (e.g. Gonzalez et al., 2019) that can be generalized across tasks (e.g. Baram et al., 2021).

One question brought by the dependency of reward and task structure relates to the requirement for reward to cause such a structural change. One aspect of our task was that the reward was roughly deterministic and stable over long periods (i.e. two full MRI sessions). Unlike stochastic rewards, where approaches as successor representation are beneficial, under these conditions, separating task structure and value might not be optimal and reward location might cause a permanent change to the structure of the environment. For example, if we frequently buy chocolate and learn over time that all types of chocolates are always positioned near the checkout, we may gradually increase the resolution of our representation of areas close to the checkout. This increase in resolution could be beneficial to finding our favorite chocolate among the options. The stability in the experience of finding our favorite chocolate might result in generalization to visits to other supermarkets. However, if the supermarket frequently changes the location of sweets, it might not be beneficial to prioritize resources toward any specific area and not generalize such changes to new environments. This idea is to some degree supported by recent findings that grid fields shift their receptive fields primarily when reward is consistently located in one place

(Boccara et al., 2019). When additional sparse rewards were interleaved, however, grid fields increased their firing rate rather than shifting their receptive fields (Butler et al., 2019). At the same time, the hippocampus only showed an increase in place cells around reward when the reward needed to be inferred, compared to explicitly cued (Dupret et al., 2010), which could hint that reward-related representational changes are not only a reaction to experiencing reward, rather a result of generalizing values under partial observability to guide future behavior. These finding potentially suggests that hippocampal formation representations need a stable constant reward in the environment to change their representations. This interpretation is also partially supported by our finding in the second project that hippocampus changes were related to across-day performance matrices and not to recent same-day experiences. However, defining what is considered a ’stable’ reward location can be challenging. Two recent studies suggest that within the same task, hippocampus place cells can shift towards the reward location even on very short periods of a few trials (Sosa et al., 2024; Tessereau et al., 2025). At the same time, both studies showed that this dynamic adaptation was stronger when reward-switching was more predictable. This suggests that the predictability of change might play a direct role in how fast representation in the hippocampus will change. The idea that representational change depends on more abstract concepts such as predictability or volatility is in line with recent ideas of the hippocampus’ role in learning meta-representations to navigate under uncertainty (Ambrogioni & Ólafsdóttir, 2023). Since the reward in our task in the second project was constant (within each group), with very low noise introduced, we will likely not be able to directly examine the role uncertainty plays in the hippocampus representational changes, leaving this interpretation mostly speculative.

Another, more nuanced, question this dependency raises is if the change is true to the entire cognitive map, as potentially represented in the medio-temporal and medio-prefrontal network (Behrens et al., 2018; Schuck et al., 2018; Stachenfeld et al., 2017), or could it be an ‘earlier‘ change to the input of the system, e.g. in how the visual cortex represents the perceptual features. Reward-induced changes in the visual cortex have been observed before (e.g. Schaffner et al., 2023) and to the best of my knowledge, work that found reward effects on place and grid cells in

the hippocampal formation did not simultaneously measure signals from the visual cortex. Notably, even if the change originates on the ‘lower‘ perceptual level, those are still task dimensions that are, in principle, independent of reward when it is absent and would still point to an influence of reward on the formation and structure of the cognitive map. In future work on the second project, we hope to look at sub-divisions of the visual cortex to potentially shed more light on the low-level perceptual processing contribution to the interaction of perception, value, and cognitive maps.

Related to this, another important anatomical subdivision relates to value, states, and cognitive map representations in the prefrontal cortex. As mentioned in the third project, there are still many open questions regarding the distinction between mOFC and vmPFC, especially when comparing findings across species. When considering representations related to cognitive maps, this distinction becomes even more important as the findings of grid-like coding in humans were found in a much more dorsal region, which might overlap with some definitions of vmPFC but is usually distinct from the mOFC (e.g. Bongioanni et al., 2021; Doeller et al., 2010; Padoa-Schioppa & Assad, 2006; Schuck et al., 2016). Future work could further investigate the functional role of these regions, ideally guided by their anatomical and functional connectivity structures and potential link to the medial temporal lobe.

Brought together, hippocampus representations can show short-term flexibility to encode salient locations, potentially especially when changes are predictable. Stability and predictability may play a role in generalizing these representational changes to new environments. Future work on the second project, e.g. analyzing fMRI data from the reward learning session, and a deeper look into further anatomical sub-regions of the visual cortex and the hippocampal formation would hopefully shed more light on some of these important questions.

vmPFC/mOFC: The intersection of values, states, and cognitive maps

In the last project, we dove into the underlying neural code in the OFC and adjacent vmPFC, bridging findings of value and states co-existing and interacting within their neural code. We support previous findings that the OFC represents states (Niv, 2019; Schuck et al., 2016). As

discussed above, states can be conceptualized as a location within a cognitive map, incorporating structural knowledge and previous experiences with current observations. If the hippocampus is to represent a somewhat more stable representation of the environment, mOFC might exhibit more dynamic representations as it needs to take a lot more changing factors into account. Our findings in the second project also partially support this intuition as mOFC changes were related to same-day recent experiences. This is supported by other recent work which suggests that OFC dynamically updates hippocampus representation to optimize future choice inference (Garvert et al., 2023).

One aspect of OFC's neuronal coding that might support such dynamics is mixed selectivity. As discussed in detail in the third project, single neurons in OFC are characterized by mixed selectivity to mixtures of value, outcome, task state, and other variables. Such complex representational space does not mean that single representational dimensions (such as value) can't be read out from it, but it does emphasize the possibility that different aspects of the task might be strongly connected. If mixed selectivity is part of binding perceptual inputs, value, and memory, then it might potentially facilitate learning in which states or locations of the map strongly co-vary with value and contribute to goal-dependent representational changes. One example of such representational change is presented in the simulations in the second project, suggesting that over time, reward acts as a gravitational pull within the cognitive map, changing representational axes that were previously unrelated to reward. Mixed selectivity might play a role in the dynamics of learning and changing task structure representations over time (see more concrete suggestions below).

Compression and competition of alternative goals within the cognitive map

In the first project, we saw multiple states co-activated within vmPFC. Although we saw evidence for simultaneously processing relevant and irrelevant context and their associated values, there were still differences in *how* they were represented within the vmPFC. Evidence for the irrelevant value in the frontal context was only found in the multivariate code, with no univariate effect. This suggests that task-irrelevant variables are represented in a compressed form which is not

easily detectable using univariate tests. At the same time, as briefly mentioned in the third project, some studies indeed found linear scaling with task-independent value-like signals, such as the beauty of faces or more general likability (Abitbol et al., 2015; Frömer et al., 2019; A. H. Harvey et al., 2010; Lebreton et al., 2009; Levy & Glimcher, 2011). This raises an interesting conundrum, namely why are some task-irrelevant aspects more compressed than others? And specifically in our task, irrelevant values were so compressed such that they have no linear trace. One aspect of our task is that contexts changed frequently, making relevant values in one trial irrelevant in subsequent trials. We also made sure that in one-third of the trials, the contexts pointed at different options (incongruent condition), further emphasizing the conflict between the contexts. One potential speculation is that in our task, the alternative context needed to be actively suppressed in order not to interfere with choice, which could explain why we could not detect univariate evidence for the irrelevant values. Previous findings showed that the degree of neural compression of irrelevant features in vmPFC predicted the ability to better perform in context-dependent decisions (Mack et al., 2020) and could suggest that the level of compression might relate to task demands. A complimentary point of view is that dimensions such as aesthetics or general likability which were observed in previous studies are so often trained in daily life, that they are automatically activated in an uncompressed manner (e.g. Lebreton et al., 2009). This would be in line with the hippocampus representing semantic relations between objects even though those were irrelevant for the task at hand (Zheng et al., 2024), since also here they were not posing a competition on which is the right choice at the tasks at hand. Future work could try to better understand under which conditions and to what extent task-irrelevant values (and states) are compressed.

Binding it all together: integrating goals and values in dynamic cognitive maps

One question that relates to all three projects is what is the representational space required to guide goal-directed behavior? Both the first and second projects incorporated tasks with different goals that often pointed to opposing actions. In the first project, those were color versus motion value mappings, whereas, in the second project, it was judging perceptual similarity versus

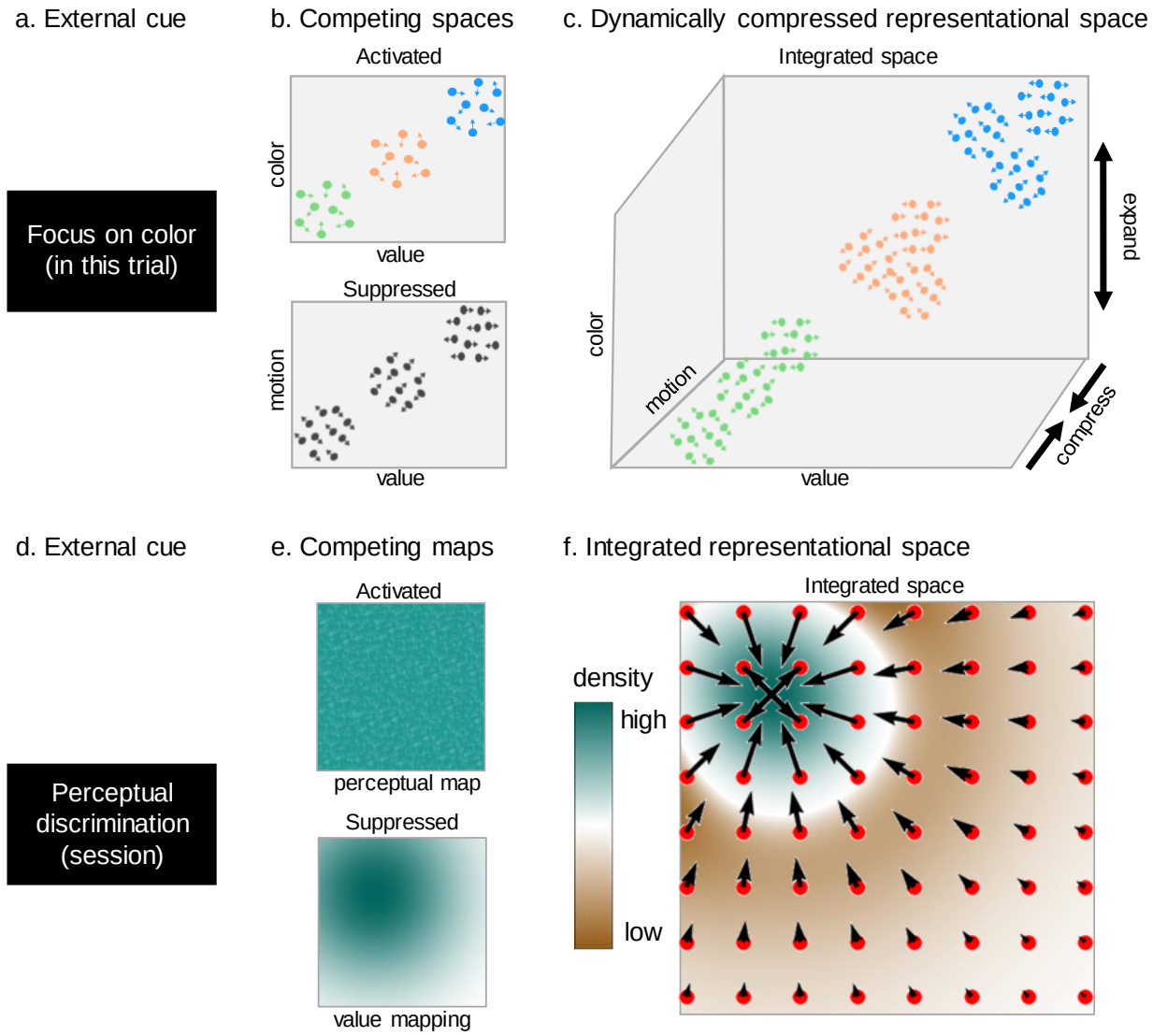


Figure 21

Potential mechanisms for competing goals and their representational spaces ***a*** *In the first project, a contextual cue appeared in every trial instructing participants which context to follow. The context changed every 4-7 trials.* ***b*** *Two separate potential spaces for the task(s) presented in the first project, each one mapping either color (top, y-axis) or motion directions (bottom, y-axis) by their values (x-axis in both). Each stimulus can be conceived as a state in this space. To achieve the right goal in the existence of two separate maps, on each trial one of the maps is activated and the other is suppressed.* ***c*** *Dynamically compressed representational space for color (z-axis), motion-directions (y-axis), and values (x-axis), co-represented within the same space. To achieve the right goal, on each trial, one of the axes is expanded and the other is compressed. In this toy example, linear value-guided relations between the different motion states remained intact only for green, whereas the blue states only kept a separation of the highest rewarding feature (horizontal motion).* ***d*** *In the second project, the goal was presented only once at the beginning of the session and remained constant throughout.* ***e*** *Two separate spaces for perceptual similarities (top) and mapping values (bottom). To achieve the right goal, in the case of two competing separated maps, on each trial, the perceptual map is activated and the value map should be suppressed. We currently can not find evidence for value mapping, pending further investigations.* ***f*** *Integrated representational space for perceptual features, after the exertion of reward-driven gravitational pull (conceptually illustrated by red points and black arrows). Unlike panel c, here the value is not an independent axis but rather integrated into a two-dimensional perceptual space.*

identifying a specific rewarding tree. Both projects also incorporated partial observability, in the sense that the context was not present while choosing, but rather before making the decision, which might encourage co-representations of multiple potential states. In the third project, we proposed that the OFC represents a multidimensional space of all *task-relevant* information, along multiple axes with value being one of them. In both projects, we found evidence of one opposing goal interfering with the other, although in a slightly different manner. The existence of simultaneously opposing goals raises the question if those goals are represented in separate and competing representational spaces, or if they co-exist in one dynamically changing space.

Focusing on the first project, one interpretation of the findings is that there exist two separate representational spaces: a color space and a motion space. Given a specific contextual cue (e.g. ’focus on color’, Fig.21a), one map might be activated whereas the other suppressed (Fig.21b). Discussed findings that the hippocampus maintains task-irrelevant semantic relationships and the mOFC representing commonly used task-irrelevant attributes support the existence of multiple co-activated maps (Garvert et al., 2023; Lebreton et al., 2009). As mentioned previously, our analyses suggest that when states switch from relevant to irrelevant their representation is compressed. Such compression could be a result of suppressing a separate map. An alternative perspective is that the representational space incorporates *all* presented information, and dynamically compresses feature dimensions depending on the current goal (Fig.21c). This perspective is not in contradiction to the conclusion of the third project but rather expands the interpretation of ’relevance’ by emphasizing that such a space might preserve more information than what is immediately goal-related. Similar to the line of argument in the third project, a read-out model could still extract the two separate maps from the full space by selectively integrating over the irrelevant axes (from Fig.21c to Fig.21b), or even further integrate to show a single dimension of value representation (as suggested by the ’common currency’ framework). However, considering the space as inherently one opens the door to phenomena that can not be explained if representation is separated. For instance, this perspective allows us to discover through experience co-variation and dependencies between dimensions previously assumed to be

independent and even assign values to task aspects that were never goal-related (e.g. Ben-Artzi et al., 2023; Shahar et al., 2019). This perspective still leaves many open questions on what governs compression, what remains intact after such compression, and what forms such compression can take (see examples in Fig.21c).

The findings of the second project provide a different perspective on the co-variation of task axes and hint at different types of representational change. Also in this project, the goal was not explicitly present during the choice period but was rather presented once at the beginning of the session (Fig.21d), making the task partially observable. This could potentially lead to two competing maps: one *only* representing the perceptual space and one *only* representing potential values. According to this perspective, the observed behavior would then be a result of failure to suppress the reward map (Fig.21e). However, we do not (currently) find clear evidence for univariate or multivariate value mapping, in contrast to the previous project, where multivariate analyses found the representation of the so-called 'suppressed' map. This suggests that in this task, the previously learned value might not be directly present in the representational space. Our findings point to a different form of representational change, namely that generalizing reward knowledge (or saliency) is integrated over time into the perceptual dimensions (Fig.21f).

The discrepancy between the first and second projects might be attributed to the time spent on each task and the frequency of change between the tasks. Whereas in the first project, contexts switched every 1-2 minutes, in the second project, the two tasks were completely separated, sometimes 1-2 hours apart, with no interleaved appearance of the alternative task. The stability of the goal during the extensive reward learning sessions could be a reason for the integration of axes in the hippocampus and OFC representational spaces. It is possible, that the co-variation of reward and perceptual dimensions over a long time enforced gravitational pull on the space, potentially to support optimal reward maximization in the future (Ginosar et al., 2021). It is important to note that none of the projects aimed to test for the effects of goal consistency and did not include controlled conditions, leaving this interpretation more speculative for the time being.

As mentioned in the third project, mixed selectivity might play an important role in the

representational changes of task spaces. This mechanism could potentially support the learning of dependencies between different representational axes. Neurons exhibiting mixed selectivity can encode multiple task-relevant dimensions simultaneously, but their responses dynamically adapt depending on the behavioral context. This property could play a role in shaping how different task dimensions interact over time. If mixed selectivity contributes to dynamically adjusting representations based on task demands, then during the extended exposure to a single goal in the second project, neural populations may gradually refine their tuning to emphasize the most behaviorally relevant features while deprioritizing others. In this case, reward information may have become increasingly integrated with perceptual features rather than being represented as an independent axis. This is supported by the (current) lack of evidence for a separate representation of the value mapping. Correspondingly, when task dimensions are engaged in competition, e.g. such as when one must be suppressed to optimize decision-making in the first project, neurons encoding one dimension may shift their tuning away from the suppressed axis, effectively reducing interference.

Notably, mixed selectivity was found in different areas of the prefrontal cortex, visual cortex, and retrosplenial cortex which is adjacent and connected to the hippocampal formation but not part of it (Kira et al., 2023; Rigotti et al., 2013; Tye et al., 2024). To the best of my knowledge, there is currently no direct evidence for mixed selectivity on the neuronal level in the hippocampal formation. It is possible that if mixed selectivity plays a role in learning co-variation of task dimensions, such learning might happen outside of the hippocampal formation (e.g. the OFC), potentially informing it on such learned dependencies (e.g. Garvert et al., 2023; Schuck & Niv, 2019). As leading candidates to represent task structures and guide our behavior, there is also still much to uncover about the specific roles of hippocampal formation and OFC, how dynamically they adapt to current and past goals, and how they potentially inform one another to guide our behavior.

Concluding remarks

Taken together, the projects in this thesis provide new insights into how cognitive maps flexibly integrate structural knowledge, value representations, and task demands to guide goal-directed behavior. By demonstrating how task state representations influence value signals in vmPFC, how reward reshapes cognitive maps in the hippocampus and mOFC, and how values might be structured in a multidimensional representational space in the OFC/vmPFC, this work contributes to a growing understanding of how the brain dynamically constructs internal models of the environment. The findings challenge simplified notions of value representation and instead highlight the flexible, context-dependent nature of these representations. At the same time, these findings challenge strong separations between value and task structure, especially when values are stable and predictable. By bridging perspectives from reinforcement learning, cognitive maps, and representational geometry, this thesis provides a framework for thinking about how multiple task dimensions, whether perceptual, structural, or value-based, interact over different timescales to guide decision-making. Future research could further explore how these representational spaces evolve with experience, how they are influenced by learning across different temporal scales or uncertainty levels, and how interactions between hippocampal and prefrontal sub-regions contribute to the continuous adaptation of goal-directed behavior.

Supplementary information for 1st project

- Fig. S1: Full procedure and experimental design for all phases, related to Fig 1
- Fig. S2: Nested RT models (EV, Context, Block and switch) related to Fig 2
- Fig. S3: Alternative RT models, extended RT model comparisons and correlation matrix of all regressors, related to Fig 2
- Fig. S4: Exploratory analysis of RT model presented in Main Text, related to Fig 2
- Fig. S5: Behavioral accuracy results: related to Fig 2
- Fig. S6: Frequency bias in the design and supplementary information for Representational Similarity Analysis: related to Fig. 5 and Fig. 3
- Fig. S7: Supplementary information for Value similarity analysis: related to Fig. 4 and Fig. 5
- Fig. S8:Supplementary information for perceptual similarity analysis: related to Fig. 4 and Fig. 5
- Fig. S9: Modelling probability assigned to the EV class: related to Fig. 5
- Fig. S10: Main effects and corresponding data, fMRI effects, related to Fig. 5
- Fig. S11: Main effects and corresponding data, link of fMRI to behavioral accuracy, related to Fig. 6
- Fig. S12: Main univariate results
- Fig. S13: Additional univariate results,
- Table 3: Detailed univariate results: Clusters for whole brain univariate analysis
- Effect sizes and confidence intervals for best explaining models:
  - Table 4: RT model
  - Table 5: fMRI model (P_{EV}, main model)
  - Table 6: fMRI model (P_{EV}, nested in EV_{back})
  - Table 7: RSA model - Main effect models
  - Table 8: RSA model - Main effect models value difference models

Source data for all figures are provided as a Source Data file.

