NEURAL AND COMPUTATIONAL REPRESENTATIONS OF DECISION VARIABLES

Thesis by

Daniel McNamee

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2015

(Defended November 20th, 2014)

ACKNOWLEDGEMENTS

I would like to thank my advisor John O’Doherty. His input was crucial in my transition from a purely theoretical background to the world of experimental neuroscience. Gaining an intuition for the interaction between theory and data and designing research to match the two are just some of the critical skills necessary for successful research that I learned in his lab. I’m also very grateful for the space he afforded me to explore several different directions of research while always making himself available to provide advice and guidance.

Over the years, I’ve had instructive and interesting interactions with all my committee members, which includes Antonio Rangel, Ralph Adolphs, Richard Andersen, and Pietro Perona, and I thank them for that. Looking back, I learned an incredible range of topics from them through various courses, rotations, and seminars, including decision theory, neurobiology, monkey electrophysiology, and neural network theory. In particular, Antonio worked with me on the project described in Chapter 2, and I’m continuing to benefit from his perspective. I’d also like to thank Peter Bossaerts for his advice on computational models related to the work in Chapter 5.

I’ve had fun times and productive scientific discussions with my CNS classmates, Nikki Sullivan, Bo Chen, Soyoung Park, Yong-Jun Lin, and Shuo Wang. Beyond the realm of CNS, Tim Blasius has been a seemingly unending source of humor and lessons on America.

My lab mates over the years have been a great source of ideas, motivation, and distraction. Special mention goes to the “originals” who were there when I started in Dublin – Vik, Mimi, Tobias, Simon, Charlotte, Ryan, and Jeff – but not forgetting the more current crewmembers, Sang Wan, Wolfgang, Sven, Shinsuke, Ryo.

Most importantly, I’d like to thank my parents for buying me every book I ever asked for, my sister for repeatedly comparing me to Sheldon, Rachel for her daily encouragements, and all for their support over the course of my studies.

ABSTRACT

These studies explore how, where, and when representations of variables critical to decision-making are represented in the brain. In order to produce a decision, humans must first determine the relevant stimuli, actions, and possible outcomes before applying an algorithm that will select an action from those available. When choosing amongst alternative stimuli, the framework of value-based decision-making proposes that values are assigned to the stimuli and that these values are then compared in an abstract “value space” in order to produce a decision. Despite much progress, in particular regarding the pinpointing of ventromedial prefrontal cortex (vmPFC) as a region that encodes the value, many basic questions remain. In Chapter 2, I show that distributed BOLD signaling in vmPFC represents the value of stimuli under consideration in a manner that is independent of the type of stimulus it is. Thus the open question of whether value is represented in abstraction, a key tenet of value-based decision-making, is confirmed. However, I also show that stimulus-dependent value representations are also present in the brain during decision-making and suggest a potential neural pathway for stimulus-to-value transformations that integrates these two results.

More broadly speaking, there is both neural and behavioral evidence that two distinct control systems are at work during action selection. These two systems compose the “goal-directed system”, which selects actions based on an internal model of the environment, and the “habitual” system, which generates responses based on antecedent stimuli only. Computational characterizations of these two systems imply that they have different informational requirements in terms of input stimuli, actions, and possible outcomes. Associative learning theory predicts that the habitual system should utilize stimulus and action information only, while goal-directed behavior requires that outcomes as well as stimuli and actions be processed. In Chapter 3, I test whether areas of the brain hypothesized to be involved in habitual versus goal-directed control represent the corresponding theorized variables.

The question of whether one or both of these neural systems drives Pavlovian conditioning is less well-studied. Chapter 4 describes an experiment in which subjects were scanned while engaged in a Pavlovian task with a simple non-trivial structure. After comparing a variety of model-based and model-free learning algorithms (thought to underpin goal-directed and habitual decision-making, respectively), it was found that subjects' reaction times were better explained by a model-based system. In addition, neural signaling of precision, a variable based on a representation of a world model, was found in the amygdala. These data indicate that the influence of model-based representations of the environment can extend even to the most basic learning processes.