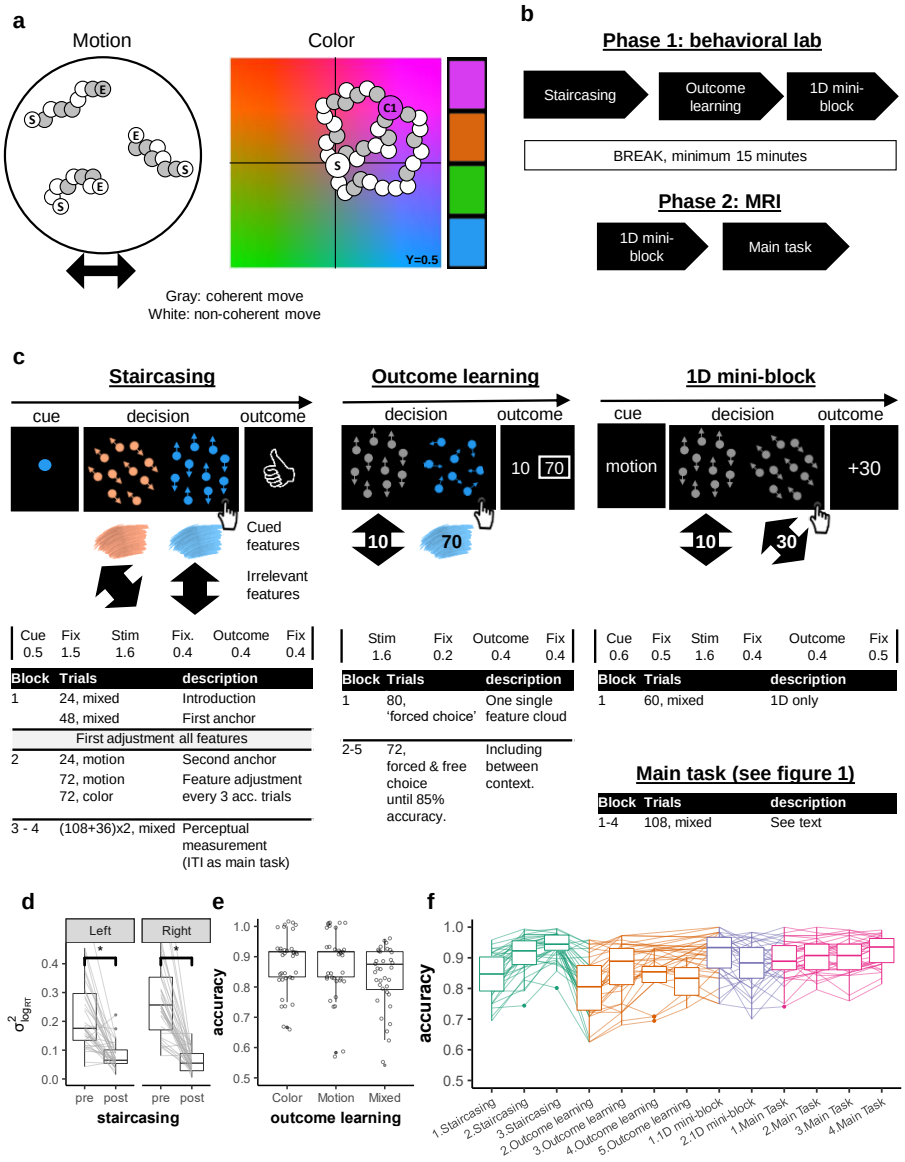


Figure S1
Full procedure and experimental design for all phases

Figure S1: Full procedure and experimental design for all phases, related to Fig 1. a. Brownian algorithm for color and motion. Each illustration shows the course of 3 example dots; 'S' and 'E' marked dots reflect Start and End positions, respectively. Remaining dots represent location in space for different frames. Left panel: Horizontal motion trial. Shown are framewise dot positions between start and end. In each frame, a different set of dots moved coherently in the designated direction (gray) with a fixed speed; remaining dots moved in a random direction (conceptually taken from Pilly & Seitz, 2009). Right panel: Example of a pink color trial. We simulated the YCbCr color space that is believed to represent the human perception in a relative accurate way (cf. Abbott et al., 2016). A fixed luminance of Y = 0.5 was used. For technical reasons we sliced the X-axis by 0.1 on each side and the Y-axis by 0.2 from the bottom of the space to ensure the middle of the space remained gray given the chosen luminance. In each frame,

a different set of dots (always 30% of the dots) moved coherently towards the target color in a certain speed whereas the rest were assigned with a random direction. All target colors were offset by 23.75% from the center towards each corner. Right bar illustrates the used target colors. **b.** Full procedure. The experiment consisted of two phases, the first one took place in the behavioral lab and included Staircasig, Outcome-learning and the first 1D mini-block. The second took place inside the MRI scanner and consisted of the second 1D mini-block and the main task. **c.** Example trial procedures and timing of the different tasks. Timing of each trial is depicted below illustrations. **Staircasing (left)** Each trial started with a cue of the relevant feature. Each cloud had one or two features (motion and/or color) and participants had to detect the cued feature. Participants' task was to choose the cued feature (here: blue). After a choice, participants received feedback if they were correct and faster than 1 second, correct and slower, or wrong. **Outcome learning (middle)** Participants were presented with either one or two single-feature clouds and asked to chose the highest valued feature. Following their choice, they were presented with the values of both clouds, with the chosen cloud's associated value marked with a square around it. The pair of shown stimuli included across contexts comparisons, e.g. between up/right and blue, as shown. **1D mini block (right)** At the end of the first phase and beginning of the second phase participants completed a mini-block of 60 1D trials during the anatomical scan (30 color-only, 30 motion-only, interleaved). Participants were again asked to make a value-based two alternative forced choice choice decision. In each trial, they were first presented with a contextual cue (color/motion), followed by the presentation of two single-feature clouds of the cued context. After a choice, they were presented with the chosen-cloud's value. No BOLD response was measured during these blocks and timing of the trials was fixed and shorter than in the main task (see Main task preparation in methods) **Main task (bottom)** This part included 4 blocks, each consisting of 36 1D and 72 2D trials trials presented in an interleaved fashion (see method and Fig. 1). **d.** Button specific reduction in RT variance following the staircasing. We verified that the staircasing procedure also reduced differences in detection speed between features when testing each button separately. Depicted is the variance of reaction times (RTs) across different color and motion features (y axis). While participants' RTs were markedly different for different features before staircasing (pre), a significant reduction in RT differences was observed after the procedure (post, paired t-test: $p < .001$, N=35) **e.** Choice accuracy in outcome learning trials. Participants achieved near ceiling accuracy in choosing the highest valued feature in the outcome learning task, also when testing for color, motion and mixed trials separately (ps$< .001$, N=35). Mixed trials only appeared in this part of the experiment to encourage mapping of the values on similar scales. **f.** Accuracy throughout the experiment, plotted for each block of each part of the experiment. *In the staircasing (left)* High accuracy for the adjustment and measurement blocks (2-3) ensured that there were no difficulties in perceptual detection of the features. *In Outcome learning* a clear increase in accuracy throughout this task indicated learning of feature-outcome associations. Note that Block 5 of this part was only included for those who did not achieve 85% accuracy beforehand. Starting the *1D mini blocks* (middle) and throughout the*main task* (right) until the end of the experiment high accuracy. μ and σ from left to right: Staircasing: .84,.07;.91,.06;.94,.04; Outcome Learning: .81,.1;.86,.09;.83,.08;.82,.06; 1D mini blocks: .91,.07;.88,.08; Main task: .89,.06;.91,.05;.9,.06;.92,.05.; N=35. In panels d-f boxes mid-line represent mean, lower and upper the 25th and 75th percentile and whiskers extend to the range of the data (no more than 1.5 of the full box range). Data beyond the whiskers are plotted as individual solid points.

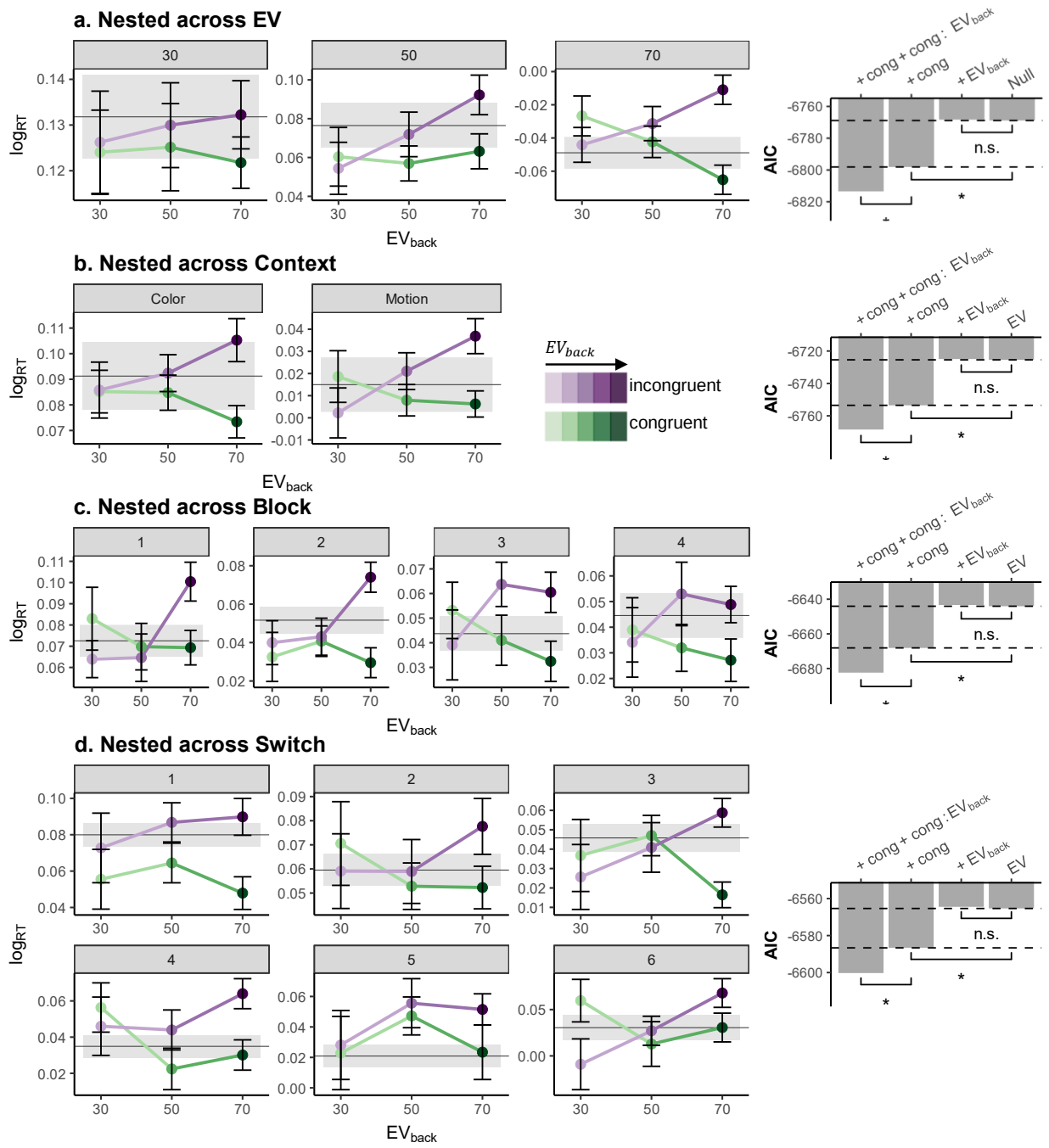


Figure S2
Nested RT models, related to Fig 2

Figure S2: Nested RT models, related to Fig 2
a-d. Nested models within Factors. Each row represents one congruency analysis, done separately for each level of expected value (a, top row), context (b, 2nd row), block (c, 3rd row) or switch (d, bottom row). The RT effect of Congruency $\times$ EV_{back} is shown on the left, corresponding AICs for mixed effect models with nested factors are shown on the right. Mean RT

(line) and SEM (shades) for the corresponding 1D trials is plotted in gray for each panel (e.g. mean across all 1D trials where EV=30 are on top left panel). Error bars assigned to colored lines and gray error band represent corrected within subject SEMs (Cousineau et al., 2005; Morey et al., 2008). Null models shown on the right are identical to Eq. 2, albeit included $\zeta_{0_{k_v}}$, which is the factor-specific (v) intercept nested within each within each subject level (see methods). Likelihood ratio tests were performed to asses improved model fit when adding (1) Congruency or (2) $\mathrm{EV}_{\mathrm{back}}$ terms to the Null model and when adding (3) Congruency $\times$ $\mathrm{EV}_{\mathrm{back}}$) in addition to Congruency. Stars represent p values less than .05. For nested within EV, the Null model did not include a main effect for EV and the likelihood ratio (LR) tests with added term: (1) $\chi^2_{(1)} = 31.22$, $p < .001$; (2) $\chi^2_{(1)} = 1.47$, $p = .226$; (3) $\chi^2_{(1)} = 19.37$, $p < .001$; For models nested within Context the LR test was: (1) $\chi^2_{(1)} = 30.01$, $p < .001$; (2) $\chi^2_{(1)} = 1.5$, $p = .22$; (3) $\chi^2_{(1)} = 18.9$, $p < .001$; For models nested within Block: (1) $\chi^2_{(1)} = 26.06$, $p < .001$; (2) $\chi^2_{(1)} = 1.27$, $p = .26$; (3) $\chi^2_{(1)} = 18.25$, $p < .001$; And for models nested within switch: (1) $\chi^2_{(1)} = 23.29$, $p < .001$; (2) $\chi^2_{(1)} = 1.13$, $p = .29$; (3) $\chi^2_{(1)} = 17.66$, $p < .001$;, N=35 for all panels and models. In the first row (nested across EV) the interaction with EV is visible, i.e. the higher the EV, the stronger our effects of interests were.

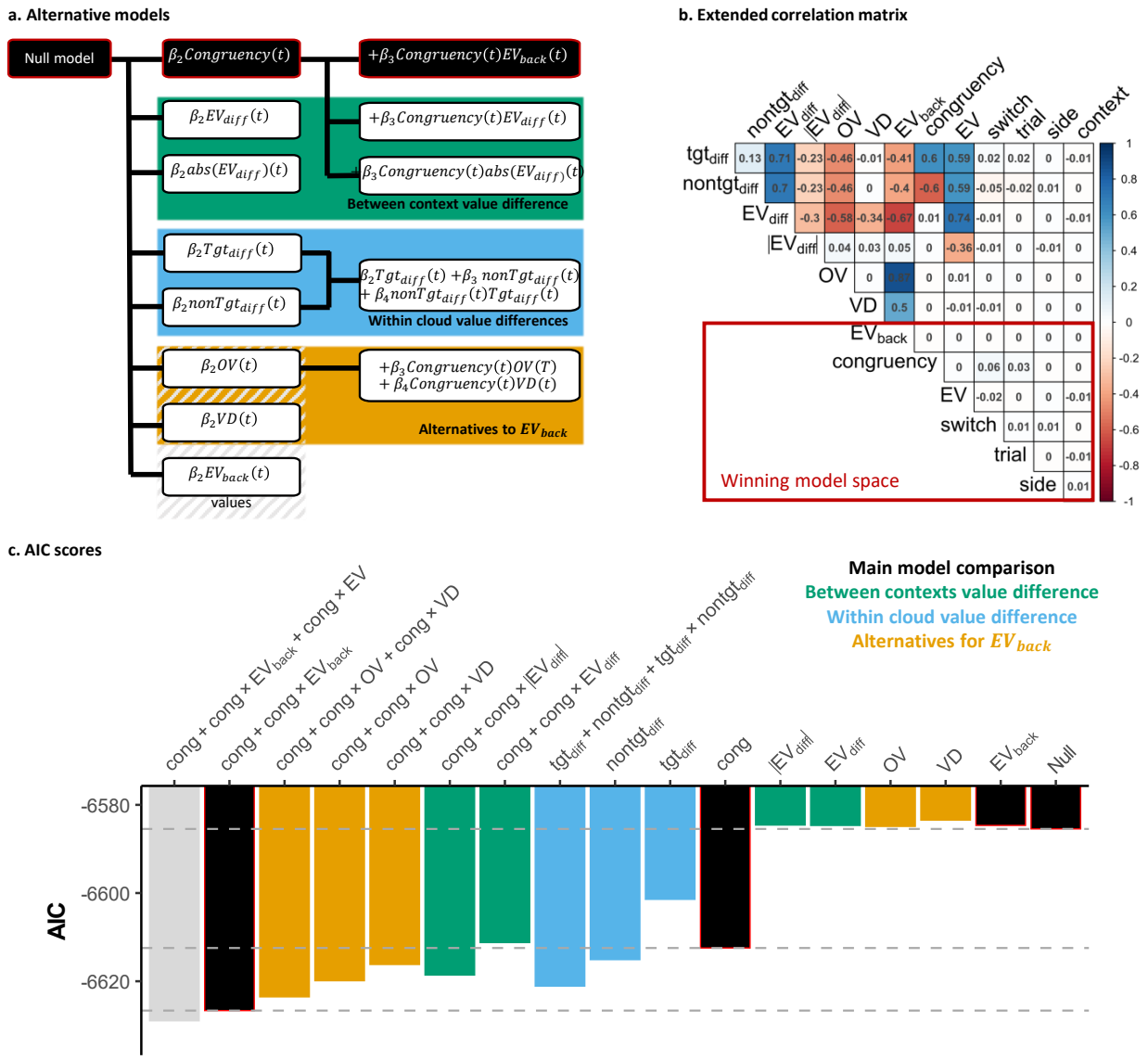


Figure S3
Alternative RT models, extended RT model comparisons and correlation matrix of all regressors, related to Fig 2.

Figure S3: Alternative RT models, extended RT model comparisons and correlation matrix of all regressors, related to Fig 2.
a. Alternative mixed effect models, each represented as a row which lists main factors of interest. We clustered different alternative models into three classes: *Green models* included factors that reflected the difference between the expected values of both contexts (EV - EV_{back}, including unsigned EV factors); *blue* models include instead factor that reflect the value-difference between context within each cloud where 'tgt' (target) is the chosen cloud with the highest value according to the relevant context and *orange* models included two alternative parameterization of values in the non-relevant context: irrelevant features' Value Difference (VD) and Overall Value (OV), which are also orthogonal to Congruency (Cong), and to each other. *In black* is the main model comparison as presented in the main text. **b. Extended correlation matrix**. Averaged correlation across subjects of all scaled regressors for accurate 2D trials (models' input). Marked in red

rectangle are main factors of the experiment which are orthogonal by design and used for the model comparison reported in the Main Text. **c. AIC scores.** We tested different alternatives shown in (a) in a stepwise hierarchical model comparison, as in the main text. Each bar represents the AIC (y-axis) of a different model (x-axis) where the labels on the x-axis depict the added terms to the Null model for that specific model. The Null model included nuisance regressors and the main effect of EV (see ν and β_1 in Eq. 2). The models described in the main text are shown in black. The gray model includes the additional term for Congruency $\times$ EV. Dashed lines correspond to the AIC values of the models used in the main text. Importantly, no main effect representing only the contextually irrelevant values (VD, OV, $\mathrm{EV}_{\mathrm{back}}$) nor the difference between the EVs ($\mathrm{EV}_{\mathrm{diff}}$,$|\mathrm{EV}_{\mathrm{diff}}|$, also when excluding EV from the null model, not presented) improved model fit over the Null model. This supports our finding that neither large irrelevant values, nor their similarity to the objective EV, influenced participants' behavior. Similar to $\mathrm{EV}_{\mathrm{back}}$, factors from the green and orange clusters are also orthogonal to Congruency, which allowed us to test their interaction. Factors from the blue cluster highly correlate with both Congruency (and $\mathrm{EV}_{\mathrm{back}}$) and therefore were tested separately. Non of the alternatives provided a better AIC score (y axis, lower is better).

a

effect	Df	AIC	likelihood ratio test	P value
Switch	1	-6490.88	15.19	<.001
Trial	1	-6499.01	7.04	.008
Context	1	-6304.79	201.27	<.001
EV	1	-6159.75	346.31	<.001
Congruency	1	-6496.76	9.29	.002
Congruency x EV	1	-6501.69	4.37	.037
Congruency x EV_{back}	1	-6488.35	17.70	<.001
Switch x Side	1	-6502.42	3.64	.0565
Switch x Context	1	-6503.25	2.80	.094
Switch x EV	1	-6500.23	5.83	.0158
Switch x Congruency	1	-6502.54	3.52	.068
Context x EV	1	-6499.16	31.07	<.001
Context x Trial	1	-6499.16	6.89	.009
Side x EV_{back}	1	-6501.28	4.77	.029

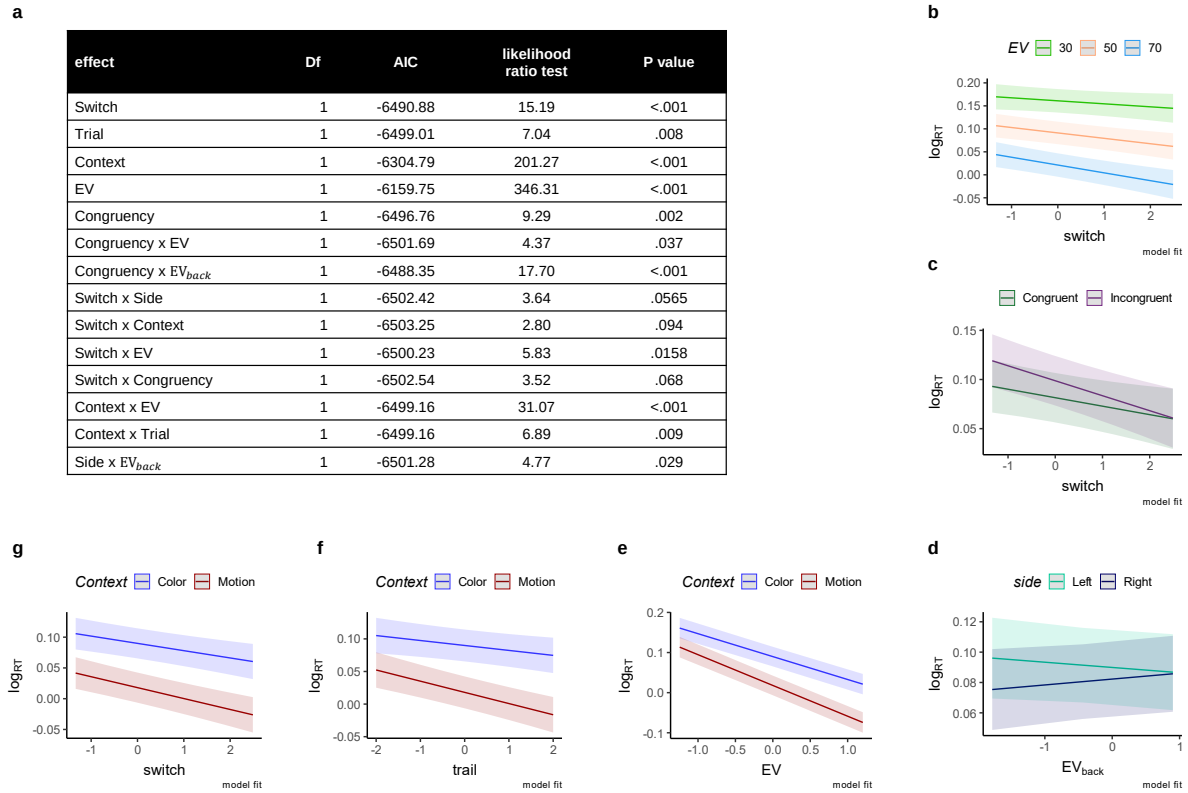
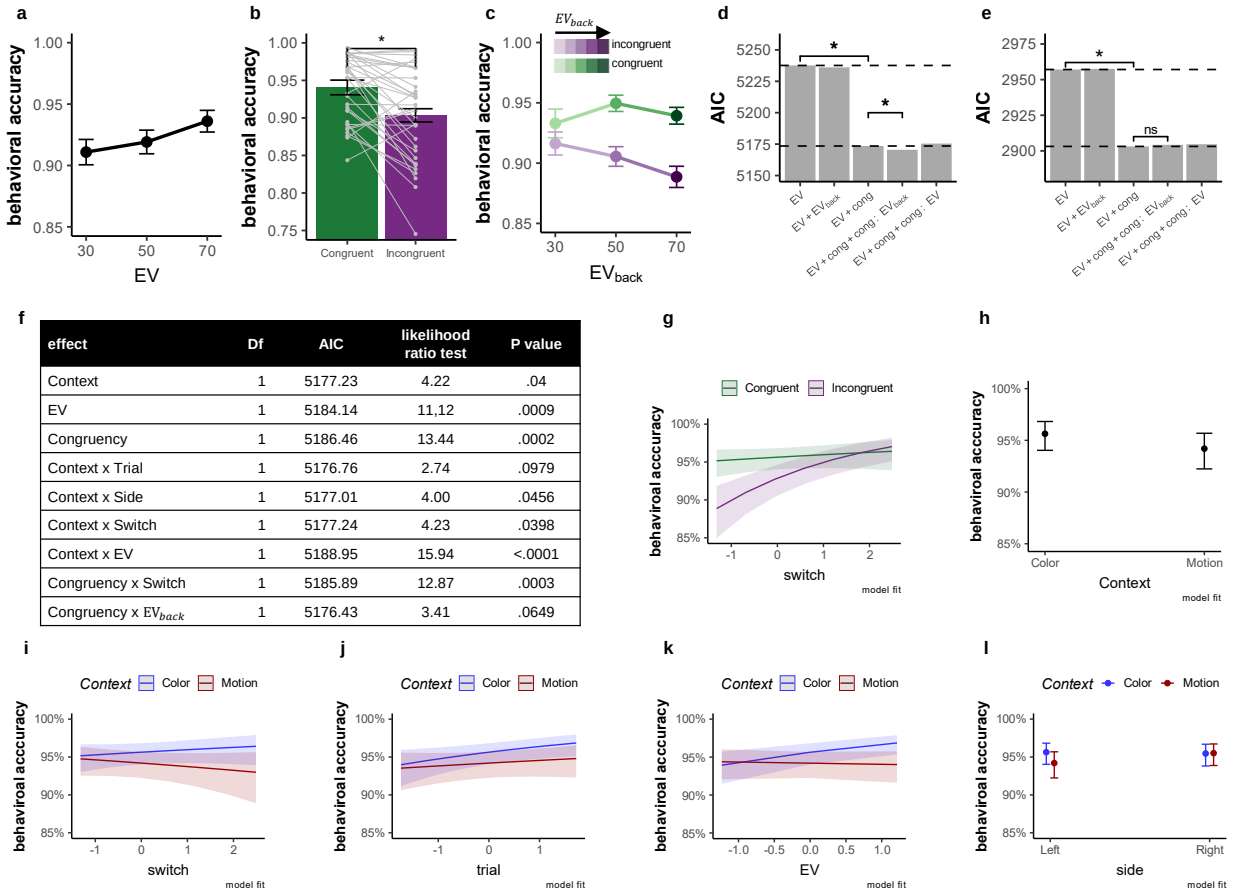


Figure S4
Exploratory analysis of RT model presented in Main Text, related to Fig 2.

Figure S4: Exploratory analysis of RT model presented in Main Text, related to Fig 2. **a.** The table presents the individual contribution of terms taken from Eq. 2 and all possible two-way interactions to the model fit using the drop1 function in R (R Core Team, 2017). In short, this exploratory analysis started with a model that included all main effects from Eq. 2 and all possible 2-way interaction between them and tested which terms contribute to the fit. If a term did not improve fit, it was dropped from the model. Presented are all effects with p value less than $p < .01$ for likelihood ratio test with added terms. Additionally, we specifically tested if the switch interacts with our main effect and found no such interaction (likelihood-ratio test with added term for Congruency x $\mathrm{EV_{back}}$ x switch:$\chi^2_{(1)} = 3.70$, $p = .157$). **b-g.** Model fits of all effects with $p < .01$ for likelihood ratio test with added terms. X-axes are normalized (as in the model) and y-axes reflect RTs on a log scale (model input). Clockwise from the top: RTs became progressively faster with increasing trials since the context switch. This effect was possibly stronger for higher EV (b) and for incongruent trials (c). We note that our experiment was not designed to test the effect of the switch. (d) An interaction of Side and $\mathrm{EV_{back}}$ was found, for which we offer no explanation. Panels (e) to (g) reflect interaction of context with EV (e), trial (f), and switch (g). In panels b-g error bands represent the 89% confidence interval. P values of each effect are found in the table in panel (a). We note that due to the used perceptual color space there might be a context-specific ceiling effect in RTs due to training throughout the task which could have induced effects of context. Specifically, since dots start gray and slowly 'gain' the color, it might take a few frames until there is any evidence for color. However, the motion could be theoretically detected already on the second frame (since coherence was very high). This could

explain why some effects that represent decrease in RT might hit a boundary for color (and not motion). Crucially, we refer the reader to supplementary Fig S2 where the main model comparison hold also when we ran the model nested within the levels of Context.



effect	Df	AIC	likelihood ratio test	P value
Context	1	5177.23	4.22	.04
EV	1	5184.14	11,12	.0009
Congruency	1	5186.46	13.44	.0002
Context x Trial	1	5176.76	2.74	.0979
Context x Side	1	5177.01	4.00	.0456
Context x Switch	1	5177.24	4.23	.0398
Context x EV	1	5188.95	15.94	<.0001
Congruency x Switch	1	5185.89	12.87	.0003
Congruency x EV_{back}	1	5176.43	3.41	.0649

Figure S5
Behavioral accuracy results: related to Fig 2.

Figure S5: Behavioral accuracy results: related to Fig 2.
a. Comparison of accuracy (y-axis) for each level of EV (x-axis) showed that participants were more accurate for higher EV, likelihood ratio test against null model: $p = .001$, N=35. **b.** Comparison of congruent versus incongruent trials also revealed a performance benefit of the former, paired t-test: $p = .001$, N=35. **c.** The effect of Congruency was modulated by $\mathrm{EV_{back}}$, i.e. the more participants could expect to receive from the ignored context, the less accurate they were when the contexts disagreed (x axis, shades of colours). Further investigations revealed that the modulation of $\mathrm{EV_{back}}$ is likely limited to Incongruent trials (likelihood ratio test with added term: $\chi^2_{(1)} = 6.91$, $p = .009$, N=35, when modeling only Incongruent trials), yet does not increase accuracy for Congruent trials (likelihood ratio test with added term: $\chi^2_{(1)} = 0.07$, $p = .794$, N=35, when modeling only congruent trials), likely due to a ceiling effect. Error bars in panels a-c represent corrected within subject SEMs (Cousineau et al., 2005; Morey et al., 2008). **d.** Hierarchical model comparison of choice accuracy, similar to the RT model reported in the main text. These analyses showed that including Congruency improved model fit (likelihood-ratio test with added term: $p < .001$, N=35). Including the additional interaction of Congruency $\times$ $\mathrm{EV_{back}}$ improved the fit even more (likelihood-ratio test with added term: $p = .03$, N=35). **e.** We replicated the choice accuracy main effect in an independent sample of 21 participants outside of

the MRI scanner, i.e. including Congruency improved model fit (likelihood-ratio test with added term: $\chi^2_{(1)} = 55.95$, $p < .001$). We did not find a main effect of EV on accuracy in this sample (likelihood-ratio test with added term: $\chi^2_{(1)} = 0.93$, $p = .333$). The interaction term Congruency $\times$ $\mathrm{EV_{back}}$ did not significantly improve fit in this sample. Modeling only Incongruent trials, as above, reveled that $\mathrm{EV_{back}}$ had a marginal effect on accuracy (likelihood-ratio test with added term: $\chi^2_{(1)} = 2.90$, $p = .088$). Near-ceiling accuracies in Congruent trials in combination with a smaller sample might have masked the effects. **f.** The table presents the individual contribution of terms taken from Eq. 3 and all possible two-way interactions to the model fit using the drop1 function in R (R Core Team, 2017). In short, this exploratory analysis started with a model that included all main effects from Eq. 3 and all possible 2-way interaction between them and tested which terms contribute to the fit. If a term did not improve fit, it was dropped from the model. Subsequent panels present all the effects corresponding to $p < .01$. Note that this is a non-hypothesis driven exploration of the data and that accuracy was very high in general throughout the main task. **g.** Accuracy as a function of time since switch. Akin to RTs, accuracy increased with number of trials since the last context switch, mainly for incongruent trials. **h.** Context effect on accuracy. According to the exploratory model, participants were slightly more accurate in color than in motion trials. However, a direct paired t test between average accuracy of color compared to motion was not significant (paired t-test: $t_{(34)} = 0.96$, $p = .345$, N=35). Error bars represent corrected within subject SEMs (Cousineau et al., 2005; Morey et al., 2008). **i-l.** Depicted are some minor interactions of no interest with Context, according to the exploratory model, N=35 for all panels. Error bars and bands in panels g-l correspond to 89% confidence interval.

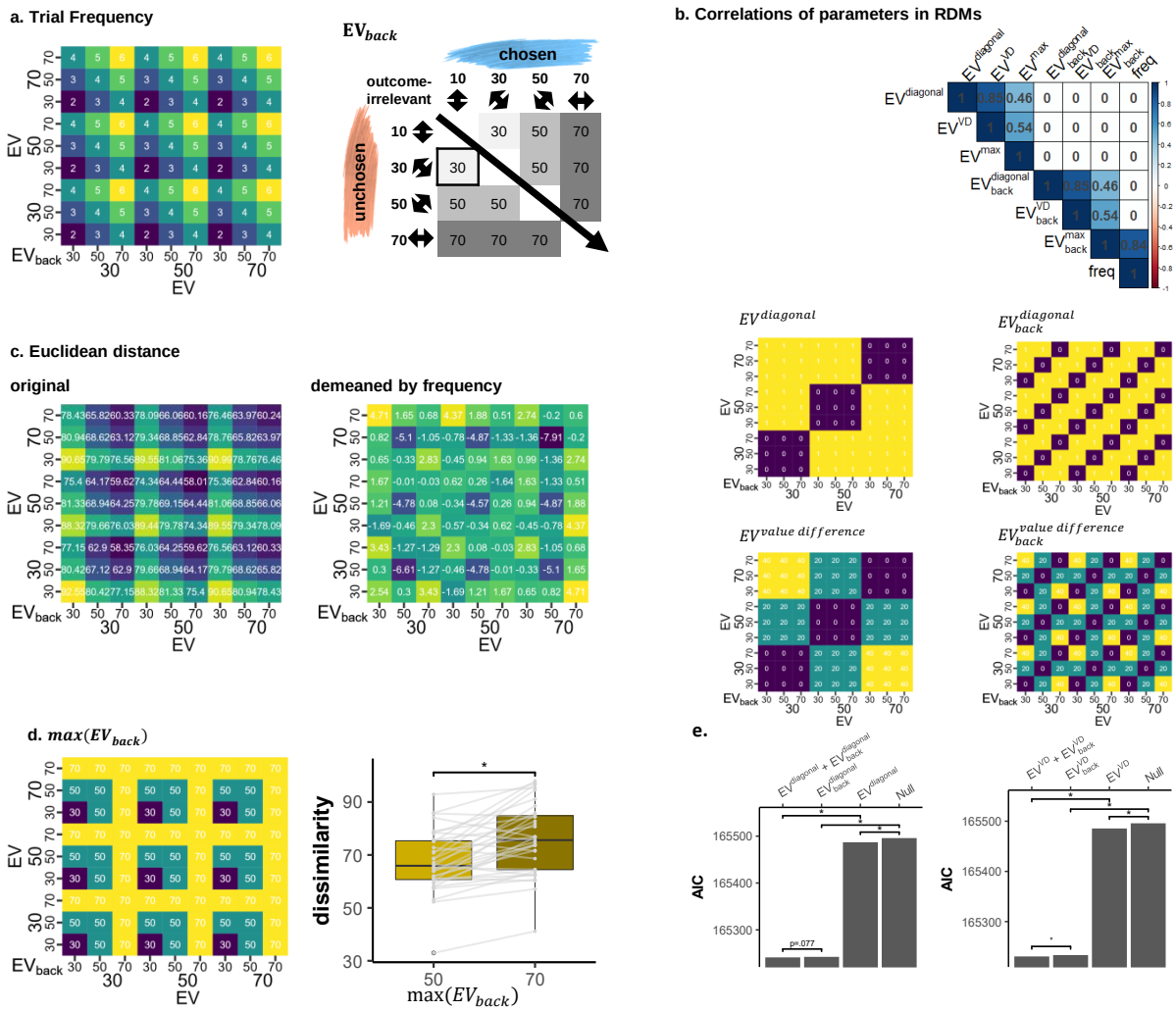
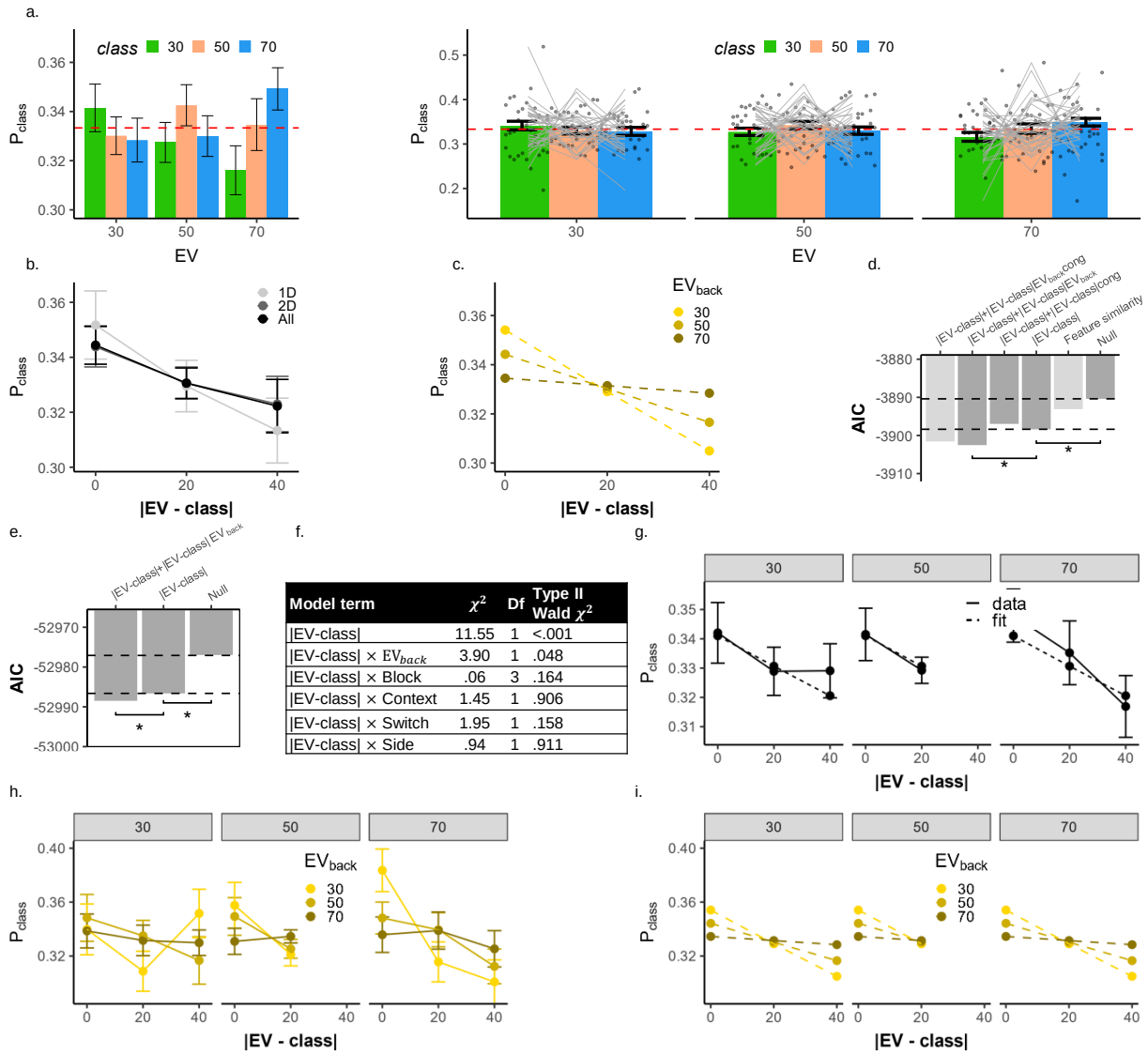


Figure S6

Frequency bias in the design and supplementary information for Representational Similarity Analysis: related to Fig. 5 and Fig. 3

Fig. S6: Frequency bias in the design and supplementary information for Representational Similarity Analysis: related to Fig. 3. a-b Panel a shows the frequency of unique examples within 2D trials (for each context). Panel b is taken from Fig. 1e. to help with visualization. Each cell shows the number of how many trials were used for to both the betas that correspond to that cell (presented as ratio relative to the rest). As can be seen, our design included more trials for higher EV_{back}. We believe this is the reason why the probabilities the classifier trained on 2D trials were biased. Note that the analyses depicted in Fig. 5g-i. was conducted nested within the levels of EV_{back}, thus eliminating influences of frequency of trials (henceforth: Frequency) from the probability of the EV_{back} classifier. Additionally, all RSA models were conducted nested within the levels of Frequency, meaning all effects found go beyond any mean difference resulting from the frequency bias. **c** Correlations of parameters used in the RSA analyses show that all the main and value difference parameters are orthogonal to the frequency effect. Added below the correlations are the effects taken from Fig. 3 to help with visualization. **d.** In order to replicate the effect found in Fig. 5b, when focusing only on the cells corresponding to the same

EV (i.e. corresponding to the diagonal in the EV main effect matrix), only one level of Frequency (4) has two separate levels of $\max_{back}^{EV}$ (parameter indicating which is the maximum $\mathrm{EV}_{\mathrm{back}}$ involved in the comparison, explaining the high correlation in panel c). Nevertheless, when comparing these two cells across subjects we find a positive effect of $\max_{back}^{EV}$ indicating an increase in dissimilarity of EV representation when $\max_{back}^{EV}$ is higher, paired t-test: $t_{(34)} = -5.42$, $p < .001$, N=35. Boxes mid-line represent median, lower and upper the 25th and 75th percentile and whiskers extend to the range of the data (no more than 1.5 of the full box range). Data beyond the whiskers are plotted individually as solid points. **e.** Hierarchical model comparison showing that the model with both Main effects (right) and with both Value similarity effects (left) explain the data best. All models are nested within the levels of frequency (see panel a). Likelihood-ratio-tests with added terms: For Diagonal effects models (left): adding $\mathrm{EV}^{\mathrm{diagonal}}$ to null model: $\chi^2_{(1)} = 10.89$, $p = .001$; adding $\mathrm{EV}^{\mathrm{diagonal}}_{\mathrm{back}}$ to null model: $\chi^2_{(1)} = 255.44$, $p < .001$; adding $\mathrm{EV}^{\mathrm{diagonal}}$ to the model with $\mathrm{EV}^{\mathrm{diagonal}}$: $\chi^2_{(1)} = 3.12$, $p = .077$. adding $\mathrm{EV}^{\mathrm{diagonal}}_{\mathrm{back}}$ to the model with $\mathrm{EV}^{\mathrm{diagonal}}_{\mathrm{back}}$: $\chi^2_{(1)} = 247.67$, $p < .001$; For Value Difference models (VD, right): adding $\mathrm{EV}^{\mathrm{VD}}$ to null model: $\chi^2_{(1)} = 12.34$, $p < .001$; adding $\mathrm{EV}^{\mathrm{VD}}_{\mathrm{back}}$ to null model: $\chi^2_{(1)} = 264.61$, $p < .001$; adding $\mathrm{EV}^{\mathrm{VD}}$ to the model with $\mathrm{EV}^{\mathrm{VD}}_{\mathrm{back}}$: $\chi^2_{(1)} = 4.71$, $p = .03$; adding $\mathrm{EV}^{\mathrm{VD}}_{\mathrm{back}}$ to the model with $\mathrm{EV}^{\mathrm{VD}}$: $\chi^2_{(1)} = 256.98$, $p < .001$. Stars in panels d-e represent the p-value is lower than conventional .05 threshold.



Model term	χ^2	Df	Type II Wald χ^2
|EV-class|	11.55	1	<.001
|EV-class| × EV_{back}	3.90	1	.048
|EV-class| × Block	.06	3	.164
|EV-class| × Context	1.45	1	.906
|EV-class| × Switch	1.95	1	.158
|EV-class| × Side	.94	1	.911

Figure S7
Supplementary information for value similarity analysis: related to Fig. 4 and Fig. 5

Fig. S7: Supplementary information for Value similarity analysis: related to Fig. 4 and Fig. 5. .

Focusing on the Value classifier we asked whether EVs affected not only the probability of the corresponding class, but also influenced the full probability distribution predicted by the Value classifier. We reasoned that if the classifier is decoding the neural code of values, then similarity between the values assigned to the classes will yield similarity in probabilities associated to those classes. Specifically, we expected not only that the probability associated with the correct class be highest (e.g. '70'), but also that the probability associated with the closest class (e.g. '50') would be higher than the probability with the least similar class (e.g. '30', panel a, note that this difference also reflects which options where displayed vs not in a given trial). The following analyses model directly the class probabilities estimated by this classifier. Probabilities were modelled with beta regression mixed effects models (Magnusson et al., 2017). For technical

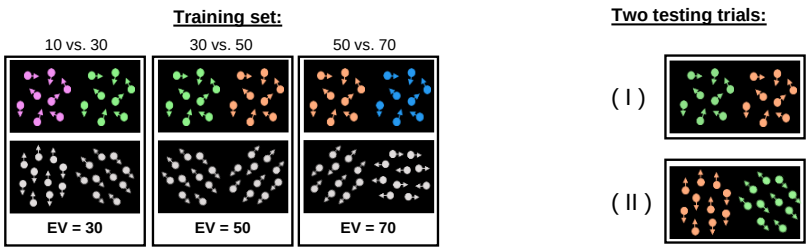
reasons, we averaged across nuisance regressors used in behavioral analyses. An exploratory analysis of raw data including nuisance variables showed that they had no influence and confirmed all model comparison results reported (see Fig. S7. To test our hypothesis, we modelled the probabilities in each trial as a function of the absolute difference between the objective EV of the trial and the class (|EV-class|, i.e. in the above example with a correct class of 70, the probability for the class 50 will be modelled as condition 70-50=20 and the probability of 30 as 70-30=40). This analysis indeed revealed such a value similarity effect ($\chi^2_{(1)} = 12.74$, $p < .001$) also when tested separately on 1D and 2D trials ($\chi^2_{(1)} = 14.22$, $p < .001$, $\chi^2_{(1)} = 9.99$, $p = .002$, respectively, panel d.). Note that the difference between |EV-class| = 20 and |EV-class| = 40 also reflects which options where displayed vs. not in a given trial. Careful analysis of perceptual overlap, however, indicated that this could not explain our results (see below and SI).