Knowledge of the state of hidden variables in an environment is required for optimal inference regarding the abstract decision structure of a given environment and therefore can be crucial to decision-making in a wide range of situations. Inferring the state of an abstract variable requires the generation and manipulation of an internal representation of beliefs over the values of the hidden variable. In Chapter 5, I describe behavioral and neural results regarding the learning strategies employed by human subjects in a hierarchical state-estimation task. In particular, a comprehensive model fit and comparison process pointed to the use of "belief thresholding". This implies that subjects tended to eliminate low-probability hypotheses regarding the state of the environment from their internal model and ceased to update the corresponding variables. Thus, in concert with incremental Bayesian learning, humans explicitly manipulate their internal model of the generative process during hierarchical inference consistent with a serial hypothesis testing strategy.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND TABLES

Figures

Supplementary Figures

Tables

Supplementary Tables

NOMENCLATURE

EUT – Expected Utility Theory

PFC – Prefrontal Cortex

dlPFC - Dorsolateral Prefrontal Cortex

vmPFC – Ventromedial Prefrontal Cortex

OFC – Orbitofrontal Cortex

cOFC – Central Orbitofrontal Cortex

dmPFC – Dorsomedial Prefrontal Cortex

FPC – Frontopolar Cortex

MDP – Markov Decision Process

TD – Temporal Difference

RL – Reinforcement Learning

ROI – Region of Interest

MNI – Montreal Neurological Institute

PPI - Psychophysiological interactions

RT – Reaction Time

MLE – Maximum Likelihood Estimation

HBA – Hierarchical Bayesian Analysis

Chapter 1

INTRODUCTION

Decision Neuroscience and Neuroimaging

Through action we exert control over the world. With control, we increase the likelihood of our survival (Fuster, 2008). From homeostatic regulation (Rangel, 2013) to risk management (Symmonds, Bossaerts, & Dolan, 2010), the selection of good actions is a critical task for any nervous system. Understanding the neural and computational mechanisms of decision-making will aid us in the treatment of its pathologies (P. Read Montague, Dolan, Friston, & Dayan, 2012) and also, in a more distant future, enable us to create machines that might exhibit some signs of intelligent behavior (Russell & Norvig, 2009). In addition, cross-species (B. Balleine & O'Doherty, 2010; Rushworth, Mars, & Summerfield, 2009) and multi-scale comparative analyses indicate that at least some neurobiological, algorithmic, and computational principles of decision-making (Franklin & Wolpert, 2011) are conserved across many domains of action. Thus, elucidating the decision-making process is of fundamental importance to understanding both the brain and the human experience.

For many years, a major obstacle in this research program was the lack of *in vivo* recordings of neural signaling in the human brain with reasonable spatial and temporal resolutions. However, advances in non-invasive brain recording techniques have allowed us unprecedented access to functional activity within the human brain. In particular, in 1992 a seminal study (Ogawa et al., 1992) showed that a neuroimaging technique known as functional magnetic resonance imaging (fMRI) allowed us to non-invasively record blood-oxygen-level dependent (BOLD) signals in the brain, effectively using our own blood as a contrast agent for magnetic resonance. Neural activity leads to increased energy demands in order to hyperpolarize cells following spiking, and the BOLD signal measures the concomitant relative increase in regional oxygenated blood flow. Due to the lack of physical damage (for example, with intracranial recording) or effects of radiation (e.g., positron emission tomography) and comparatively excellent data resolutions, fMRI came to dominate modern neuroimaging. Combined with the application of sophisticated machine learning algorithms to analyze this neural data and the computational modeling of behavior in order to make predictions regarding the internal variables used by the brain, recent years have seen rapid advances in our understanding of human decision-making from a coarse neuro-computational standpoint (John P. O'Doherty, 2011; Rangel & Hare, 2010; Rushworth & Behrens, 2008).