Our main hypothesis was that context-irrelevant values might directly influence neural codes of expected value in the vmPFC. The experimentally manipulated background values in our task should therefore interact with the EV probabilities decoded from vmPFC. We thus asked whether the above described value similarity effect was influenced by EV_{back} and/ or Congruency in 2D trials. Analogous to our RT analyses, we used a hierarchical model comparison approach and tested if the interaction of value similarity with these factors improved model fit. We found that EV_{back}, but not Congruency, modulated the value similarity effect ($\chi^2_{(1)} = 6.16$, $p = .013$, $\chi^2_{(1)} = .58$, $p = .446$, respectively, panel d). This effect indicated that the higher the EV_{back} was, the less steep was the value similarity effect. These results also hold when running the models nested within the levels of EV (panels g-i). Additional control analyses included perceptual models that merely encoded the amount of perceptual overlap between each training class and 2D testing as well as the presence of the perceptual feature corresponding to EV_{back} in the training class. These analyses indicated that our classifier was indeed sensitive to values and not only to the perceptual features the values were associated with, see S8 for details.

a. Analyses of all probabilities by the Value classifier revealed gradual value similarities. The y-axis represents the probability assigned to each class, colors indicate the classifier class and the x-axis represents the trial type (the objective EV of the trial). As can be seen, the highest probability was assigned to the class corresponding to the objective EV of the trial (i.e. when the color label matched the X axis label). N=35. **b.** Larger difference between the decoded class and the objective EV of the trial (x axis) was related to a lower probability assigned to that class (y axis) when tested in 1D, 2D or all trials (likelihood ratio test compared to null model: all $p < .002$, N=35, grey shades). Hence, the multivariate classifier reflected gradual value similarities. Note that when |EV - class|=0, P_{class} is the probability assigned to the objective EV of the trial. **c.** EV_{back} modulated the value similarity effect (likelihood-ratio test with added term: $p = .013$, N=35) indicating weaker simialrity between EV representations for higher EV_{back}. **d.** AIC values of competing models of value probabilities classified from vmPFC. Hierarchical model comparison of 2D trials revealed not only the differences between decoded class and objective EV (|EV-class|) improved model fit (likelihood-ratio test: $p < .002$, N=35), but rather that EV_{back} modulated this effect. Crucially, Congruency did not directly modulate the value similarity (likelihood-ratio test: $p = .446$, N=35). Asterisks represent p-value lower than conventional .05 threshold. Light gray bars represent models outside the hierarchical comparison. Including a 3-way interaction (with both EV_{back} and Congruency) did not provide better AIC score (-3902.5,-3901.6, respectively). A perceptual model encoding the feature similarity between each

testing trial and the training classes (irrespective of values) did not provide a better AIC score than the value similarity model (|EV-class|), see Fig S8 for details. **e.** Main value similarity model comparison replicated when fitting the models to unaveraged data. Adding a term for |EV-class| improved model fit (likelihood-ratio test with added term: $\chi^2_{(1)} = 11.56$, $p < .001$). Adding an additional term for |EV-class| $\times$ $\mathrm{EV_{back}}$ further improved the fit (likelihood-ratio test: $\chi^2_{(1)} = 3.86$, $p = .049$, N=35), as in the model reported in panel c). Asterisks represent p-value lower than conventional .05 threshold. **f.** Effect of Nuisance regressors on unaveraged data (t, Side, Switch and Context). Same as Congruency and $\mathrm{EV_{back}}$, all of the nuisance regressors don't discriminate between the classes, but rather assign the same value to all three probabilities from that trial (which sum to 1). We therefore tested if any of them modulated the value similarity effect. As can be seen in the table, none of the nuisance regressors modulated the value similarity effect. **g-i.** Replication of the value similarity model comparison reported in the main text, averaged across nuisance regressors and nested within the levels of EV, i.e. including EV-specific intercepts nested within each within each subject level ($\zeta_{0_{kv}}$, see methods). As in the analysis reported in the Main Text, adding a main effect for |EV-Class| improves model fit (likelihood-ratio test against null model: $\chi^2_{(1)} = 16.15$, $p < .001$, N=35, first row) as well as adding an additional interaction term |EV-class| $\times$ $\mathrm{EV_{back}}$ (likelihood-ratio test with added term: $\chi^2_{(1)} = 6.16$, $p = .013$, N=35). Panel g shows the value simialrity effect across levels of EV, panel h and g show data and fit of the effect of $\mathrm{EV_{back}}$ interaction across levels of EV, respectively. Error bars throughout the figure represent corrected within subject SEMs (Cousineau et al., 2005; Morey et al., 2008).

a.



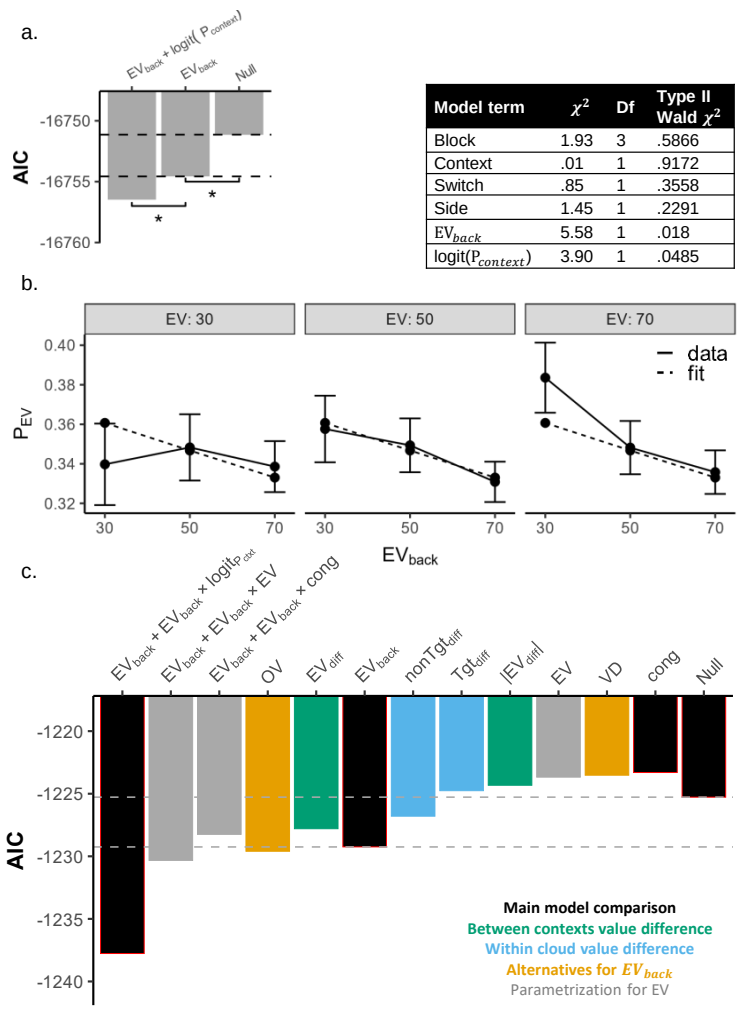
b.

	Test (color)		Test (motion)		EV	EV_{back}	Train Class (color & motion)		|EV-class|	Feature Similarity	$Similarity_{back}$
(I) 1D color	30	50			50	--	10	30	**20**	**1** 1x30	-
							30	50	**0**	**2** 1x30, 1x50	-
							50	70	**20**	**1** 1x50	-
(II) 2D color	30	50	10	30	50 color	30 motion	10	30	**20**	**3** 2x30, 1x10	**1**
							30	50	**0**	**3** 2x30, 1x50	**1**
							50	70	**20**	**1** 1x50	**0**

Figure S8
Supplementary information for perceptual similarity analysis: related to Fig. 4 and Fig. 5

Fig. S8: Supplementary information for perceptual similarity analysis: related to Fig. 4. To control that our EV classifier was indeed sensitive to values and not only to the perceptual features the values were associated with, we compared this value similarity model to a perceptual models that merely encodes the amount of perceptual overlap between each training class and 2D testing (irrespective of their corresponding values) and found that our model explained the data best (see panel d). Replacing the EV_{back} with a parameter that encodes the presence of the perceptual feature corresponding to EV_{back} in the training class ($Similarity_{back}$: 1 if the feature was preset, 0 otherwise) did not provide a better AIC score (-3897.1) than including the value of EV_{back} (-3902.5). **a.** Left: training set consisting of 1D trials provided for the classifier for each class (in the experiment the sides were pseudorandomised). Note that each class had the same amount of color and motion 1D trials and that the value difference between the values was always 20. Right: two examples of 2D trials that constituted the classifier test set. **b.** The table illustrates the calculation of feature similarity between classifier test and training in two example trials in one 1D and one 2D trial. Specifically, shown are the corresponding values and features for each trial with the predicted values at each class for the parameters value similarity (|EV-class|), feature similarity and $similarity_{back}$. Feature similarity encodes the perceptual overlap between the shown test example and the training examples underlying with each value class. The first row shows a case in which the classifier was tested on a 1D green vs. orange color trial (30 vs 50, EV = 50). Considering in this case for instance the predicted probability that EV=30, the table illustrates the training example underlying the EV = 30 cases (10 vs 30, dark gray shading), the |EV-class| (here: 20, because 50-30), and the feature similarity i.e. how many features from the training class appeared in the test example (here: 1). The second row shows a 2D color trial, reflecting the same value based choice between 30 and 50. The value similarity between training and test stays the same as for the 1D trial shown above. However, the feature similarity between test and training changes because of the motion features. If we take class 30 for example (which is 10 vs 30, dark gray shading), the feature 30 appeared twice (color and motion) and the feature 10 appeared once

(motion), i.e. feature similarity now takes on the value 3. $\text{Similarity}_{\text{back}}$ was used to test a perceptual-based alternative to the EV_{back} parameter. $\text{Similarity}_{\text{back}}$ takes on 1 if the perceptual feature corresponding to the EV_{back} appeared in the training class and 0 otherwise (red text in table). As described in the main text, none of the perceptual-similarity encoding alternatives provided a better fit than the reported models that focused on the values the features represent.



Model term	χ^2	Df	Type II Wald χ^2
Block	1.93	3	.5866
Context	.01	1	.9172
Switch	.85	1	.3558
Side	1.45	1	.2291
EV_{back}	5.58	1	.018
logit($P_{context}$)	3.90	1	.0485

Figure S9
Modelling probability assigned to the EV class.

Fig. S9: Modelling probability assigned to the EV class: related to Fig. 5.
a. We replicated the main results using the unaveraged data. The Null model was: $P^k_{t,EV} = \beta_0 + \gamma_{0k} + \nu_1 side(t) + \nu_2 switch(t) + \nu_3 context(t)$, where $P^k_{t,EV}$ is the probability assigned to the class corresponding to the EV of trial t for subject k, β_0 and γ_{0k} represent global and subject-specific intercepts. Side, Switch and Context are the same as in the RT model (Eq. 2); None of these variables had a main effect, $p > 0.4$ (Type II Wald χ^2 tests, N=35, see table, right), N=35. The factor $trial$ could not be included due to model convergence issues. Adding a term representing EV_{back} improved model fit (likelihood-ratio test including term: $\chi^2_{(1)} = 5.42$, $p = .019$). Adding an additional term for context decodability further improved the fit (likelihood-ratio test with added term: $\chi^2_{(1)} = 3.9$, $p = .048$). The table (right) displays the Type 2 Wald χ^2 test for all main effects from the model. **b.** Depicted is the effect of EV_{back} (x-axis) on the probability assignd to the EV class (P_{EV}, y axis). Solid lines represent the data and dashed lines the model fit of a model that included random effects of subject and EV nested within subject (data averaged across nuisance regressors, adding a main effect for EV_{back} improved model fit (likelihood-ratio test with added term: $\chi^2_{(1)} = 5.99$, $p = .014$, N=35). Error bars represent

corrected within subject SEMs Cousineau et al., 2005; Morey et al., 2008. **c.** Similar to our analysis of alternative models of RT, we clustered models reflecting alternative explanations into three conceptual groups (see color legend; cf. Fig. S3a). All models were fitted to the probability assigned to the objective EV in accurate 2D trials, similar to Eq. 7. Each column represents the AIC (y-axis) of a different model (x-axis) where the labels on the x-axis depict all the main effects included in that specific model (i.e. added to the Null, i.e. Eq. 7 without any main effects). We found no evidence that any other parameters explain the data better than the ones we used in the main text. Specifically, only including main effect of $\mathrm{EV}_{\mathrm{back}}$, Overall Value of the irrelevant values (OV) and the difference of both EVs (EV_{diff}) provided a better AIC score than the Null model. Note that adding OV (-1229.6) only slightly surpassed $\mathrm{EV}_{\mathrm{back}}$ (-1229.26). Crucially, the correlation of $\mathrm{EV}_{\mathrm{back}}$ and OV is very high (Pearson correlation: $\rho = .87$, see main text). We then looked at possible interactions with the $\mathrm{EV}_{\mathrm{back}}$ effect. Congruency did not seem to modulate the main effect of $\mathrm{EV}_{\mathrm{back}}$ and adding an interaction term EV $\times$ $\mathrm{EV}_{\mathrm{back}}$ provided a slightly better AIC (-1230.33), yet this effect was not significant (likelihood-ratio test: $\chi^2_{(1)} = 3.08$, $p = .079$). Section (b) also visualizes this effect. Lastly, adding a term for the Context decodability provided the lowest (i.e. best) AIC score. This exploratory analysis revealed that our model provides the best fit for P_{EV} in all cases except when $\mathrm{EV}_{\mathrm{back}}$ was replaced with the sum of irrelevant values (-1229.6, -1229.2, respectively, Fig. S9). In contrast, AIC scores of behavioral models' favored $\mathrm{EV}_{\mathrm{back}}$ as modulator of Congruency, over the sum of irrelevant values (-6626.6, -6619.9, respectively, Fig.S3). However, both parameters were strongly correlated ($\rho = .87$, $\sigma = .004$) and therefore our task was not designed to distinguish between these two alternatives.

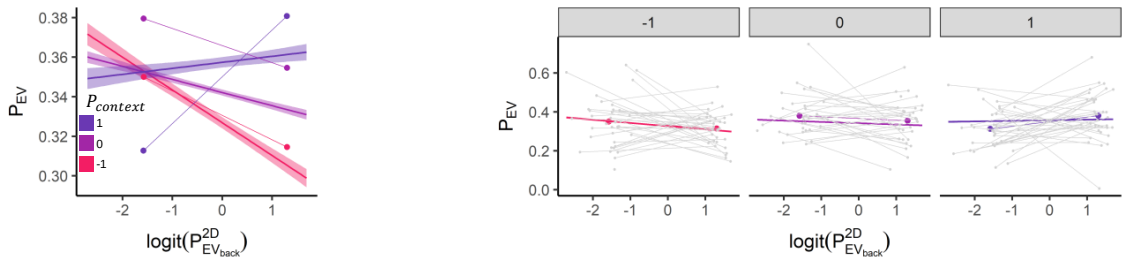


Figure S10
Main effects and corresponding data, fMRI effects, related to Fig. 5

Fig. S10: Main effects and corresponding data, fMRI effects, related to Fig. 5. Since the effects describe data and predictors that are beta-distributed, visualization of simply imposing the true data over the predictions is not very informative. To solve this, and only for visualization purposes here and in the main paper, we took for each effect the mean of top and bottom 20% of the true probabilities from the classifiers (not transformed) for each participant. Context signal ($P_{context}$) moderated the negative effect of EV_{back} decodability ($P^{2D}_{EV_{back}}$) on EV decodability (P_{EV}). Model prediction of multilogit($P^{2D}_{EV_{back}}$) x $P_{context}$ (left, taken from Fig. 5h.) and top and bottom 20% for each subject for three levels of $P_{context}$ (right, the split to three levels is for visualization whereas in the model the predictor was continuous). In all panels error bands represent the 89% confidence interval.

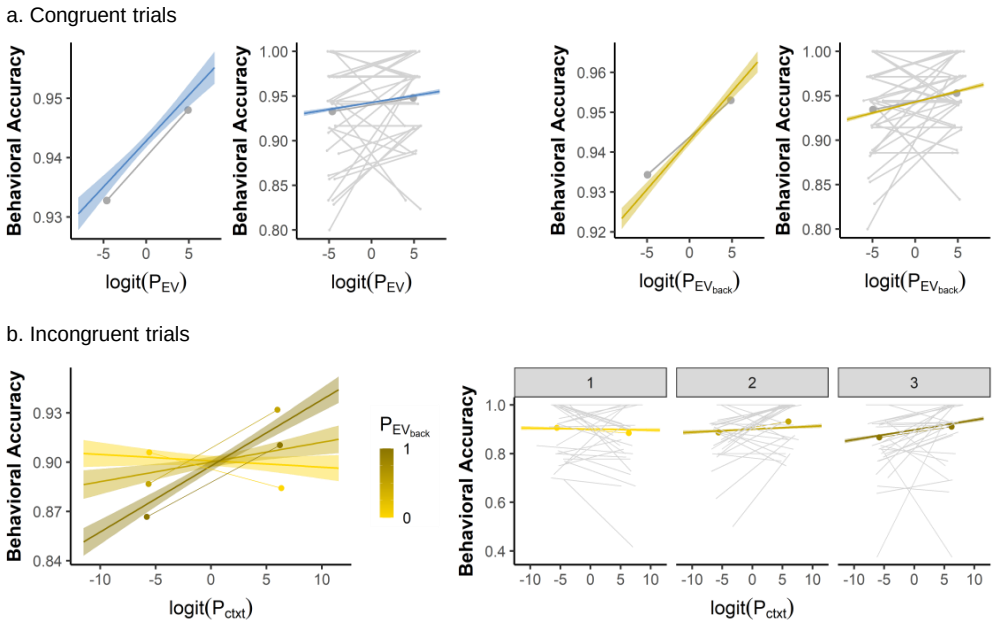
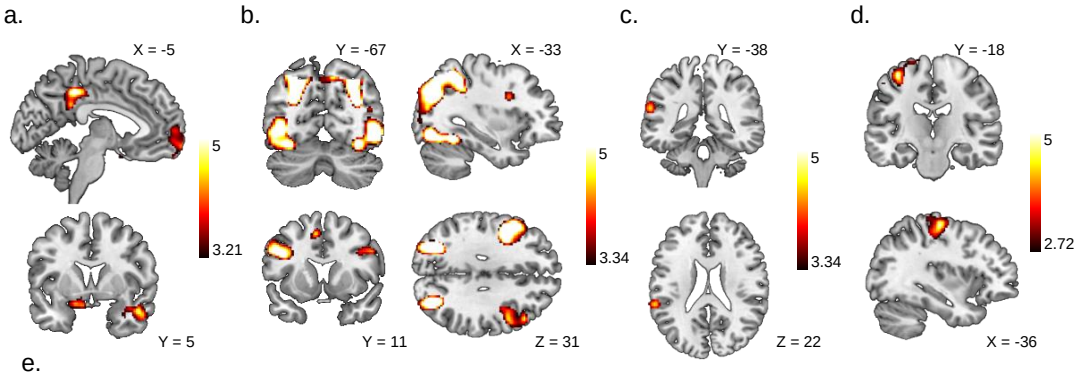


Figure S11
Main effects and corresponding data, link of fMRI to behavioral accuracy, related to Fig. 6

Fig. S11: Main effects and corresponding data, link of fMRI to behavioral accuracy, related to Fig. 6. Since the effects describe data and predictors that are beta-distributed, visualization of simply imposing the true data over the predictions is not very informative. To solve this, and only for visualization purposes here and in the main paper, we took for each effect the mean of top and bottom 20% of the true probabilities from the classifiers (not transformed) for each participant. **a** Congruent trials. Stronger EV decodability (left) and stronger $\mathrm{EV_{back}}$ decodability (right) increases behavioral accuracy. The left side of each panel is taken from Fig. 6. The right side depicts the same plot with additional individual subject-specfic lines that represent the top and bottom 20% of the data for each subject (meaning that the gray line on the left side is the mean of the individual lines on the right). **b.** Incongruent trials. Stronger Context decodability ($\mathrm{P}_{context}$) increases behavioral accuracy, modulated by $\mathrm{EV_{back}}$ decodability ($\mathrm{P}_{EV_{back}}$) such that when $\mathrm{P}_{EV_{back}}$ was low, the effect of $\mathrm{P}_{context}$ diminished. For visualization purpose, Right panel is split by 3 equal sized bins of $\mathrm{P}_{EV_{back}}$ (left is the lowest bin, increasing to the right, the split to three levels is for visualization whereas in the model the predictor was continuous). In all panels error bands represent the 89% confidence interval.



e.

Cue (split)	Accurate Stimuli		Non-accurate stimuli	Outcome	+ 13 fmriprep	+ 18 Physiological
Color + Motion	+ 1D	+ 2D	+ Wrong + no-answer	+ Correct + wrong + no-answer	See online methods	

GLM	Parametric modulators 1D	2D
GLM1	EV	EV
GLM2	EV	EV $+EV_{back}$ +Congruency
GLM3	EV	EV $+(EV_{back}$ x Congruency)
GLM4	EV	(EV x Congruency)

Parametric modulators (demeaned):	
EV	{30, 50, 70}
Congruency	{+1, -1}
EV_{back}	{30, 50, 70}
EV_{back} x Congruency	{-70,-50,-30,30, 50, 70}
EV x Congruency	{-70,-50,-30,30, 50, 70}

GLM	Contrasts: below threshold, *p<0.005, **p<0.001
GLM1	2D > 1D** , 1D > 2D*
GLM2	Congruency > 0, Congruency < 0, EV_{back} > 0, EV_{back} < 0**
GLM3	(EV_{back} x Congruency) > 0, (EV_{back} x Congruency) < 0 *
GLM4	(EV x Congruency) > 0, (EV x Congruency) < 0

Figure S12
Main univariate results

Fig. S12: Main univariate results.
The main analyses indicated that multiple value expectations are represented in parallel within vmPFC. Here, we asked whether whole-brain univariate analyses could also uncover evidence for processing of multiple value representations. In particular, we asked whether we could find evidence for a single representation that integrates the multiple value expectations into one signal. To this end, we first analyzed the fMRI data using GLMs with separate onsets and EV parametric modulators for 1D and 2D trials (see below for detailed description).
a. The intersection of the EV parametric modulators of 1D and 2D trials ($EV_{1D} > 0 \cap EV_{2D} > 0$) revealed several regions including right Amygdala, bilateral Hippocampus and Angular Gyrus, the lateral and medial OFC and overlapping vmPFC. Hence, the vmPFC signaled the expected value of the current context in both trial types as expected – even though 2D trials likely required higher attentional demands (see panel b). Voxelwise threshold $p < .001$, FDR cluster-corrected.
b 2D trials were characterized by increased activation in an attentional network involving occipital, parietal and frontal clusters (2D > 1D, $p < .001$ FDR cluster corrected).
Next, we searched for univariate evidence of processing irrelevant values by modifying the

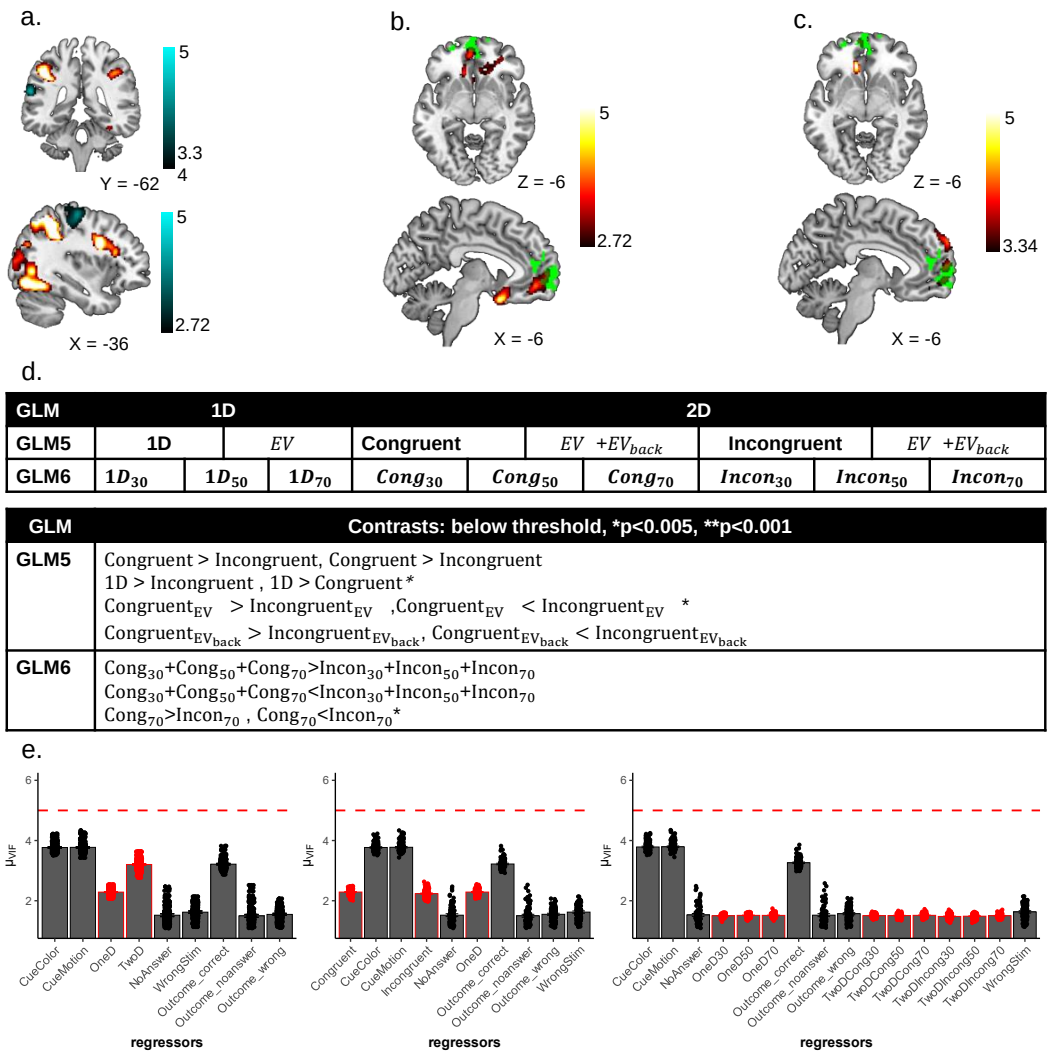
parametric modulators assigned to 2D trials in the above-mentioned GLM. Specifically, in addition to EV_{2D}, we added Congruency (+1 for congruent and -1 for incongruent) and EV_{back} as additional modulators of the activity in 2D trials. This GLM revealed no evidence for a Congruency contrast anywhere in the brain (even at a liberal voxel-wise threshold of $p < .005$). **c.** An unexpected negative effect of EV_{back} was found in the Superior Temporal Gyrus ($p < .001$), i.e. the higher the EV_{back}, the lower the signal in this region. $p < .001$, FDR cluster-corrected. No overlap with (b), see S13. We note that this is similar to previous reports implicating this region in modelling choices of others (Nicolle et al., 2012)). Notably, unlike the multivariate analysis, no effect in any frontal region was observed.

Motivated by our behavioral analysis, we then turned to look for the interaction of each relevant or irrelevant value with Congruency. An analysis including only a Congruency $\times$ EV_{2D} parametric modulator revealed no cluster (even at $p < .005$).

d. A cluster in the primary motor cortex was negatively modulated by Congruency $\times$ EV_{back}, i.e. the difference between Incongruent and Congruent trials increased with higher EV_{back}, similar to the RT effect and akin to a response conflict, $p < .005$, FDR cluster-corrected. No overlap with (b), see S13

Lastly, we re-ran all above analyses concerning Congruency and EV_{back} only inside the identified vmPFC ROI. No voxel survived for Congruency, EV_{back} nor the interactions, even at threshold of $p < .005$.

e. Visualization of GLMs. The tables depict the structure of GLMs1-4 which were mainly motivated by the behavioral analysis; onset regressors are shown in the top table, parametric modulators assigned to 1D and 2D onsets (middle-left), the values they were modeled with (demeaned, middle-right) are shown below. The contrasts of interest are shown in the bottom table. The GLMs differed only in their modulations of the 2D trials: GLM1 included only modulators of the objective outcome, GLM2 included one modulator for Congruency and one for EV_{back}, GLM3 included a modulator for the Congruency $\times$ EV_{back} interaction and GLM4 included instead of the EV modulator a modulator of the EV $\times$ Congruency interaction. In the contrast table (bottom) contrasts that only revealed effects at a liberal threshold of $p < .005$ are marked with one star, and contrasts significant at $p < .001$ are marked with two stars. All statistical tests represent one-sided t-test either larger or smaller than 0, see lower table in panel e for details of each contrast.



GLM	1D			2D					
GLM5	1D		EV	Congruent		EV $+EV_{back}$	Incongruent		EV $+EV_{back}$
GLM6	$1D_{30}$	$1D_{50}$	$1D_{70}$	$Cong_{30}$	$Cong_{50}$	$Cong_{70}$	$Incon_{30}$	$Incon_{50}$	$Incon_{70}$

GLM	Contrasts: below threshold, *p<0.005, **p<0.001
GLM5	Congruent > Incongruent, Congruent > Incongruent 1D > Incongruent , 1D > Congruent * Congruent_{EV} > Incongruent_{EV} ,Congruent_{EV} < Incongruent_{EV} * $\text{Congruent}_{EV_{back}}$ > $\text{Incongruent}_{EV_{back}}$, $\text{Congruent}_{EV_{back}}$ < $\text{Incongruent}_{EV_{back}}$
GLM6	$\text{Cong}_{30}+\text{Cong}_{50}+\text{Cong}_{70}>\text{Incon}_{30}+\text{Incon}_{50}+\text{Incon}_{70}$ $\text{Cong}_{30}+\text{Cong}_{50}+\text{Cong}_{70}<\text{Incon}_{30}+\text{Incon}_{50}+\text{Incon}_{70}$ $\text{Cong}_{70}>\text{Incon}_{70}$, $\text{Cong}_{70}<\text{Incon}_{70}$*

Figure S13
Additional univariate results

Fig. S13: Additional univariate results.
a. Overlap of effects of EV_{back} and trial type (2D > 1D). Main effects of EV_{back}<0 (GLM2, $p < 0.001$ FDR cluster corrected, top, blue shades) and EV_{back} X Congruency < 0 (GLM3, $p < 0.005$, FDR cluster corrected, bottom, blue shades, t values) did not overlap with the 2D network (red shades in both panels, t values). **b.** Main effect of 1D > 2D. A stronger signal in vmPFC for 1D over 2D trials revealed weak activation in a PFC network ($p < .005$, red shades,t values). This included the vmPFC (our functional ROI is depicted in green). Interestingly, at a liberal threshold of $p < .005$ we found stronger activity for 1D over 2D trials in a cluster overlapping with vmPFC (1D > 2D, $p < .005$). Although this could be interpreted as a general preference for 1D trials, splitting the 2D onsets by Congruency revealed no cluster for 1D > Incongruent (also at $p < .005$) but a stronger cluster for 1D > Congruent ($p < .001$,Fig. S13). In other words, the signal in the vmPFC was *weaker* when both contexts indicate the same action, compared to when only one context is present. **c.** Stronger signal in vmPFC for 1D over congruent but not incongruent trials. When we split the onset of the 2D into Congruent and

Incongruent trials (GLM5), we found no significant cluster for the 1D > Incongruent contrast, but an overlapping and stronger cluster for the 1D > Congruent contrast ($p < .001$, FDR cluster corrected, red shades, t values). We found very similar results when contrasting the onsets of 1D and Congruent in GLM6 (not presented), confirming the same results also when controlling for the number of trials for each level of EV (i.e. $1D_{30}$+$1D_{50}$+$1D_{70}$> $Congruent_{30}$+$Congruent_{50}$+$Congruent_{70}$). Our functional ROI is depicted in green. **d.** Additional exploratory analyses such as contrasting the onsets of congruent and incongruent trials, confirmed the lack of Congruency modulation in any frontal region. Specifically, We constructed additional GLMs to verify the results of GLMs 1-4. In GLM5 we split the onset of 2D trials into congruent and incongruent trials and assigned a parametric modulator of EV and EV_{back} to each. As in GLM2, we found no effect of congruency; no voxel survived when contrasting the congruency onsets nor their EV_{back} modulators. Only the contrast $Congruent_{EV}$<$Incongruent_{EV}$ revealed a weak cluster in the right visual cortex (peak 38,-80,16, p<0.005 not presented). In GLM6 we split the onsets of the 1D and 2D trials by levels of EV and the 2D trials further by Congruency. No Congruency main effect survived correction. Only when the onsets of Congruent and Incongruent 2D trials with EV=70 were contrasted, a cluster in the primary motor cortex was found (also at $p < .005$). Unsurprisingly, this cluster largely overlapped with the Congruency $\times$ EV_{back} effect reported in the Main Text. Except the contrast of 1D > Congruent (see Main Text) none of the other contrasts shown in the table revealed any cluster, even at $p < .005$. All statistical tests in panels a-d represent one-sided t-test either larger or smaller than 0, see lower table in panel d for details of each contrast. **e.**Variance Inflation Factor (VIF) of the different regressors in all GLMs. None of the regressors (x axis) had a mean VIF value (y axis) across blocks and participants above the threshold of 4. Regressors involved in GLMs 1-4 shown on the left (Fig. S12); GLM5 and GLM6 are shown in the middle and on the right, respectively. See Methods for details. N=35. Error bars represent corrected within subject SEMs (Cousineau et al., 2005; Morey et al., 2008)

Anatomical region		Peak (MNI)					peak	
Label		Distance	X	Y	Z	Cluster size	t$_34$	p$_unc$
EV_{1D} > 0 ∩ EV_{2D} >0 , p<001, k = 280								
R	Inferior Temporal Gyrus	4.90	60	-18	-14	1770	6.53	< .0001
R	Middle Temporal Gyrus	0	50	-6	-20		5.49	< .0001
R	Middle Temporal Gyrus	0	56	-30	-8		5.27	< .0001
R	Superior Frontal Gyrus, medial Orbital	0	8	68	-12	1045	6.09	< .0001
L	Inferior Frontal Gyrus pars orbitalis	0	-50	30	-10		4.67	< .0001
L	Superior Frontal Gyrus	0	-24	58	-6		4.35	< .0001
L	Middle Temporal Gyrus	0	-60	-30	-6	1318	5.85	< .0001
L	Middle Temporal Gyrus	0	-66	-24	-8		5.78	< .0001
L	Hippocampus	2	-40	-26	-12		4.96	< .0001
L	Angular Gyrus	0	-50	-60	38	875	5.58	< .0001
L	Angular Gyrus	0	-46	-52	30		4.86	< .0001
L	Angular Gyrus	0	-46	-70	34		3.66	.0002
L	Middle Cingulate & Paracingulate Gyri	0	-4	-40	44	1065	5.51	< .0001
L	Posterior Cingulate Gyrus	0	0	-44	32		4.52	< .0001
R	Middle Cingulate & Paracingulate Gyri	0	12	-48	32		4.52	< .0001
L	Hippocampus	0	-18	-6	-20	280	4.59	< .0001
L	Olfactory Cortex	2	-10	6	-18		4.34	< .0001
R	Angular Gyrus	0	50	-56	30	474	4.27	< .0001
R	Superior Temporal Gyrus	0	62	-54	22		4.26	< .0001
2D > 1D, p<.001, k=158								
L	Superior Occipital Gyrus	2.83	-28	-76	38	5367	8.71	< .0001
L	Inferior Occipital Gyrus	0	-48	-76	-4		7.69	< .0001
L	Superior Parietal Gyrus	0	-28	-66	52		7.62	< .0001
L	Precentral Gyrus	0	-46	4	30	1766	7.69	< .0001
L	Inferior Frontal Gyrus, triangular part	0	-44	34	22		5.88	< .0001
L	Inferior Frontal Gyrus, triangular part	0	-40	26	22		5.59	< .0001
R	Inferior Parietal Gyrus	0	32	-56	54	3876	7.23	< .0001
R	Fusiform Gyrus	0	30	-76	-10		7.16	< .0001
R	Inferior Temporal Gyrus	0	48	-70	-8		7.13	< .0001
R	Inferior Frontal Gyrus, triangular part	0	48	26	26	616	5.17	< .0001
R	Precentral Gyrus	0	48	8	32		4.50	< .0001
R	Precentral Gyrus	0	38	2	30		4.23	.0001
L	Supplementary Motor Area	0	-8	14	50	159	4.69	< .0001
EV_{back}<0, p<.001, k = 240								
L	SupraMarginal Gyrus	2	-62	-38	22	240	4.50	< .0001
L	Superior Temporal Gyrus	0	-60	-32	10		4.26	.0001
L	Superior Temporal Gyrus	0	-60	-22	8		3.71	.0004
Congruency × EV_{back}<0, p<.005, k=632								
L	Postcentral Gyrus	6.93	-36	-18	60	632	4.03	.0002
L	Postcentral Gyrus	0	-48	-22	52		3.11	.0019
L	Postcentral Gyrus	0	-24	-20	74		3.08	.0020
EV_{1D} + EV_{2D} >0, within functional ROI, p<.001, k=979								
R	Anterior Orbital Gyrus	4.47	8	68	-12	979	7.89	< .0001
L	Superior Frontal Gyrus, Medial Orbital	2	-6	68	-12		6.86	< .0001
L	Superior Frontal Gyrus, Medial	0	-10	64	2		5.86	< .0001

Table 3

Detailed univariate results: *Clusters for whole brain univariate analysis, related to Fig. S12. Presented are the closest labels to the local maxima of each cluster and each contrast using AAL3v1 (Rolls et al., 2015, 2020; Tzourio-Mazoyer et al., 2002). All contrasts are FDR cluster corrected. p and k values presented for each cluster. p values represent one sided t-test.*

Effect sizes and confidence intervals for best explaining models. In all the following tables, CI marks Confidence Interval, CIl and CIh the low and high ends of the confidence interval respectively. P-values in all tables showing best explaining models represent Type II Wald χ^2 tests.

	Parameter	Coef	CI	CIl	CIh	t	df_er	p	Effects	Std_{Coef}	Std^{CIl}_{Coef}	Std^{CIh}_{Coef}
1	(Intercept)	0.08	0.95	0.06	0.11	6.48	9035	0.00	fixed	0.17	0.04	0.29
2	Switch	-0.01	0.95	-0.01	-0.01	-6.57	9035	0.00	fixed	-0.06	-0.08	-0.04
3	Trial	-0.01	0.95	-0.02	-0.01	-6.92	9035	0.00	fixed	-0.06	-0.08	-0.04
4	side [R]	-0.01	0.95	-0.01	-0.00	-2.29	9035	0.02	fixed	-0.04	-0.08	-0.01
5	Context [M]	-0.07	0.95	-0.08	-0.07	-20.73	9035	0.00	fixed	-0.37	-0.40	-0.33
6	EV	-0.07	0.95	-0.08	-0.07	-29.14	9035	0.00	fixed	-0.36	-0.39	-0.34
7	Congruency [1]	0.02	0.95	0.01	0.03	5.40	9035	0.00	fixed	0.10	0.06	0.13
8	Congruency [-1] * EV_{back}	-0.01	0.95	-0.01	-0.00	-2.07	9035	0.04	fixed	-0.03	-0.05	-0.00
9	Congruency [1] * EV_{back}	0.01	0.95	0.00	0.01	3.73	9035	0.00	fixed	0.05	0.02	0.07
10	EV * Congruency [1]	0.01	0.95	0.00	0.01	2.08	9035	0.04	fixed	0.04	0.00	0.07
11		0.07	0.95						random-sub			
12		0.17	0.95						random-Residual			

Table 4
Effect sizes and confidence intervals for best explaining RT model

	Parameter	Coef	CI	CIl	CIh	z	p	Effects	Std_{Coef}	Std_{Coef}^{CIl}	Std_{Coef}^{CIh}
1	(Intercept)	-0.65	0.95	-0.69	-0.61	-30.83	0.00	fixed	-0.65	-0.69	-0.61
2	EV_{back}	-0.05	0.95	-0.08	-0.01	-2.58	0.01	fixed	-0.05	-0.08	-0.01
3	logit($P_{context}$)	0.06	0.95	0.02	0.10	3.25	0.00	fixed	0.06	0.02	0.10
4		0.06	0.95	0.02	0.18			random-sub			
5		8.79	0.95					random-Residual			

Table 5
Effect sizes and confidence intervals for best explaining fMRI model (main model)l

	Parameter	Coef	CI	CIl	CIh	z	p	Effects	Std_{Coef}	Std_{Coef}^{CIl}	Std_{Coef}^{CIh}
1	(Intercept)	-0.65	0.95	-0.69	-0.61	-31.36	0.00	fixed	-0.65	-0.69	-0.61
2	logit($P_{context}$)	0.06	0.95	0.02	0.09	3.02	0.00	fixed	0.06	0.02	0.09
3	mlogit($P_{EV_{back}}$	-0.03	0.95	-0.07	0.01	-1.63	0.10	fixed	-0.03	-0.07	0.01
4	logit($P_{context}$) * mlogit($P_{EV_{back}}$)	0.04	0.95	0.00	0.08	2.17	0.03	fixed	0.04	0.00	0.08
5		0.05	0.95	0.00	0.62			random-EV_{back}:sub			
6		0.04	0.95	0.00	0.70			random-sub			
7		8.82	0.95	8.16	9.54			random-Residual			

Table 6
Effect sizes and confidence intervals for best explaining fMRI model (model nested in EV_{back}

	Parameter	Coef	CI	CIl	CIh	z	df_error	p	Effects	Std_{Coef}	Std_{Coef}^{CIl}	Std_{Coef}^{CIh}
1	(Intercept)	0.01	0.95	0.01	0.02	30.66	Inf	0.00	fixed	0.01	0.01	0.02
2	$EV^{diagonal}$	-0.00	0.95	-0.00	0.00	-1.77	Inf	0.08	fixed	-0.00	-0.00	0.00
3	$EV_{back}^{diagonal}$	-0.00	0.95	-0.00	-0.00	-15.61	Inf	0.00	fixed	-0.00	-0.00	-0.00
4		0.00	0.95	0.00	0.00				random-freq:sub			
5		0.00	0.95	0.00	0.00				random-sub			
6		0.27	0.95	0.27	0.27				random-freq			

Table 7
Effect sizes and confidence intervals for best explaining RSA model - diagonal models

	Parameter	Coef	CI	CIl	CIh	z	df_error	p	Effects	Std_{Coef}	Std_{Coef}^{CIl}	Std_{Coef}^{CIh}
1	(Intercept)	0.01	0.95	0.01	0.02	30.66	Inf	0.00	fixed	0.01	0.01	0.02
2	VDev	-0.00	0.95	-0.00	-0.00	-2.17	Inf	0.03	fixed	-0.00	-0.00	-0.00
3	VDevback	-0.00	0.95	-0.00	-0.00	-15.93	Inf	0.00	fixed	-0.00	-0.00	-0.00
4		0.00	0.95	0.00	0.00				random-freq:sub			
5		0.00	0.95	0.00	0.00				random-sub			
6		0.27	0.95	0.27	0.27				random-freq			

Table 8

Effect sizes and confidence intervals for best explaining RSA model - value difference models

Supplementary information for 2nd project

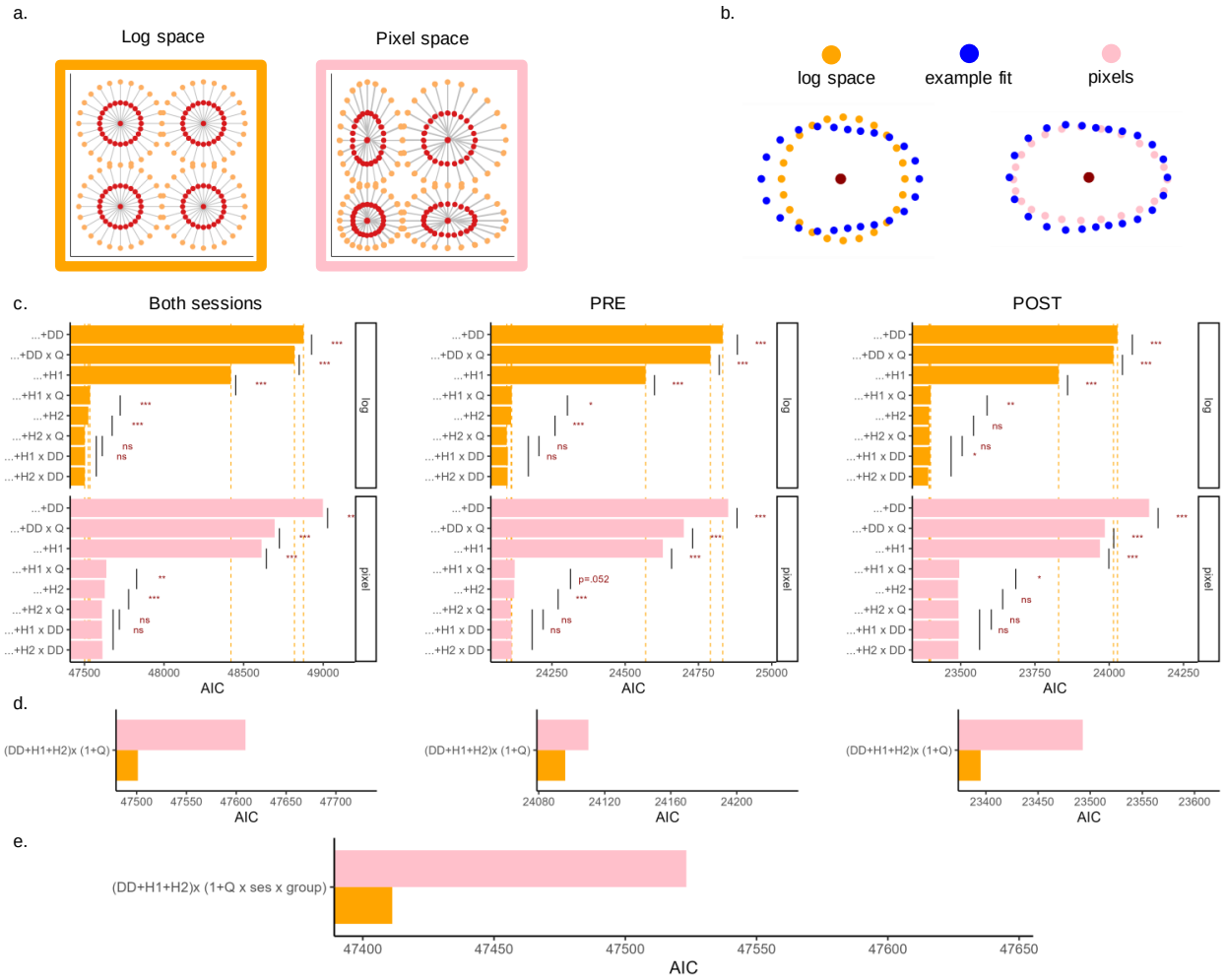


Figure S14

Fit on PRE and POST separately and logging of space a We logged the space of leaves and fruits (left). On the right is the true number of leaves and fruits ('pixel' space). **b** By adding H_1 and H_2 to coordinates in logged space, the model can modify the space such that it would capture even if participants perceive the distances exactly as the un-logged space. Orange dots show one toy quadrant in logged space and pink in pixel. There exists a set of coefficients that would, in practice, transform the coordinates 'back' to the pixel space. Blue points show the model prediction of such a hypothetical set of parameters, after moving each point according to the model prediction, similar to Fig.8. **c** General effect of DD, H_1 and $H2$ and their interaction with Quadrant on both sessions, only PRE and only POST (columns) for pixel of logged coordinates (rows), without any interaction with group or session (equivalent to Fig9a). For each space, we re-calculated DD and the angles generating $H1$ and $H2$. Most of the comparisons hold across all combinations of sessions, showing each component had a main effect in each session and almost always an interaction with the quadrant. Vertical lines show the best-explaining model in log space. There are some minor discrepancies with respect to H_2 and its interaction with Quadrant or DD which we plan to investigate in the future. The main effects related to reward,DD and H_1, hold across all combinations of sessions and spaces. **d** Our best-explaining model (without session and group interaction) provided a better fit (lower AIC) in log space compared to pixel space, for both session (left), PRE (middle), or POST (right). **e** Our best explaining model, including session and group interaction, also provided a better fit (lower AIC) in log space compared to pixel space.