Multi-Voxel Pattern Analysis

A typical fMRI-based experiment consists of three steps: (i) the acquisition of imaging data while subjects perform a task, (ii) the pre-processing of said data, usually automated by a software package such as FSL (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/) or SPM (http://www.fil.ion.ucl.ac.uk/spm/), and (iii) the general linear modeling (GLM) of voxel responses via a haemodynamic response model as a function of task variables. This allows one to map predicted neural signals onto distinct regions of the brain, for example, BOLD signals in visual cortex would transiently increase when a stimulus is presented on-screen and motor cortex would be more active when a response is performed. In 2001, a novel approach to fMRI analysis was successfully attempted (Haxby et al., 2001) which is now known as multi-voxel pattern analysis (MVPA). In this study, distinct exemplars of faces and houses were “decoded” from patterns of BOLD signaling while subjects passively viewed images of these stimuli. More specifically, volumes of neural data were labeled with the exemplar being viewed during that scan, then the data was split into “training” and “testing” data before a classification algorithm (Bishop, 2006) or “classifier” learned how to distinguish between neural samples contained in the training data as a function of the labels. Finally, the classifier was asked to predict the labels of the held-out data samples in order to estimate a generalization accuracy score. If this score was significantly above chance then it was determined that there existed multi-voxel activity patterns within the neural data which encoded the identities of the exemplars in question.

To this day, most MVPA is based on variations of this basic analysis template. The progress made using this approach is exemplified by the fact that less than ten years later, a movie being watched by subjects in the scanner could be re-constructed (Nishimoto, Vu, & Gallant, 2010) from independent clips of film (drawn from the Youtube video website). Essentially, MVPA reverses step (iii) of general linear modeling by attempting to predict decision variables from the neural data rather than predicting neural data from decision variables (Naselaris, Kay, Nishimoto, & Gallant, 2011; Pereira, Mitchell, & Botvinick, 2009). Since a model mapping data to variables does not necessarily require a response model defined *a priori*, MVPA attempts to take advantage of fine-grained statistical correlations across multiple voxels in order to makes its predictions (Haynes & Rees, 2006). The source of these distributed patterns of BOLD signaling, whether it be based on sampling from varied distributions of neuronal populations (Kamitani & Tong, 2005), coarse regional connectivity profiles and neural organization (Op de Beeck, 2010), or complex spatiotemporal filtering (Kriegeskorte, Cusack, & Bandettini, 2010), is still an open question. However many predictions regarding the representational contents of brain algorithms (which do not depend on the specification of a neuronal model) can be tested using MVPA only and not via GLM. For example, if we hypothesized that a brain region X contains motor instructions for a particular response A, a GLM might conclude that region X is active while such a representation is made, but only the analysis of the pattern of activity within X would be conclude that a specific motor command A is being represented and not another motor command B. Furthermore, representational dissimilarity analysis (Kriegeskorte, 2009; Mur, Bandettini, & Kriegeskorte, 2009) (one of myriad of MVPA techniques) can produce a distance function which relates objects cognitively in a “space” of internal representations, thus addressing the question of *how* stimuli are represented in cognition and not just where in the brain.

Feature selection forms a crucial part of MVPA, that is, what voxels do we input into our pattern analysis algorithm? The first paper (Haxby et al., 2001) to apply MVPA to neuroimaging data simply entered all available voxels into their decoding algorithm. Although this increases the chance of a significant classification by maximizing the number of features that the algorithm can take advantage of, it does so at the expense of localizing the representations to a specific region. One particular approach to this issue, which I employ in two studies in this thesis, is the “searchlight” technique (Chen et al., 2011; Kriegeskorte, Goebel, & Bandettini, 2006). Briefly, MVPA is performed within a sphere of voxels centered at each voxel in the brain. This allows one to assign a decoding score to each voxel in the brain based on the information available locally around that voxel. After smoothing and performing group-level statistical tests on these “decoding maps” of the brain, the experimenter can conclude where in the brain specific identities are being represented. In summary then, MVPA can estimate what, where, and how stimuli and variables are represented in the brain. Knowing what is being represented in the brain at specific timepoints during a decision progress significantly restricts the model space of plausible algorithms being implemented.

Value-Based Decision-Making

Many decision-making processes are modeled computationally as a value comparison problem. That is, a subjective value is assigned to each potential option and then these values are compared in order to produce a decision. This notion of value-based decision-making has a long history in machine learning (R. Sutton & Barto, 1998) and psychology (Schultz, 2006) and forms the cornerstone of micro-economic theory in the form of expected utility (von Neumann & Morgenstern, 1944). Though expected utility theory (EUT) has been shown to fail in some circumstances (Allais paradox) and leads to paradoxical predictions in others

(Ellsberg paradox), the core theoretical assumption of value representation and comparison has remained intact. In fact, refinements of EUT are generally thought of as context- and subject-dependent variations in what is considered valuable, and thus are modeled as adjustments to the "final common signal" of value (prospect theory). Thus, centuries of decision-making theories and experiments have deemed value to be of fundamental importance and therefore support a strong hypothesis that value is of neurobiological relevance.