References

Abbott, J. T., Griffiths, T. L., & Regier, T. (2016). Focal colors across languages are representative members of color categories. *Proceedings of the National Academy of Sciences*, *113*(40), 11178–11183.

Abitbol, R., Lebreton, M., Hollard, G., Richmond, B. J., Bouret, S., & Pessiglione, M. (2015). Neural mechanisms underlying contextual dependency of subjective values: Converging evidence from monkeys and humans. *Journal of Neuroscience*, *35*(5), 2308–2320.

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*. https://doi.org/10.3389/fninf.2014.00014

Ambrogioni, L., & Ólafsdóttir, H. F. (2023). Rethinking the hippocampal cognitive map as a meta-learning computational module. *Trends in Cognitive Sciences*, *27*(8), 702–712.

Anand, A., Racah, E., Ozair, S., Bengio, Y., Côté, M.-A., & Hjelm, R. D. (2019). Unsupervised state representation learning in atari. *Advances in neural information processing systems*, *32*.

Anderson, B. A. (2013). A value-driven mechanism of attentional selection. *Journal of vision*, *13*(3), 7–7.

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

Asaad, W. F., Lauro, P. M., Perge, J. A., & Eskandar, E. N. (2017). Prefrontal Neurons Encode a Solution to the Credit-Assignment Problem. *Journal of Neuroscience*, *37*(29), 6995–7007. https://doi.org/10.1523/JNEUROSCI.3311-16.2017

Aurelio, C., Asuka, Y., Maryam, H., Pradyumna, S., Mitsuo, K., et al. (2021). Value signals guide abstraction during learning. *eLife*, *10*.

Avants, B., Epstein, C., Grossman, M., & Gee, J. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41. https://doi.org/10.1016/j.media.2007.06.004

Averbeck, B., & O'Doherty, J. P. (2022). Reinforcement-learning in fronto-striatal circuits [Publisher: Nature Publishing Group]. *Neuropsychopharmacology*, *47*(1), 147–162. https://doi.org/10.1038/s41386-021-01108-0

Ballesta, S., Shi, W., Conen, K. E., & Padoa-Schioppa, C. (2020). Values encoded in orbitofrontal cortex are causally related to economic choices [Number: 7838 Publisher: Nature Publishing Group]. *Nature*, *588*(7838), 450–453. https://doi.org/10.1038/s41586-020-2880-x

Banino, A., Barry, C., Uria, B., Blundell, C., Lillicrap, T., Mirowski, P., Pritzel, A., Chadwick, M. J., Degris, T., Modayil, J., Wayne, G., Soyer, H., Viola, F., Zhang, B., Goroshin, R., Rabinowitz, N., Pascanu, R., Beattie, C., Petersen, S., … Kumaran, D. (2018). Vector-based navigation using grid-like representations in artificial agents [Publisher: Nature Publishing Group]. *Nature*, *557*(7705), 429–433. https://doi.org/10.1038/s41586-018-0102-6

Bao, X., Gjorgieva, E., Shanahan, L. K., Howard, J. D., Kahnt, T., & Gottfried, J. A. (2019). Grid-like neural representations support olfactory navigation of a two-dimensional odor space. *Neuron*, *102*(5), 1066–1075.

Baram, A. B., Muller, T. H., Nili, H., Garvert, M. M., & Behrens, T. E. J. (2021). Entorhinal and ventromedial prefrontal cortices abstract and generalize the structure of reinforcement learning problems. *Neuron*, *109*(4), 713–723.e7. https://doi.org/10.1016/j.neuron.2020.11.024

Barron, H. C., Dolan, R. J., & Behrens, T. E. (2013). Online evaluation of novel choices by simultaneous representation of multiple memories. *Nature neuroscience*, *16*(10), 1492–1498.

Barron, H. C., Garvert, M. M., & Behrens, T. E. (2015). Reassessing vmpfc: Full of confidence? *Nature neuroscience*, *18*(8), 1064–1066.

Barron, H. C., Garvert, M. M., & Behrens, T. E. (2016). Repetition suppression: A means to index neural representations using bold? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1705), 20150355.

Barron, H. C., Reeve, H. M., Koolschijn, R. S., Perestenko, P. V., Shpektor, A., Nili, H., Rothaermel, R., Campo-Urriza, N., O'Reilly, J. X., Bannerman, D. M., et al. (2020).

Neuronal computation underlying inferential reasoning in humans and mice. *Cell*, *183*(1), 228–243.

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, *76*, 412–427. https://doi.org/10.1016/j.neuroimage.2013.02.063

Basten, U., Biele, G., Heekeren, H. R., & Fiebach, C. J. (2010). How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Sciences*, *107*(50), 21767–21772. https://doi.org/10.1073/pnas.0908104107

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bavard, S., Lebreton, M., Khamassi, M., Coricelli, G., & Palminteri, S. (2018). Reference-point centering and range-adaptation enhance human reinforcement learning at the cost of irrational preferences [Number: 1 Publisher: Nature Publishing Group]. *Nature Communications*, *9*(1), 4503. https://doi.org/10.1038/s41467-018-06781-2

Bavard, S., & Palminteri, S. (2023). The functional form of value normalization in human reinforcement learning (T. Kahnt, M. J. Frank, & N. Garrett, Eds.) [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *12*, e83891. https://doi.org/10.7554/eLife.83891

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron*, *100*(2), 490–509. https://doi.org/10.1016/j.neuron.2018.10.002

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fmri. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Bein, O., & Niv, Y. (2023, September). Schemas, reinforcement learning, and the medial prefrontal cortex. https://doi.org/10.31234/osf.io/spxq9

Bellemare, M. G., Dabney, W., & Munos, R. (2017). A distributional perspective on reinforcement learning. *International conference on machine learning*, 449–458.

Bellmund, J. L., De Cothi, W., Ruiter, T. A., Nau, M., Barry, C., & Doeller, C. F. (2020). Deforming the metric of cognitive maps distorts memory. *Nature human behaviour*, *4*(2), 177–188.

Bellmund, J. L., Deuker, L., Navarro Schröder, T., & Doeller, C. F. (2016). Grid-cell representations in mental simulation. *Elife*, *5*, e17089.

Bellmund, J. L., Gärdenfors, P., Moser, E. I., & Doeller, C. F. (2018). Navigating cognition: Spatial codes for human thinking. *Science*, *362*(6415), eaat6766.

Ben-Artzi, I., Kessler, Y., Nicenboim, B., & Shahar, N. (2023). Computational mechanisms underlying latent value updating of unchosen actions. *Science advances*, *9*(42), eadi2704.

Bengio, Y., Courville, A., & Vincent, P. (2014, April). Representation Learning: A Review and New Perspectives [arXiv:1206.5538 [cs]]. https://doi.org/10.48550/arXiv.1206.5538

Berns, G. S., Laibson, D., & Loewenstein, G. (2007). Intertemporal choice–toward an integrative framework. *Trends in cognitive sciences*, *11*(11), 482–488.

Best, D., & Roberts, D. (1975). Algorithm as 89: The upper tail probabilities of spearman's rho. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *24*(3), 377–379.

Biderman, N., & Shohamy, D. (2021). Memory and decision making interact to shape the value of unchosen options. *Nature Communications*, *12*(1), 4648. https://doi.org/10.1038/s41467-021-24907-x

Boccara, C. N., Nardin, M., Stella, F., O'Neill, J., & Csicsvari, J. (2019). The entorhinal cognitive map is attracted to goals. *Science*, *363*(6434), 1443–1447.

Bongioanni, A., Folloni, D., Verhagen, L., Sallet, J., Klein-Flügge, M. C., & Rushworth, M. F. (2021). Activation and disruption of a neural mechanism for novel choice in monkeys. *Nature*, *591*(7849), 270–274.

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), e1001093. https://doi.org/10.1371/journal.pbio.1001093

Boorman, E. D., Witkowski, P. P., Zhang, Y., & Park, S. A. (2021). The orbital frontal cortex, task structure, and inference [Place: US Publisher: American Psychological Association]. *Behavioral Neuroscience*, *135*(2), 291–300. https://doi.org/10.1037/bne0000465

Botvinick, M., Wang, J. X., Dabney, W., Miller, K. J., & Kurth-Nelson, Z. (2020). Deep Reinforcement Learning and Its Neuroscientific Implications. *Neuron*, *107*(4), 603–616. https://doi.org/10.1016/j.neuron.2020.06.014

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), 8591–8601.

Bradfield, L. A., & Hart, G. (2020). Rodent medial and lateral orbitofrontal cortices represent unique components of cognitive maps of task space. *Neuroscience & Biobehavioral Reviews*, *108*, 287–294. https://doi.org/10.1016/j.neubiorev.2019.11.009

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial vision*, *10*, 433–436.

Brand, A., Allen, L., Altman, M., Hlava, M., & Scott, J. (2015). Beyond authorship: Attribution, contribution, collaboration, and credit. *Learned Publishing*, *28*(2), 151–155.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*(2), 378–400. https://journal.r-project.org/archive/2017/RJ-2017-066/index.html

Butler, W. N., Hardcastle, K., & Giocomo, L. M. (2019). Remembered reward locations restructure entorhinal spatial maps. *Science*, *363*(6434), 1447–1452.

Castegnetti, G., Zurita, M., & De Martino, B. (2021). How usefulness shapes neural representations during goal-directed behavior [Publisher: American Association for the Advancement of Science]. *Science Advances*, *7*(15), eabd5363. https://doi.org/10.1126/sciadv.abd5363

Cavada, C., Compañy, T., Tejedor, J., Cruz-Rizzolo, R. J., & Reinoso-Suárez, F. (2000). The anatomical connections of the macaque monkey orbitofrontal cortex. a review. *Cerebral cortex*, *10*(3), 220–242.

Chan, S. C. Y., Niv, Y., & Norman, K. A. (2016). A Probability Distribution over Latent Causes, in the Orbitofrontal Cortex. *The Journal of Neuroscience*, *36*(30), 7817–7828. https://doi.org/10.1523/JNEUROSCI.0659-16.2016

Chan, S. C., Schuck, N. W., Lopatina, N., Schoenbaum, G., & Niv, Y. (2021). Orbitofrontal cortex and learning predictions of state transitions. *Behavioral neuroscience*, *135*(4), 487.

Chib, V. S., Rangel, A., Shimojo, S., & O'Doherty, J. P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal

cortex. *Journal of Neuroscience*, *29*(39), 12315–12320. https://doi.org/10.1523/JNEUROSCI.2575-09.2009

Clithero, J. A., & Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. *Social Cognitive and Affective Neuroscience*, *9*(9), 1289–1302. https://doi.org/10.1093/scan/nst106

Conen, K. E., & Padoa-Schioppa, C. (2019). Partial Adaptation to the Value Range in the Macaque Orbitofrontal Cortex [Publisher: Society for Neuroscience Section: Research Articles]. *Journal of Neuroscience*, *39*(18), 3498–3513. https://doi.org/10.1523/JNEUROSCI.2279-18.2019

Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science*, *352*(6292), 1464–1468.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience*, *3*(3), 201–215.

Costa, K. M., Scholz, R., Lloyd, K., Moreno-Castilla, P., Gardner, M. P. H., Dayan, P., & Schoenbaum, G. (2023). The role of the lateral orbitofrontal cortex in creating cognitive maps [Publisher: Nature Publishing Group]. *Nature Neuroscience*, *26*(1), 107–115. https://doi.org/10.1038/s41593-022-01216-0

Cousineau, D., et al. (2005). Confidence intervals in within-subject designs: A simpler solution to loftus and masson's method. *Tutorials in quantitative methods for psychology*, *1*(1), 42–45.

Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fmri data. *NMR in Biomedicine*, *10*(4-5), 171–178. https://doi.org/10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L

Cromwell, H. C., Tremblay, L., & Schultz, W. (2018). Neural encoding of choice during a delayed response task in primate striatum and orbitofrontal cortex. *Experimental Brain Research*, *236*(6), 1679–1688. https://doi.org/10.1007/s00221-018-5253-z

Cross, L., Cockburn, J., Yue, Y., & O'Doherty, J. P. (2021). Using deep reinforcement learning to reveal how the brain encodes abstract state-space representations in high-dimensional environments. *Neuron*, *109*(4), 724–738.e7. https://doi.org/10.1016/j.neuron.2020.11.021

Dabney, W., Kurth-Nelson, Z., Uchida, N., Starkweather, C. K., Hassabis, D., Munos, R., & Botvinick, M. (2020). A distributional code for value in dopamine-based reinforcement learning. *Nature*, *577*(7792), 671–675.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. segmentation and surface reconstruction. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Daniele Bernoulli. (1738). *Specimen Theoriae Novae de Mensura Sortis*. Retrieved February 15, 2024, from http://archive.org/details/SpecimenTheoriaeNovaeDeMensuraSortis

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

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience*, *8*(12), 1704–1711.

Daw, N. D., O'doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879.

Dayan, P. (1993). Improving generalization for temporal difference learning: The successor representation. *Neural computation*, *5*(4), 613–624.

de Bruin, T., Kober, J., Tuyls, K., & Babuška, R. (2018). Integrating State Representation Learning Into Deep Reinforcement Learning [Conference Name: IEEE Robotics and Automation Letters]. *IEEE Robotics and Automation Letters*, *3*(3), 1394–1401. https://doi.org/10.1109/LRA.2018.2800101

De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nature neuroscience*, *16*(1), 105–110.

Derdikman, D., Whitlock, J. R., Tsao, A., Fyhn, M., Hafting, T., Moser, M.-B., & Moser, E. I. (2009). Fragmentation of grid cell maps in a multicompartment environment. *Nature neuroscience*, *12*(10), 1325–1332.

Doeller, C. F., Barry, C., & Burgess, N. (2010). Evidence for grid cells in a human memory network. *Nature*, *463*(7281), 657–661.

Duan, Y., Schulman, J., Chen, X., Bartlett, P. L., Sutskever, I., & Abbeel, P. (2016). Rl^2: Fast reinforcement learning via slow reinforcement learning. *arXiv preprint arXiv:1611.02779*.

Dupret, D., O'neill, J., Pleydell-Bouverie, B., & Csicsvari, J. (2010). The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nature neuroscience*, *13*(8), 995–1002.

Ebitz, R. B., Tu, J. C., & Hayden, B. Y. (2020). Rules warp feature encoding in decision-making circuits [Publisher: Public Library of Science]. *PLOS Biology*, *18*(11), e3000951. https://doi.org/10.1371/journal.pbio.3000951

Eichenbaum, H., Kuperstein, M., Fagan, A., & Nagode, J. (1987). Cue-sampling and goal-approach correlates of hippocampal unit activity in rats performing an odor-discrimination task. *Journal of Neuroscience*, *7*(3), 716–732.

Elliott Wimmer, G., & Büchel, C. (2019). Learning of distant state predictions by the orbitofrontal cortex in humans. *Nature communications*, *10*(1), 2554.

Eppinger, B., Ruel, A., & Bolenz, F. (2023). Diminished state space theory of human aging. *Perspectives on Psychological Science*, 17456916231204811.

Epstein, R. A., Patai, E. Z., Julian, J. B., & Spiers, H. J. (2017). The cognitive map in humans: Spatial navigation and beyond. *Nature neuroscience*, *20*(11), 1504–1513.

Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). Mriqc: Advancing the automatic prediction of image quality in mri from unseen sites. *PloS one*, *12*(9), e0184661.

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., Isik, A. I., Erramuzpe, A., Kent, M., James D. andGoncalves, DuPre, E., Sitek, K. R., Gomez, D. E. P., Lurie, D. J., Ye, Z., Poldrack, R. A., & Gorgolewski, K. J. (2018). Fmriprep. *Software*. https://doi.org/10.5281/zenodo.852659

Esteban, O., Markiewicz, C., Blair, R. W., Moodie, C., Isik, A. I., Erramuzpe Aliaga, A., Kent, J., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S., Wright, J., Durnez, J., Poldrack, R., & Gorgolewski, K. J. (2018). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*. https://doi.org/10.1038/s41592-018-0235-4

Fang, C., & Stachenfeld, K. L. (2023). Predictive auxiliary objectives in deep rl mimic learning in the brain. *arXiv preprint arXiv:2310.06089*.

Farovik, A., Place, R. J., McKenzie, S., Porter, B., Munro, C. E., & Eichenbaum, H. (2015). Orbitofrontal Cortex Encodes Memories within Value-Based Schemas and Represents

Contexts That Guide Memory Retrieval [Publisher: Society for Neuroscience Section: Articles]. *Journal of Neuroscience*, *35*(21), 8333–8344. https://doi.org/10.1523/JNEUROSCI.0134-15.2015

Fehr, E., & Rangel, A. (2011). Neuroeconomic Foundations of Economic Choice–Recent Advances. *Journal of Economic Perspectives*, *25*(4), 3–30. https://doi.org/10.1257/jep.25.4.3

Fellows, L. K. (2007). The Role of Orbitofrontal Cortex in Decision Making [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1196/annals.1401.023]. *Annals of the New York Academy of Sciences*, *1121*(1), 421–430. https://doi.org/10.1196/annals.1401.023

Flesch, T., Balaguer, J., Dekker, R., Nili, H., & Summerfield, C. (2018). Comparing continual task learning in minds and machines. *Proceedings of the National Academy of Sciences*, *115*(44), E10313–E10322.

Flesch, T., Juechems, K., Dumbalska, T., Saxe, A., & Summerfield, C. (2022). Orthogonal representations for robust context-dependent task performance in brains and neural networks. *Neuron*, *110*(7), 1258–1270.e11. https://doi.org/10.1016/j.neuron.2022.01.005

Fonov, V., Evans, A., McKinstry, R., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, *47, Supplement 1*, S102. https://doi.org/10.1016/S1053-8119(09)70884-5

Frank, M. J., & Claus, E. D. (2006). Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychological review*, *113*(2), 300.

Frömer, R., Lin, H., Dean Wolf, C., Inzlicht, M., & Shenhav, A. (2021). Expectations of reward and efficacy guide cognitive control allocation. *Nature Communications*, *12*(1), 1–11.

Frömer, R., Dean Wolf, C. K., & Shenhav, A. (2019). Goal congruency dominates reward value in accounting for behavioral and neural correlates of value-based decision-making [Number: 1 Publisher: Nature Publishing Group]. *Nature Communications*, *10*(1), 4926. https://doi.org/10.1038/s41467-019-12931-x

Frömer, R., & Shenhav, A. (2021). Filling the gaps: Cognitive control as a critical lens for understanding mechanisms of value-based decision-making. *Neuroscience & Biobehavioral Reviews*.

Fusi, S., Miller, E. K., & Rigotti, M. (2016). Why neurons mix: High dimensionality for higher cognition. *Current opinion in neurobiology*, *37*, 66–74.

Gardner, R. J., Hermansen, E., Pachitariu, M., Burak, Y., Baas, N. A., Dunn, B. A., Moser, M.-B., & Moser, E. I. (2022). Toroidal topology of population activity in grid cells. *Nature*, *602*(7895), 123–128.

Garvert, M. M., Dolan, R. J., & Behrens, T. E. (2017). A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *Elife*, *6*, e17086.

Garvert, M. M., Saanum, T., Schulz, E., Schuck, N. W., & Doeller, C. F. (2023). Hippocampal spatio-predictive cognitive maps adaptively guide reward generalization [Publisher: Nature Publishing Group]. *Nature Neuroscience*, *26*(4), 615–626. https://doi.org/10.1038/s41593-023-01283-x

Gaschler, R., Schuck, N. W., Reverberi, C., Frensch, P. A., & Wenke, D. (2019). Incidental covariation learning leading to strategy change. *PloS one*, *14*(1), e0210597.

Gauthier, J. L., & Tank, D. W. (2018). A dedicated population for reward coding in the hippocampus. *Neuron*, *99*(1), 179–193.

Gershman, S. J., Markman, A. B., & Otto, A. R. (2014). Retrospective revaluation in sequential decision making: A tale of two systems. *Journal of Experimental Psychology: General*, *143*(1), 182.

Gherman, S., & Philiastides, M. G. (2018). Human vmpfc encodes early signatures of confidence in perceptual decisions. *elife*, *7*, e38293.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making [_eprint: https://doi.org/10.1146/annurev-psych-120709-145346]. *Annual Review of Psychology*, *62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Gilboa, A., & Marlatte, H. (2017). Neurobiology of Schemas and Schema-Mediated Memory [Publisher: Elsevier]. *Trends in Cognitive Sciences*, *21*(8), 618–631. https://doi.org/10.1016/j.tics.2017.04.013

Ginosar, G., Aljadeff, J., Burak, Y., Sompolinsky, H., Las, L., & Ulanovsky, N. (2021). Locally ordered representation of 3d space in the entorhinal cortex. *Nature*, *596*(7872), 404–409.

Ginosar, G., Aljadeff, J., Las, L., Derdikman, D., & Ulanovsky, N. (2023). Are grid cells used for navigation? on local metrics, subjective spaces, and black holes. *Neuron*, *111*(12), 1858–1875.

Glover, G. H., Li, T.-Q., & Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fmri: Retroicor. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *44*(1), 162–167.

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*(2), 178.

Gonzalez, W. G., Zhang, H., Harutyunyan, A., & Lois, C. (2019). Persistence of neuronal representations through time and damage in the hippocampus. *Science*, *365*(6455), 821–825.

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics*, *5*, 13. https://doi.org/10.3389/fninf.2011.00013

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., et al. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific data*, *3*(1), 1–9.

Gorgolewski, K. J., Esteban, O., Markiewicz, C. J., Ziegler, E., Ellis, D. G., Notter, M. P., Jarecka, D., Johnson, H., Burns, C., Manhães-Savio, A., Hamalainen, C., Yvernault, B., Salo, T., Jordan, K., Goncalves, M., Waskom, M., Clark, D., Wong, J., Loney, F., … Ghosh, S. (2018). Nipype. *Software*. https://doi.org/10.5281/zenodo.596855

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Grieves, R. M., Jedidi-Ayoub, S., Mishchanchuk, K., Liu, A., Renaudineau, S., Duvelle, É., & Jeffery, K. J. (2021). Irregular distribution of grid cell firing fields in rats exploring a 3d volumetric space. *Nature neuroscience*, *24*(11), 1567–1573.

Gross, J., Woelbert, E., Zimmermann, J., Okamoto-Barth, S., Riedl, A., & Goebel, R. (2014). Value Signals in the Prefrontal Cortex Predict Individual Preferences across Reward Categories [Publisher: Society for Neuroscience Section: Articles]. *Journal of Neuroscience*, *34*(22), 7580–7586. https://doi.org/10.1523/JNEUROSCI.5082-13.2014

Grueschow, M., Polania, R., Hare, T. A., & Ruff, C. C. (2015). Automatic versus Choice-Dependent Value Representations in the Human Brain. *Neuron*, *85*(4), 874–885. https://doi.org/10.1016/j.neuron.2014.12.054

Gupta, A. S., Van Der Meer, M. A., Touretzky, D. S., & Redish, A. D. (2010). Hippocampal replay is not a simple function of experience. *Neuron*, *65*(5), 695–705.

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*(7052), 801–806.

Hare, T. A., Camerer, C. F., & Rangel, A. (2009). Self-Control in Decision-Making Involves Modulation of the vmPFC Valuation System [Publisher: American Association for the Advancement of Science]. *Science*, *324*(5927), 646–648. https://doi.org/10.1126/science.1168450

Harnad, S. (2003). Categorical perception.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Harvey, A. H., Kirk, U., Denfield, G. H., & Montague, P. R. (2010). Monetary favors and their influence on neural responses and revealed preference. *Journal of Neuroscience*, *30*(28), 9597–9602.

Harvey, A. K., Pattinson, K. T., Brooks, J. C., Mayhew, S. D., Jenkinson, M., & Wise, R. G. (2008). Brainstem functional magnetic resonance imaging: Disentangling signal from physiological noise. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *28*(6), 1337–1344.

Hattori, R., Hedrick, N. G., Jain, A., Chen, S., You, H., Hattori, M., Choi, J.-H., Lim, B. K., Yasuda, R., & Komiyama, T. (2023). Meta-reinforcement learning via orbitofrontal cortex

[Publisher: Nature Publishing Group]. *Nature Neuroscience*, *26*(12), 2182–2191. https://doi.org/10.1038/s41593-023-01485-3

Hayden, B. Y., & Niv, Y. (2021). The case against economic values in the orbitofrontal cortex (or anywhere else in the brain) [Place: US Publisher: American Psychological Association]. *Behavioral Neuroscience*, *135*(2), 192–201. https://doi.org/10.1037/bne0000448

Hayden, B. Y., Pearson, J. M., & Platt, M. L. (2009). Fictive reward signals in the anterior cingulate cortex. *Science (New York, N.Y.)*, *324*(5929), 948–950. https://doi.org/10.1126/science.1168488

Heess, N., Tb, D., Sriram, S., Lemmon, J., Merel, J., Wayne, G., Tassa, Y., Erez, T., Wang, Z., Eslami, S., et al. (2017). Emergence of locomotion behaviours in rich environments. *arXiv preprint arXiv:1707.02286*.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the Dimensionality of Data with Neural Networks [Publisher: American Association for the Advancement of Science]. *Science*, *313*(5786), 504–507. https://doi.org/10.1126/science.1127647

Hogeveen, J., Hauner, K. K., Chau, A., Krueger, F., & Grafman, J. (2017). Impaired Valuation Leads to Increased Apathy Following Ventromedial Prefrontal Cortex Damage. *Cerebral Cortex*, *27*(2), 1401–1408. https://doi.org/10.1093/cercor/bhv317

Hollander, M., Wolfe, D. A., & Chicken, E. (2013). *Nonparametric statistical methods*. John Wiley & Sons.

Hollup, S. A., Molden, S., Donnett, J. G., Moser, M.-B., & Moser, E. I. (2001). Accumulation of hippocampal place fields at the goal location in an annular watermaze task. *Journal of Neuroscience*, *21*(5), 1635–1644.

Horner, A. J., Bisby, J. A., Zotow, E., Bush, D., & Burgess, N. (2016). Grid-like processing of imagined navigation. *Current Biology*, *26*(6), 842–847.

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.

Hutton, C., Josephs, O., Stadler, J., Featherstone, E., Reid, A., Speck, O., Bernarding, J., & Weiskopf, N. (2011). The impact of physiological noise correction on fmri at 7 t. *Neuroimage*, *57*(1), 101–112.

Izquierdo, A. (2017). Functional heterogeneity within rat orbitofrontal cortex in reward learning and decision making. *Journal of Neuroscience*, *37*(44), 10529–10540.

Jacobs, J., Weidemann, C. T., Miller, J. F., Solway, A., Burke, J. F., Wei, X.-X., Suthana, N., Sperling, M. R., Sharan, A. D., Fried, I., et al. (2013). Direct recordings of grid-like neuronal activity in human spatial navigation. *Nature neuroscience*, *16*(9), 1188–1190.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

Juechems, K., & Summerfield, C. (2019). Where Does Value Come From? [Publisher: Elsevier]. *Trends in Cognitive Sciences*, *23*(10), 836–850. https://doi.org/10.1016/j.tics.2019.07.012

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk [Publisher: [Wiley, Econometric Society]]. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Kahnt, T., Heinzle, J., Park, S. Q., & Haynes, J.-D. (2011). Decoding different roles for vmpfc and dlpfc in multi-attribute decision making. *Neuroimage*, *56*(2), 709–715.

Kaplan, R., Schuck, N. W., & Doeller, C. F. (2017). The role of mental maps in decision-making. *Trends in Neurosciences*, *40*(5), 256–259.

Kasper, L., Bollmann, S., Diaconescu, A. O., Hutton, C., Heinzle, J., Iglesias, S., Hauser, T. U., Sebold, M., Manjaly, Z.-M., Pruessmann, K. P., et al. (2017). The physio toolbox for modeling physiological noise in fmri data. *Journal of neuroscience methods*, *276*, 56–72.

Kennerley, S. W., Dahmubed, A. F., Lara, A. H., & Wallis, J. D. (2009). Neurons in the frontal lobe encode the value of multiple decision variables. *Journal of Cognitive Neuroscience*, *21*(6), 1162–1178. https://doi.org/10.1162/jocn.2009.21100

Kira, S., Safaai, H., Morcos, A. S., Panzeri, S., & Harvey, C. D. (2023). A distributed and efficient population code of mixed selectivity neurons for flexible navigation decisions. *Nature communications*, *14*(1), 2121.

Kiran, B. R., Sobh, I., Talpaert, V., Mannion, P., Sallab, A. A. A., Yogamani, S., & Pérez, P. (2022). Deep Reinforcement Learning for Autonomous Driving: A Survey [Conference Name: IEEE Transactions on Intelligent Transportation Systems]. *IEEE Transactions on*

Intelligent Transportation Systems, *23*(6), 4909–4926. https://doi.org/10.1109/TITS.2021.3054625

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E. C., & Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, *13*(2), e1005350. https://doi.org/10.1371/journal.pcbi.1005350

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., et al. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1.

Knudsen, E. B., & Wallis, J. D. (2022). Taking stock of value in the orbitofrontal cortex [Number: 7 Publisher: Nature Publishing Group]. *Nature Reviews Neuroscience*, *23*(7), 428–438. https://doi.org/10.1038/s41583-022-00589-2

Kobayashi, K., & Hsu, M. (2019). Common neural code for reward and information value [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *116*(26), 13061–13066. https://doi.org/10.1073/pnas.1820145116

Kozma, R., Ilin, R., & Siegelmann, H. T. (2018). Evolution of abstraction across layers in deep learning neural networks. *Procedia computer science*, *144*, 203–213.

Krupic, J., Bauza, M., Burton, S., Barry, C., & O'Keefe, J. (2015). Grid cell symmetry is shaped by environmental geometry. *Nature*, *518*(7538), 232–235.

Krupic, J., Bauza, M., Burton, S., Lever, C., & O'Keefe, J. (2014). How environment geometry affects grid cell symmetry and what we can learn from it. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1635), 20130188.

Krupic, J., Bauza, M., Burton, S., & O'Keefe, J. (2018). Local transformations of the hippocampal cognitive map. *Science*, *359*(6380), 1143–1146.

Kuperwajs, I., Schütt, H. H., & Ma, W. J. (2023). Using deep neural networks as a guide for modeling human planning. *Scientific reports*, *13*(1), 20269.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and brain sciences*, *40*, e253.

Lanczos, C. (1964). Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, *1*(1), 76–85. https://doi.org/10.1137/0701007

Lebreton, M., Abitbol, R., Daunizeau, J., & Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. *Nature neuroscience*, *18*(8), 1159–1167.

Lebreton, M., Jorge, S., Michel, V., Thirion, B., & Pessiglione, M. (2009). An Automatic Valuation System in the Human Brain: Evidence from Functional Neuroimaging. *Neuron*, *64*(3), 431–439. https://doi.org/10.1016/j.neuron.2009.09.040

Lee, D., Seo, H., & Jung, M. W. (2012). Neural basis of reinforcement learning and decision making. *Annual review of neuroscience*, *35*(1), 287–308.

Lee, H., Ghim, J.-W., Kim, H., Lee, D., & Jung, M. (2012). Hippocampal neural correlates for values of experienced events. *Journal of Neuroscience*, *32*(43), 15053–15065.

LeGates, T. A., Kvarta, M. D., Tooley, J. R., Francis, T. C., Lobo, M. K., Creed, M. C., & Thompson, S. M. (2018). Reward behaviour is regulated by the strength of hippocampus–nucleus accumbens synapses. *Nature*, *564*(7735), 258–262.

Lenth, R. V. (2025). *Emmeans: Estimated marginal means, aka least-squares means* [R package version 1.10.7-100001, https://rvlenth.github.io/emmeans/]. https://rvlenth.github.io/emmeans/

Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017). Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron*, *93*(2), 451–463.

Lesort, T., Díaz-Rodríguez, N., Goudou, J.-F., & Filliat, D. (2018). State representation learning for control: An overview. *Neural Networks*, *108*, 379–392. https://doi.org/10.1016/j.neunet.2018.07.006

Levens, S. M., Larsen, J. T., Bruss, J., Tranel, D., Bechara, A., & Mellers, B. A. (2014). What might have been? the role of the ventromedial prefrontal cortex and lateral orbitofrontal cortex in counterfactual emotions and choice. *Neuropsychologia*, *54*, 77–86.

Levy, D. J., & Glimcher, P. W. (2011). Comparing Apples and Oranges: Using Reward-Specific and Reward-General Subjective Value Representation in the Brain [Publisher: Society for Neuroscience Section: Articles]. *Journal of Neuroscience*, *31*(41), 14693–14707. https://doi.org/10.1523/JNEUROSCI.2218-11.2011

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology*, *22*(6), 1027–1038. https://doi.org/10.1016/j.conb.2012.06.001

Li, V., Michael, E., Balaguer, J., Castañón, S. H., & Summerfield, C. (2018). Gain control explains the effect of distraction in human perceptual, cognitive, and economic decision making. *Proceedings of the National Academy of Sciences*, *115*(38), E8825–E8834.

Li, X., Morgan, P. S., Ashburner, J., Smith, J., & Rorden, C. (2016). The first step for neuroimaging data analysis: Dicom to nifti conversion. *Journal of neuroscience methods*, *264*, 47–56.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of experimental psychology*, *54*(5), 358.

Lipton, P. A., Alvarez, P., & Eichenbaum, H. (1999). Crossmodal Associative Memory Representations in Rodent Orbitofrontal Cortex [Publisher: Elsevier]. *Neuron*, *22*(2), 349–359. https://doi.org/10.1016/S0896-6273(00)81095-8

Loewe, A. T., Petzka, M., Tzegka, M., & Schuck, N. W. (2024). N2 sleep inspires insight. *bioRxiv*, 2024–06.

Lopatina, N., McDannald, M. A., Styer, C. V., Sadacca, B. F., Cheer, J. F., & Schoenbaum, G. (2015). Lateral orbitofrontal neurons acquire responses to upshifted, downshifted, or blocked cues during unblocking (H. Eichenbaum, Ed.) [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *4*, e11299. https://doi.org/10.7554/eLife.11299

Lopatina, N., Sadacca, B. F., McDannald, M. A., Styer, C. V., Peterson, J. F., Cheer, J. F., & Schoenbaum, G. (2017). Ensembles in medial and lateral orbitofrontal cortex construct cognitive maps emphasizing different features of the behavioral landscape. *Behavioral Neuroscience*, *131*(3), 201.

Lopez-Persem, A., Bastin, J., Petton, M., Abitbol, R., Lehongre, K., Adam, C., Navarro, V., Rheims, S., Kahane, P., Domenech, P., et al. (2020). Four core properties of the human brain valuation system demonstrated in intracranial signals. *Nature Neuroscience*, *23*(5), 664–675.

Löwe, A. T., Touzo, L., Muhle-Karbe, P. S., Saxe, A. M., Summerfield, C., & Schuck, N. W. (2024, March). Abrupt and spontaneous strategy switches emerge in simple regularised neural networks [arXiv:2302.11351 [cs, q-bio]]. https://doi.org/10.48550/arXiv.2302.11351

Mack, M. L., Preston, A. R., & Love, B. C. (2020). Ventromedial prefrontal cortex compression during concept learning [Number: 1 Publisher: Nature Publishing Group]. *Nature Communications*, *11*(1), 46. https://doi.org/10.1038/s41467-019-13930-8

MacLeod, C. M. (1991). Half a century of research on the stroop effect: An integrative review. *Psychological bulletin*, *109*(2), 163.

Magnusson, A., Skaug, H., Nielsen, A., Berg, C., Kristensen, K., Maechler, M., van Bentham, K., Bolker, B., Brooks, M., & Brooks, M. M. (2017). Package 'glmmtmb'. *R Package Version 0.2. 0*.

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex [Number: 7474 Publisher: Nature Publishing Group]. *Nature*, *503*(7474), 78–84. https://doi.org/10.1038/nature12742

Martino, B. D., & Cortese, A. (2023). Goals, usefulness and abstraction in value-based choice [Publisher: Elsevier]. *Trends in Cognitive Sciences*, *27*(1), 65–80. https://doi.org/10.1016/j.tics.2022.11.001

MATLAB version (R2012b). (2017). The Mathworks, Inc. Natick, Massachusetts.

MATLAB version 9.3.0.713579 (R2017b). (2017). The Mathworks, Inc. Natick, Massachusetts.

McCormick, C., Barry, D. N., Jafarian, A., Barnes, G. R., & Maguire, E. A. (2020). Vmpfc drives hippocampal processing during autobiographical memory recall regardless of remoteness. *Cerebral Cortex*, *30*(11), 5972–5987.

McGrath, T., Kapishnikov, A., Tomašev, N., Pearce, A., Wattenberg, M., Hassabis, D., Kim, B., Paquet, U., & Kramnik, V. (2022). Acquisition of chess knowledge in AlphaZero [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *119*(47), e2206625119. https://doi.org/10.1073/pnas.2206625119

McMurray, B. (2022). The myth of categorical perception. *The Journal of the Acoustical Society of America*, *152*(6), 3819–3842.

McNamee, D., Rangel, A., & O'Doherty, J. P. (2013). Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex [Number: 4 Publisher: Nature Publishing Group]. *Nature Neuroscience*, *16*(4), 479–485. https://doi.org/10.1038/nn.3337

Mehta, P. S., Tu, J. C., LoConte, G. A., Pesce, M. C., & Hayden, B. Y. (2019). Ventromedial Prefrontal Cortex Tracks Multiple Environmental Variables during Search [Publisher: Society for Neuroscience Section: Research Articles]. *Journal of Neuroscience*, *39*(27), 5336–5350. https://doi.org/10.1523/JNEUROSCI.2365-18.2019

Miller, K. J., Botvinick, M. M., & Brody, C. D. (2017). Dorsal hippocampus contributes to model-based planning. *Nature neuroscience*, *20*(9), 1269–1276.

Miller, K. J., Shenhav, A., & Ludvig, E. A. (2019). Habits without values [Place: US Publisher: American Psychological Association]. *Psychological Review*, *126*(2), 292–311. https://doi.org/10.1037/rev0000120

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning [Publisher: Nature Publishing Group]. *Nature*, *518*(7540), 529–533. https://doi.org/10.1038/nature14236

Moita, M. A., Rosis, S., Zhou, Y., LeDoux, J. E., & Blair, H. T. (2003). Hippocampal place cells acquire location-specific responses to the conditioned stimulus during auditory fear conditioning. *Neuron*, *37*(3), 485–497.

Molinaro, G., & Collins, A. G. E. (2023a). Intrinsic rewards explain context-sensitive valuation in reinforcement learning [Publisher: Public Library of Science]. *PLOS Biology*, *21*(7), e3002201. https://doi.org/10.1371/journal.pbio.3002201

Molinaro, G., & Collins, A. G. E. (2023b). A goal-centric outlook on learning [Publisher: Elsevier]. *Trends in Cognitive Sciences*, *27*(12), 1150–1164. https://doi.org/10.1016/j.tics.2023.08.011

Momennejad, I., Otto, A. R., Daw, N. D., & Norman, K. A. (2018). Offline replay supports planning in human reinforcement learning. *elife*, *7*, e32548.

Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N. D., & Gershman, S. J. (2017). The successor representation in human reinforcement learning. *Nature human behaviour*, *1*(9), 680–692.

Moneta, N., Garvert, M. M., Heekeren, H. R., & Schuck, N. W. (2023). Task state representations in vmPFC mediate relevant and irrelevant value signals and their behavioral influence [Number: 1 Publisher: Nature Publishing Group]. *Nature Communications*, *14*(1), 3156. https://doi.org/10.1038/s41467-023-38709-w

Moneta, N., Grossman, S., & Schuck, N. W. (2024). Representational spaces in orbitofrontal and ventromedial prefrontal cortex: Task states, values, and beyond. *Trends in Neurosciences*.

Monk, A. M., Dalton, M. A., Barnes, G. R., & Maguire, E. A. (2021). The role of hippocampal–ventromedial prefrontal cortex neural dynamics in building mental representations. *Journal of cognitive neuroscience*, *33*(1), 89–103.

Monsell, S. (2003). Task switching. *Trends in cognitive sciences*, *7*(3), 134–140.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive hebbian learning. *Journal of neuroscience*, *16*(5), 1936–1947.

Morey, R. D., et al. (2008). Confidence intervals from normalized data: A correction to cousineau (2005). *reason*, *4*(2), 61–64.

Moser, E. I., Kropff, E., & Moser, M.-B. (2008). Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci.*, *31*, 69–89.

Muhle-Karbe, P. S., Sheahan, H., Pezzulo, G., Spiers, H. J., Chien, S., Schuck, N. W., & Summerfield, C. (2023). Goal-seeking compresses neural codes for space in the human hippocampus and orbitofrontal cortex. *Neuron*, *111*(23), 3885–3899.e6. https://doi.org/10.1016/j.neuron.2023.08.021

Muller, T. H., Butler, J. L., Veselic, S., Miranda, B., Wallis, J. D., Dayan, P., Behrens, T. E., Kurth-Nelson, Z., & Kennerley, S. W. (2024). Distributional reinforcement learning in prefrontal cortex. *Nature Neuroscience*, 1–6.

Mumford, J. A., Poline, J.-B., & Poldrack, R. A. (2015). Orthogonalization of regressors in fmri models. *PloS one*, *10*(4), e0126255.

Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving bold activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage*, *59*(3), 2636–2643.

Nassar, M. R., McGuire, J. T., Ritz, H., & Kable, J. W. (2019). Dissociable forms of uncertainty-driven representational change across the human brain. *Journal of Neuroscience*, *39*(9), 1688–1698.

Nelli, S., Braun, L., Dumbalska, T., Saxe, A., & Summerfield, C. (2023). Neural knowledge assembly in humans and neural networks. *Neuron*, *111*(9), 1504–1516.e9. https://doi.org/10.1016/j.neuron.2023.02.014

Neupane, S., Fiete, I., & Jazayeri, M. (2024). Mental navigation in the primate entorhinal cortex. *Nature*, 1–8.

Nicolle, A., Klein-Flügge, M. C., Hunt, L. T., Vlaev, I., Dolan, R. J., & Behrens, T. E. (2012). An agent independent axis for executed and modeled choice in medial prefrontal cortex. *Neuron*, *75*(6), 1114–1121.

Nitsch, A., Garvert, M. M., Bellmund, J. L., Schuck, N. W., & Doeller, C. F. (2024). Grid-like entorhinal representation of an abstract value space during prospective decision making. *Nature Communications*, *15*(1), 1198.

Niv, Y. (2019). Learning task-state representations [Number: 10 Publisher: Nature Publishing Group]. *Nature Neuroscience*, *22*(10), 1544–1553. https://doi.org/10.1038/s41593-019-0470-8

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *Journal of Neuroscience*, *35*(21), 8145–8157.

Noonan, M. P., Chau, B. K. H., Rushworth, M. F. S., & Fellows, L. K. (2017). Contrasting Effects of Medial and Lateral Orbitofrontal Cortex Lesions on Credit Assignment and Decision-Making in Humans. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(29), 7023–7035. https://doi.org/10.1523/JNEUROSCI.0692-17.2017

OC Jordan, H., Navarro, D. M., & Stringer, S. M. (2020). The formation and use of hierarchical cognitive maps in the brain: A neural network model. *Network: Computation in Neural Systems*, *31*(1-4), 37–141.

O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex [Number: 1 Publisher: Nature Publishing Group]. *Nature Neuroscience*, *4*(1), 95–102. https://doi.org/10.1038/82959

O'doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: Insights from neuroimaging. *Current opinion in neurobiology*, *14*(6), 769–776.

O'Keefe, J. (1978). The hippocampus as a cognitive map.

O'Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. *Experimental neurology*, *51*(1), 78–109.

O'Keefe, J., & Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurons. *Nature*, *381*(6581), 425–428.

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

Öngür, D., & Price, J. (2000). The Organization of Networks within the Orbital and Medial Prefrontal Cortex of Rats, Monkeys and Humans. *Cerebral Cortex*, *10*(3), 206–219. https://doi.org/10.1093/cercor/10.3.206

Padoa-Schioppa, C. (2009). Range-adapting representation of economic value in the orbitofrontal cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(44), 14004–14014. https://doi.org/10.1523/JNEUROSCI.3751-09.2009

Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value [Number: 7090 Publisher: Nature Publishing Group]. *Nature*, *441*(7090), 223–226. https://doi.org/10.1038/nature04676

Padoa-Schioppa, C., & Assad, J. A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes of menu [Publisher: Nature Publishing Group]. *Nature Neuroscience*, *11*(1), 95–102. https://doi.org/10.1038/nn2020

Palminteri, S., Khamassi, M., Joffily, M., & Coricelli, G. (2015). Contextual modulation of value signals in reward and punishment learning [Number: 1 Publisher: Nature Publishing Group]. *Nature Communications*, *6*(1), 8096. https://doi.org/10.1038/ncomms9096

Palminteri, S., & Lebreton, M. (2021). Context-dependent outcome encoding in human reinforcement learning. *Current Opinion in Behavioral Sciences*, *41*, 144–151. https://doi.org/10.1016/j.cobeha.2021.06.006

Park, S. A., Miller, D. S., & Boorman, E. D. (2021). Inferences on a multidimensional social hierarchy use a grid-like code. *Nature neuroscience*, *24*(9), 1292–1301.

Park, S. A., Miller, D. S., Nili, H., Ranganath, C., & Boorman, E. D. (2020). Map making: Constructing, combining, and inferring on abstract cognitive maps. *Neuron*, *107*(6), 1226–1238.

Pastor-Bernier, A., Stasiak, A., & Schultz, W. (2021). Reward-specific satiety affects subjective value signals in orbitofrontal cortex during multicomponent economic choice [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *118*(30), e2022650118. https://doi.org/10.1073/pnas.2022650118

Peasgood, T. (2014). Expected utility theory. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 2092–2096). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_962

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pelletier, G., & Fellows, L. K. (2019). A critical role for human ventromedial frontal lobe in value comparison of complex objects based on attribute configuration. *Journal of Neuroscience*, *39*(21), 4124–4132.