For many years, lesion studies were the only means of investigating the neurobiological foundry of decision-making. Despite the imprecise localization of lesion mapping both in the anatomical and functional domains, a broad consensus was formed that ventromedial prefrontal cortex (vmPFC) was of critical importance in action selection and inhibitory control. Lesions of this region lead to inappropriate and impulsive decision-making both from economic and social points of view, most famously exemplified in the case of Phineas Gage (Damasio, Grabowski, Frank, Galaburda, & Damasio, 1994; Harlow, 1848). Through carefully designed behavioral experiments, it became understood that, algorithmically, patients who suffered from a lesion of this brain region had trouble prospectively evaluating decisions and their outcomes and also integrating environment feedback into future decisions (Bechara, Damasio, Damasio, & Anderson, 1994; Szczepanski & Knight, 2014). Computationally, this effect is probably best observed in reversal learning tasks where patients perseverate in making decisions that are no longer rewarding (Fellows & Farah, 2003). Both of these deficits are consistent with an impaired ability to assign values to decisions. Thus, it was not unexpected to find that one of the most consistent BOLD signals subsequently identified in the human brain was a parametric correlate of subjective value in vmPFC (John P. O'Doherty, 2004).

Value Abstraction

A critical assumption is that values are represented and compared in a manner divorced from the stimulus to which they are associated, thus addressing the problem of value comparison between distinct stimuli with no common features (for example, a risky gamble and a chocolate bar). Neurally, one would hypothesize that there exists a region of the brain in which value is represented in the same manner regardless of the specific stimulus under consideration. In a series of electrophysiological experiments in which monkeys made binary choices between options which varied in the identity and amount of juice, it was shown that distinct sets of cells in orbitofrontal cortex (OFC) represented, via firing rate, the offer values of the two juice options and also the value of the chosen option regardless of the motor response requirements or sensory aspects of the task (Padoa-Schioppa & Assad, 2006). In a follow-up, the same experimenters found that these neural representations of economic value were “menu-invariant” (Padoa-Schioppa & Assad, 2008), consistent with the requirement of transitivity in EUT. That is, cells represented the value of a juice regardless of the alternative option in the choice. Interestingly, in contrast, some cells in OFC exhibited binary responses to the identity of a juice and did not reflect the amount or value of said juice. Furthermore, in other monkey electrophysiological experiments (S. Kennerley, Behrens, & Wallis, 2011; S. W. Kennerley, Dahmubed, Lara, & Wallis, 2009), abstract value encoding schemes were observed in anterior cingulate cortex (ACC, including the subgenual cingulate which forms a component of vmPFC in humans). Neurons in this region rate coded value in a multiplexed fashion over three qualitatively distinct decision variables, namely reward probability, reward magnitude, and effort cost. In contrast, in more ventral potions of the prefrontal cortex (PFC), neurons tended to only encode the value of two of the three or just one of the three variables. In summary, these data indicated that stimulus-dependent signaling was present in OFC during the choice process, in concert with abstract value representations in more dorsal portions of medial PFC.

In humans, this hypothesis was confirmed in a task requiring subjects to make decisions regarding three different categories of items, namely food items, sums of money, and material goods (Vikram S. Chib, Antonio Rangel, Shinsuke Shimojo, & John P. O'Doherty, 2009). Using GLMs, it was found that BOLD signaling in an overlapping region of vmPFC correlated with the subjective value of items drawn from all three categories. However, there remained the possibility that the neural correlates of subjective value in vmPFC may have been based on stimulus-specific value representations in distributed activity patterns which were spatially smoothed as one of the standard pre-processing steps. Amongst other results, work described in Chapter 2 tested and rejected this possibility in an analogous experiment using multi-voxel pattern analysis.