Pelli, D. G., & Vision, S. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, *10*, 437–442.

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (2011). *Statistical parametric mapping: The analysis of functional brain images*. Elsevier.

Pessiglione, M., & Daunizeau, J. (2021). Bridging across functional models: The ofc as a value-making neural network. *Behavioral Neuroscience*, *135*(2), 277.

Pilly, P. K., & Seitz, A. R. (2009). What a difference a parameter makes: A psychophysical comparison of random dot motion algorithms. *Vision research*, *49*(13), 1599–1612.

Piray, P., & Daw, N. D. (2021). Linear reinforcement learning in planning, grid fields, and cognitive control. *Nature communications*, *12*(1), 4942.

Plassmann, H., O'doherty, J., & Rangel, A. (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *Journal of neuroscience*, *27*(37), 9984–9988.

Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science*, *310*(5756), 1963–1966.

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fmri. *NeuroImage*, *84*(Supplement C), 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Radvansky, B. A., & Dombeck, D. A. (2018). An olfactory virtual reality system for mice. *Nature communications*, *9*(1), 839.

Reuter, M., Rosas, H. D., & Fischl, B. (2010). Highly accurate inverse consistent registration: A robust approach. *NeuroImage*, *53*(4), 1181–1196. https://doi.org/10.1016/j.neuroimage.2010.07.020

Rich, E. L., & Wallis, J. D. (2016). Decoding subjective decisions from orbitofrontal cortex. *Nature neuroscience*, *19*(7), 973–980.

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks [Publisher: Nature Publishing Group]. *Nature*, *497*(7451), 585–590. https://doi.org/10.1038/nature12160

Rolls, E. T., Deco, G., Huang, C.-C., & Feng, J. (2023). The human orbitofrontal cortex, vmpfc, and anterior cingulate cortex effective connectome: Emotion, memory, and action. *Cerebral cortex*, *33*(2), 330–356.

Rolls, E. T., Huang, C.-C., Lin, C.-P., Feng, J., & Joliot, M. (2020). Automated anatomical labelling atlas 3. *NeuroImage*, *206*, 116189.

Rolls, E. T., Joliot, M., & Tzourio-Mazoyer, N. (2015). Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *Neuroimage*, *122*, 1–5.

RStudio Team. (2020). *Rstudio: Integrated development environment for r*. RStudio, PBC. Boston, MA. http://www.rstudio.com/

Rudebeck, P. H., & Murray, E. A. (2014). The orbitofrontal oracle: Cortical mechanisms for the prediction and evaluation of specific behavioral outcomes. *Neuron*, *84*(6), 1143–1156.

Rushworth, M. F., Kolling, N., Sallet, J., & Mars, R. B. (2012). Valuation and decision-making in frontal cortex: One or many serial or parallel systems? [Decision making]. *Current Opinion in Neurobiology*, *22*(6), 946–955. https://doi.org/https://doi.org/10.1016/j.conb.2012.04.011

Sadacca, B. F., Wied, H. M., Lopatina, N., Saini, G. K., Nemirovsky, D., & Schoenbaum, G. (2018). Orbitofrontal neurons signal sensory associations underlying model-based inference in a sensory preconditioning task (M. J. Frank, Ed.) [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *7*, e30373. https://doi.org/10.7554/eLife.30373

Samuelson, P. A. (1947). Some Implications of "Linearity." 1. *The Review of Economic Studies*, *15*(2), 88–90. https://doi.org/10.2307/2295997

Sandbrink, K., & Summerfield, C. (2024). Modelling cognitive flexibility with deep neural networks. *Current Opinion in Behavioral Sciences*, *57*, 101361.

Sanguinetti-Scheck, J. I., & Brecht, M. (2020). Home, head direction stability, and grid cell distortion. *Journal of neurophysiology*.

Sarel, A., Finkelstein, A., Las, L., & Ulanovsky, N. (2017). Vectorial representation of spatial goals in the hippocampus of bats. *Science*, *355*(6321), 176–180.

Saxe, A. M., McClelland, J. L., & Ganguli, S. (2019). A mathematical theory of semantic development in deep neural networks. *Proceedings of the National Academy of Sciences*, *116*(23), 11537–11546.

Saxe, A. M., Bansal, Y., Dapello, J., Advani, M., Kolchinsky, A., Tracey, B. D., & Cox, D. D. (2018). On the Information Bottleneck Theory of Deep Learning. Retrieved April 23, 2024, from https://openreview.net/forum?id=ry_WPG-A-

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature reviews neuroscience*, *8*(9), 657–661.

Schaffner, J., Bao, S. D., Tobler, P. N., Hare, T. A., & Polania, R. (2023). Sensory perception relies on fitness-maximizing codes. *Nature Human Behaviour*, *7*(7), 1135–1151.

Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature communications*, *6*(1), 1–10.

Schoenbaum, G., & Roesch, M. (2005). Orbitofrontal cortex, associative learning, and expectancies. *Neuron*, *47*(5), 633–636.

Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron*, *91*(6), 1402–1412. https://doi.org/10.1016/j.neuron.2016.08.019

Schuck, N. W., Gaschler, R., Wenke, D., Heinzle, J., Frensch, P. A., Haynes, J.-D., & Reverberi, C. (2015). Medial prefrontal cortex predicts internally driven strategy shifts. *Neuron*, *86*(1), 331–340.

Schuck, N. W., & Niv, Y. (2019). Sequential replay of nonspatial task states in the human hippocampus. *Science*, *364*(6447), eaaw5181.

Schuck, N. W., Wilson, R., & Niv, Y. (2018, January). Chapter 12 - A State Representation for Reinforcement Learning and Decision-Making in the Orbitofrontal Cortex. In R. Morris, A. Bornstein, & A. Shenhav (Eds.), *Goal-Directed Decision Making* (pp. 259–278). Academic Press. https://doi.org/10.1016/B978-0-12-812098-9.00012-7

Schultz, W. (1998). Predictive reward signal of dopamine neurons [PMID: 9658025]. *Journal of Neurophysiology*, *80*(1), 1–27. https://doi.org/10.1152/jn.1998.80.1.1

Shahar, N., Moran, R., Hauser, T. U., Kievit, R. A., McNamee, D., Moutoussis, M., Dolan, R. J., Consortium, N., et al. (2019). Credit assignment to state-independent task representations and its relationship with model-based decision making. *Proceedings of the National Academy of Sciences*, *116*(32), 15871–15876.

Shapiro, A. D., & Grafton, S. T. (2020). Subjective value then confidence in human ventromedial prefrontal cortex. *Plos one*, *15*(2), e0225617.

Sharpe, M. J., Stalnaker, T., Schuck, N. W., Killcross, S., Schoenbaum, G., & Niv, Y. (2019). An integrated model of action selection: Distinct modes of cortical control of striatal decision making. *Annual review of psychology*.

Shenhav, A., Straccia, M. A., Musslick, S., Cohen, J. D., & Botvinick, M. M. (2018). Dissociable neural mechanisms track evidence accumulation for selection of attention versus action. *Nature communications*, *9*.

Shi, W., Meisner, O. C., Blackmore, S., Jadi, M. P., Nandy, A. S., & Chang, S. W. C. (2023). The orbitofrontal cortex: A goal-directed cognitive map framework for social and non-social behaviors. *Neurobiology of Learning and Memory*, *203*, 107793. https://doi.org/10.1016/j.nlm.2023.107793

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search [Publisher: Nature Publishing Group]. *Nature*, *529*(7587), 484–489. https://doi.org/10.1038/nature16961

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2017, December). Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm [arXiv:1712.01815 [cs]]. https://doi.org/10.48550/arXiv.1712.01815

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play [Publisher: American Association for the Advancement of Science]. *Science*, *362*(6419), 1140–1144. https://doi.org/10.1126/science.aar6404

Silver, D., Singh, S., Precup, D., & Sutton, R. S. (2021). Reward is enough. *Artificial Intelligence*, *299*, 103535. https://doi.org/10.1016/j.artint.2021.103535

Solstad, T., Boccara, C. N., Kropff, E., Moser, M.-B., & Moser, E. I. (2008). Representation of geometric borders in the entorhinal cortex. *Science*, *322*(5909), 1865–1868.

Song, H. F., Yang, G. R., & Wang, X.-J. (2017). Reward-based training of recurrent neural networks for cognitive and value-based tasks (T. E. Behrens, Ed.) [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *6*, e21492. https://doi.org/10.7554/eLife.21492

Sosa, M., & Giocomo, L. M. (2021). Navigating for reward. *Nature Reviews Neuroscience*, *22*(8), 472–487.

Sosa, M., Plitt, M. H., & Giocomo, L. M. (2024). Hippocampal sequences span experience relative to rewards. *bioRxiv*, 2023–12.

Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature neuroscience*, *20*(11), 1643–1653.

Stoewer, P., Schlieker, C., Schilling, A., Metzner, C., Maier, A., & Krauss, P. (2022). Neural network based successor representations to form cognitive maps of space and language. *Scientific Reports*, *12*(1), 11233.

Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, *78*(2), 364–375.

Stoll, F. M., & Rudebeck, P. H. (2024). Preferences reveal dissociable encoding across prefrontal-limbic circuits [Publisher: Elsevier]. *Neuron*, *0*(0). https://doi.org/10.1016/j.neuron.2024.03.020

Suhaimi, A., Lim, A. W. H., Chia, X. W., Li, C., & Makino, H. (2022). Representation learning in the artificial and biological neural networks underlying sensorimotor integration [Publisher: American Association for the Advancement of Science]. *Science Advances*, *8*(22), eabn0984. https://doi.org/10.1126/sciadv.abn0984

Sutton, R. S. (1991). Dyna, an integrated architecture for learning, planning, and reacting. *ACM Sigart Bulletin*, *2*(4), 160–163.

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

Suzuki, S., Cross, L., & O'Doherty, J. P. (2017). Elucidating the underlying components of food valuation in the human orbitofrontal cortex. *Nature neuroscience*, *20*(12), 1780–1786.

Takagi, Y., Hunt, L. T., Woolrich, M. W., Behrens, T. E., & Klein-Flugge, M. (2020). Projections of non-invasive human recordings into state space show unfolding of spontaneous and over-trained choice. *bioRxiv*.

Takagi, Y., Hunt, L. T., Woolrich, M. W., Behrens, T. E., & Klein-Flügge, M. C. (2021). Adapting non-invasive human recordings along multiple task-axes shows unfolding of spontaneous and over-trained choice (J. M. Palva, J. I. Gold, & L. C. Parra, Eds.) [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *10*, e60988. https://doi.org/10.7554/eLife.60988

Tang, W., Shin, J. D., & Jadhav, S. P. (2021). Multiple time-scales of decision-making in the hippocampus and prefrontal cortex. *Elife*, *10*, e66227.

Taube, J. S., Muller, R. U., & Ranck, J. B. (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. i. description and quantitative analysis. *Journal of Neuroscience*, *10*(2), 420–435.

Terada, S., Sakurai, Y., Nakahara, H., & Fujisawa, S. (2017). Temporal and rate coding for discrete event sequences in the hippocampus. *Neuron*, *94*(6), 1248–1262.

Tessereau, C., Xuan, F., Mellor, J. R., Dayan, P., & Dombeck, D. (2025). Navigating uncertainty: Reward location variability induces reorganization of hippocampal spatial representations. *bioRxiv*.

Thalmann, M., Schäfer, T. A., Theves, S., Doeller, C. F., & Schulz, E. (2024). Task imprinting: Another mechanism of representational change? *Cognitive Psychology*, *152*, 101670.

Theves, S., Fernandez, G., & Doeller, C. F. (2019). The hippocampus encodes distances in multidimensional feature space. *Current Biology*, *29*(7), 1226–1231.

Theves, S., Fernández, G., & Doeller, C. F. (2020). The hippocampus maps concept space, not feature space. *Journal of Neuroscience*, *40*(38), 7318–7325.

Theves, S., Schäfer, T. A., Reisner, V., de Cothi, W., & Barry, C. (2024). Category boundaries modulate memory in a place-cell-like manner. *Current Biology*, *34*(23), 5546–5553.

Tobia, M. J., Guo, R., Schwarze, U., Boehmer, W., Gläscher, J., Finckh, B., Marschner, A., Büchel, C., Obermayer, K., & Sommer, T. (2014). Neural systems for choice and valuation with counterfactual learning signals. *NeuroImage*, *89*, 57–69. https://doi.org/10.1016/j.neuroimage.2013.11.051

Tobia, M. J., Guo, R., Schwarze, U., Böhmer, W., Gläscher, J., Finckh, B., Marschner, A., Büchel, C., Obermayer, K., & Sommer, T. (2014). Neural systems for choice and valuation with counterfactual learning signals. *Neuroimage*, *89*, 57–69.

Tolman, E. C., & Honzik, C. H. (1930a). Introduction and removal of reward, and maze performance in rats. *University of California Publications in Psychology*, *4*, 257–275.

Tolman, E. C. (1948). Cognitive maps in rats and men [Place: US Publisher: American Psychological Association]. *Psychological Review*, *55*(4), 189–208. https://doi.org/10.1037/h0061626

Tolman, E. C., Ritchie, B. F., & Kalish, D. (1946). Studies in spatial learning. i. orientation and the short-cut. *Journal of experimental psychology*, *36*(1), 13.

Tolman, E. C., & Honzik, C. H. (1930b). Introduction and removal of reward, and maze performance in rats. *University of California publications in psychology*.

Tremblay, L., & Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. *Nature*, *398*(6729), 704–708.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4itk: Improved n3 bias correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908

Tye, K. M., Miller, E. K., Taschbach, F. H., Benna, M. K., Rigotti, M., & Fusi, S. (2024). Mixed selectivity: Cellular computations for complexity. *Neuron*, *112*(14), 2289–2303. https://doi.org/https://doi.org/10.1016/j.neuron.2024.04.017

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, *15*(1), 273–289.

Vaidya, A. R., & Fellows, L. K. (2020, February). *Under construction: Ventral and lateral frontal lobe contributions to value-based decision-making and learning* (tech. rep. No. 9:158) (Type: article). F1000Research. https://doi.org/10.12688/f1000research.21946.1

Vlaev, I., Chater, N., Stewart, N., & Brown, G. D. (2011). Does the brain calculate value? *Trends in cognitive sciences*, *15*(11), 546–554.

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage*, *137*, 188–200.

Wang, J. X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J. Z., Hassabis, D., & Botvinick, M. (2018). Prefrontal cortex as a meta-reinforcement learning system

[Publisher: Nature Publishing Group]. *Nature Neuroscience*, *21*(6), 860–868. https://doi.org/10.1038/s41593-018-0147-8

Wang, M. Z., Hayden, B. Y., & Heilbronner, S. R. (2022). A structural and functional subdivision in central orbitofrontal cortex. *Nature communications*, *13*(1), 3623.

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 3t and 1.5t. *NeuroImage*, *33*(2), 493–504. https://doi.org/https://doi.org/10.1016/j.neuroimage.2006.07.029

Westbrook, A., Lamichhane, B., & Braver, T. (2019). The Subjective Value of Cognitive Effort is Encoded by a Domain-General Valuation Network [Publisher: Society for Neuroscience Section: Research Articles]. *Journal of Neuroscience*, *39*(20), 3934–3947. https://doi.org/10.1523/JNEUROSCI.3071-18.2019

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

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wierda, T., Dora, S., Pennartz, C. M. A., & Mejias, J. F. (2023, November). Diverse and flexible behavioral strategies arise in recurrent neural networks trained on multisensory decision making [Pages: 2023.10.28.564511 Section: New Results]. https://doi.org/10.1101/2023.10.28.564511

Wikenheiser, A. M., Marrero-Garcia, Y., & Schoenbaum, G. (2017). Suppression of Ventral Hippocampal Output Impairs Integrated Orbitofrontal Encoding of Task Structure. *Neuron*, *95*(5), 1197–1207.e3. https://doi.org/10.1016/j.neuron.2017.08.003

Wikenheiser, A. M., & Schoenbaum, G. (2016). Over the river, through the woods: Cognitive maps in the hippocampus and orbitofrontal cortex [Publisher: Nature Publishing Group]. *Nature Reviews Neuroscience*, *17*(8), 513–523. https://doi.org/10.1038/nrn.2016.56

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. https://doi.org/10.1016/j.neuron.2013.11.005

Wimmer, G. E., & Büchel, C. (2019). Learning of distant state predictions by the orbitofrontal cortex in humans. *Nature communications*, *10*(1), 1–11.

Winston, J. S., Vlaev, I., Seymour, B., Chater, N., & Dolan, R. J. (2014). Relative valuation of pain in human orbitofrontal cortex. *Journal of Neuroscience*, *34*(44), 14526–14535.

Wirth, S., Avsar, E., Chiu, C. C., Sharma, V., Smith, A. C., Brown, E., & Suzuki, W. A. (2009). Trial outcome and associative learning signals in the monkey hippocampus. *Neuron*, *61*(6), 930–940.

Wittkuhn, L., Chien, S., Hall-McMaster, S., & Schuck, N. W. (2021). Replay in minds and machines. *Neuroscience & Biobehavioral Reviews*, *129*, 367–388.

Wittkuhn, L., Krippner, L. M., Koch, C., & Schuck, N. W. (2022). Replay in human visual cortex is linked to the formation of successor representations and independent of consciousness. *BioRxiv*, 2022–02.

Wittkuhn, L., & Schuck, N. W. (2021). Dynamics of fmri patterns reflect sub-second activation sequences and reveal replay in human visual cortex. *Nature communications*, *12*(1), 1–22.

Yamada, H., Louie, K., Tymula, A., & Glimcher, P. W. (2018). Free choice shapes normalized value signals in medial orbitofrontal cortex [Number: 1 Publisher: Nature Publishing Group]. *Nature Communications*, *9*(1), 162. https://doi.org/10.1038/s41467-017-02614-w

Yao, Y.-W., Song, K.-R., Schuck, N. W., Li, X., Fang, X.-Y., Zhang, J.-T., Heekeren, H. R., & Bruckner, R. (2023). The dorsomedial prefrontal cortex represents subjective value across effort-based and risky decision-making. *NeuroImage*, *279*, 120326. https://doi.org/10.1016/j.neuroimage.2023.120326

Yoo, S. B. M., & Hayden, B. Y. (2020). The Transition from Evaluation to Selection Involves Neural Subspace Reorganization in Core Reward Regions. *Neuron*, *105*(4), 712–724.e4. https://doi.org/10.1016/j.neuron.2019.11.013

Yoo, S. B. M., Sleezer, B. J., & Hayden, B. Y. (2018). Robust Encoding of Spatial Information in Orbitofrontal Cortex and Striatum. *Journal of Cognitive Neuroscience*, *30*(6), 898–913. https://doi.org/10.1162/jocn_a_01259

Yoshimoto, T., Okazaki, S., Sumiya, M., Takahashi, H. K., Nakagawa, E., Koike, T., Kitada, R., Okamoto, S., Nakata, M., Yada, T., Kosaka, H., Sadato, N., & Chikazoe, J. (2022).

Coexistence of sensory qualities and value representations in human orbitofrontal cortex. *Neuroscience Research*, *180*, 48–57. https://doi.org/10.1016/j.neures.2022.02.004

Zeithamova, D., Gelman, B. D., Frank, L., & Preston, A. R. (2018). Abstract representation of prospective reward in the hippocampus. *Journal of Neuroscience*, *38*(47), 10093–10101.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57. https://doi.org/10.1109/42.906424

Zhang, Z., Cheng, Z., Lin, Z., Nie, C., & Yang, T. (2018). A neural network model for the orbitofrontal cortex and task space acquisition during reinforcement learning [Publisher: Public Library of Science]. *PLOS Computational Biology*, *14*(1), e1005925. https://doi.org/10.1371/journal.pcbi.1005925

Zhang, Z., Fanning, J., Ehrlich, D. B., Chen, W., Lee, D., & Levy, I. (2017). Distributed neural representation of saliency controlled value and category during anticipation of rewards and punishments [Number: 1 Publisher: Nature Publishing Group]. *Nature Communications*, *8*(1), 1907. https://doi.org/10.1038/s41467-017-02080-4

Zheng, X. Y., Hebart, M. N., Grill, F., Dolan, R. J., Doeller, C. F., Cools, R., & Garvert, M. M. (2024). Parallel cognitive maps for multiple knowledge structures in the hippocampal formation. *Cerebral Cortex*, *34*(2), bhad485.

Zhou, J., Gardner, M. P. H., Stalnaker, T. A., Ramus, S. J., Wikenheiser, A. M., Niv, Y., & Schoenbaum, G. (2019). Rat Orbitofrontal Ensemble Activity Contains Multiplexed but Dissociable Representations of Value and Task Structure in an Odor Sequence Task. *Current Biology*, *29*(6), 897–907.e3. https://doi.org/10.1016/j.cub.2019.01.048

Zhou, J., Jia, C., Montesinos-Cartagena, M., Gardner, M. P. H., Zong, W., & Schoenbaum, G. (2021). Evolving schema representations in orbitofrontal ensembles during learning [Number: 7847 Publisher: Nature Publishing Group]. *Nature*, *590*(7847), 606–611. https://doi.org/10.1038/s41586-020-03061-2

Zielinski, M. C., Shin, J. D., & Jadhav, S. P. (2019). Coherent coding of spatial position mediated by theta oscillations in the hippocampus and prefrontal cortex. *Journal of Neuroscience*, *39*(23), 4550–4565.

Zimmermann, J., Glimcher, P. W., & Louie, K. (2018). Multiple timescales of normalized value coding underlie adaptive choice behavior [Number: 1 Publisher: Nature Publishing Group]. *Nature Communications*, *9*(1), 3206. https://doi.org/10.1038/s41467-018-05507-8

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR PSYCHOLOGIE UND
BEWEGUNGSWISSENSCHAFT
Institut für Bewegungswissenschaft
Institut für Psychologie

Declaration according to *(please mark with a cross where applicable)*

☐ **§ 4 (1c) of the Doctoral Degree Regulations of the Institute of Human Movement Science at the Universität Hamburg dated 18.08.2010**

☒ **§ 5 (4d) of the Doctoral Degree Regulations of the Institute of Psychology at the Universität Hamburg dated 20.08.2003**

I herby declare,

Nir Moneta (surname, name),

that I have not yet taken a doctoral examination at another university or faculty or applied for admission to a doctoral examination.

Berlin, 01.03.25
place, date

signature

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR PSYCHOLOGIE UND
BEWEGUNGSWISSENSCHAFT
Institut für Bewegungswissenschaft
Institut für Psychologie

Statutory declaration according to *(please mark with a cross where applicable)*

☐ **§ 7 (4) of the Doctoral Degree Regulations of the Institute of Human Movement Science at the Universität Hamburg dated 18.08.2010**

☒ **§ 9 (1c and 1d) of the Doctoral Degree Regulations of the Institute of Psychology at the Universität Hamburg dated 20.08.2003**

I hereby declare on oath,

1. that the dissertation I have submitted has not been the subject of another examination procedure or has not been assessed as unsatisfactory in such a procedure.
2. that I have written the dissertation I have submitted myself, have not used any sources or aids other than those specified and have not taken advantage of any commercial doctoral advice. I have marked the passages taken verbatim or in terms of content as such.

Berlin 01.03.25

place, date

signature