Value Construction

A measure on this abstract uni-dimensional space of values is sometimes referred to as a “common currency” (P Read Montague & Berns, 2002). Assuming that stimuli x are composed of elemental attributes or features with stimulus-specific parameters x_i and weights v_i associated with each attribute, the integration hypothesis of subjective value (Rangel & Clithero, 2014) proposes a weighted L1 norm for stimulus value computation

$$V(x) := \sum_i v_i x_i$$

Qualitatively, this theory has some empirical support. Studies have found neural signals (and modulation of functional connectivity) corresponding to the valuation of a specific attribute, namely “healthiness” in dlPFC (T. Hare, Camerer, & Rangel, 2009). Also, the same authors found that this attribute could be exogenously manipulated via putative attentional mechanisms (T. A. Hare, J. Malmaud, & A. Rangel, 2011). Together, these studies indicate that stimulus features make dissociable contributions to the overall value of a stimulus, that these contributions may be computed in distinct brain regions, and that attention can modulate this contribution. Further evidence for regional differences in reward processing comes from an early neuroimaging meta-analysis (Kringelbach & Rolls, 2004) which found that neural signals for more abstract reinforcers (e.g., money) tended to be concentrated in more anterior portions of OFC as opposed to representations of primary rewards (e.g., food) in posterior OFC. Taken together, these results suggest the possibility that value representations which are dependent on the identity of a stimulus may be detectable in the human brain while subjects are evaluating a choice. In Chapter 2, I tested this hypothesis and indeed found stimulus-dependent value representations in OFC organized along a posterior/anterior gradient whereby value for food items were encoded more posteriorly while material goods such as DVDs and books were represented more anteriorly (McNamee, Rangel, & O'Doherty 2013). The use of MVPA was critical in this endeavor since these value representations were distributed across voxels rather than independently rate coded (Jimura & Poldrack, 2012).

Learning from Reinforcement

In order to control the world through our action, we must understand how it works. Decisions are made based on information acquired through experience which we use to (i) build internal models of the generative structure of the world, (ii) understand the effect our actions have on it, and (iii) develop policies, or action priorities in given situations. The simplest learning algorithms model how we estimate associations between distinct stimuli, a process which is referred to as Pavlovian conditioning in classical psychology (Pavlov, 1927). For example, if stimulus X (known as the conditioned stimulus) repeatedly precedes stimulus Y (the unconditioned stimulus), organisms learn to predict Y given the presence of X. In probabilistic terms, a representation of $P(Y|X)$ is acquired. Although humans and animals are capable of encoding probabilistic distributions, the equations of classical conditioning theory developed in the early half of the twentieth century focused on the relatively primitive representation of an association strength A between stimuli X and Y, that is updated for each experience:

$$A \leftarrow A + \alpha(I - A)$$

where I is an indicator variable for the presence of Y soon after or in conjunction with X. This incremental updating rule is the basic computational template for many conditioning phenomena. One notable modification is the Rescorla-Wagner rule which incorporates “blocking” or competition between stimuli for predictions,

$$A_j \leftarrow A_j + \alpha\left(I - \sum_i A_i\right)$$

where A_i represent the associability for a stimulus X_i. This rule implies that if the presence or occurrence of a stimulus is fully explained by alternative stimuli already, then no associability is assigned to stimulus X_j. This is equivalent to an important principle in Bayesian inference where the likelihood of a given event is scaled by the probability of that event based on the current model (see Appendix A for a detailed derivation).

Even such a basic ability to predict the world can mean the difference between life and death. Rustling bushes X could imply that the appearance of a lion Y is imminent to a gazelle. Compared to $P(Z|X)$ where Z is a harmless elephant, this example highlights the importance of the value of the predicted outcome from a decision-theoretic point of view. Thus, we would like to estimate the reward or loss associated with a stimulus or action (as in instrumental conditioning). This moves us from the domain of associative learning to reward learning. For example, one might compute a value estimate $\tilde{V}(a)$ from rewards R obtained by selecting a particular action a as the average reward obtained on previous performances of that action:

$$\tilde{V}(a) := \frac{R_1 + R_2 + \cdots + R_n}{n}$$

In this way, one can assign a subjective value to an action that inputs into the decision process. However, this averaged representation of reward based on a “batched” update is not consistent with real “online” interactions in the environment that occur continuously. This motivates the use of the aforementioned incremental updating algorithms for reward learning:

$$V_{t+1}(a) = V_t(a) + \alpha(R - V_t(a))$$

The relationship between a reward variable R and its estimated value is non-trivial, primarily because a reward may have different values depending on the context in which it is considered. For example, the value of food is dependent on an animal’s internal state of hunger. Another complication is that reward may not be directly contingent on an action but require multiple sequential actions to be performed. Such sequential decision-making environments are modeled as Markov Decision Processes (MDPs), which consist of the following 5-tuple (R. Sutton & Barto, 1998):

$$(S, \{A_s, s \in S\}, P, R, \gamma)$$

S denotes a set of states which could refer to parametric combinations of external factors and variables internal to an agent. A_s is the set of actions available in state $s \in S$, $P: S \times A \times S \rightarrow [0,1]$ is a transition probability function $P(s'|a, s)$, which describes the probability of arriving in state s' after taking action a in state s. R is a reward function where $R(s', a, s)$ describes the immediate reward acquired after the transition (s', a, s), where $\gamma \in [0,1]$ is a factor that exponentially discounts future rewards and thus weighs the relative importance of immediate and future rewards. Given that no agent in a naturalistic environment can be considered immortal, $\gamma < 1$ is a reasonable assumption and should be tuned to the expected time horizon of the environment. The goal of an agent in such an environment is to compute a control policy which maximizes expected future reward

$$V(t) = E[R_t + \gamma^1 R_{t+1} + \gamma^2 R_{t+2} + \cdots]$$

$$V(t) = E\left[\sum_i \gamma^i R_{t+i}\right]$$

where time is indexed by t and R implicitly depends on the selected state-action transitions. Within this framework, a decision trajectory through many transitions may forgo many unrewarded actions in favor of a large final reward. Learning such policies raises the temporal "credit assignment" problem of delayed reinforcement. If an agent has to select many actions, only the last of which is rewarded, how to does it assign positive values to the distal non-rewarded actions? A popular approach to this problem is temporal difference learning (TD-learning) (R. Sutton & Barto, 1998), which is described by the following equations in the Pavlovian case (where an agent passively receives rewards at times indexed by t):

$$\tilde{V}(t) \leftarrow \tilde{V}(t) + \alpha\delta_t$$

$$\delta_t = R_{t+1} + \gamma\tilde{V}(t+1) - \tilde{V}(t)$$

$\tilde{V}(t)$ is the algorithm's estimate of expected future reward, which is linearly updated by a prediction error δ_t scaled by a learning rate α. The prediction error δ_t is composed of the difference between the reward received plus the estimated future reward at time $t+1$ and the current estimate of $\tilde{V}(t)$. This local updating rule is possible due to the fact that the expected reward function satisfies Bellman's principle of optimality (Bellman, 1952). We can derive a new form of the expected reward function in order to expose Bellman's principle in this situation:

$$V(t) = E\left[\sum_{i=0\ldots} \gamma^i R_{t+i}\right]$$

$$V(t) = E\left[R_t + \sum_{i=1\ldots} \gamma^i R_{t+i}\right]$$

$$V(t) = E[R_t] + \gamma V(t+1)$$

TD-updating works by propagating reward prediction errors backwards through experienced trajectories. Initially, early states will be incorrect in their estimate of expected future reward while states proximal to the actual rewards will become more accurate. Over time, the estimates of expected rewards at future states will be used for reward prediction errors for earlier states and eventually they will be associated with accurate predictions of reward. Despite the simplicity of this approach, it has been successfully used to train an artificial agent to learn to play backgammon at an elite level (Tesauro, 1995), exhibiting its capacity for generating complex, apparently intelligent behaviors. Importantly, neural signals corresponding to the aforementioned prediction errors δ_t have been identified in dopaminergic neurons in the nervous system of several species (McClure, Berns, & Montague, 2003; Schultz, Dayan, & Montague, 1997), lending weight to the notion that temporal difference learning is being used in the brain.

Multiple Mechanisms

Based on classic early animal experiments we have described a reward-based associative account of reinforcement learning. However, there are many alternative algorithms which seek to learn a model of the MDP environment itself by estimating the reward and transitions functions directly (Engel, 2005). Based on this information, one can compute the expected future reward contingent on an action by brute force or sampling and thus produce a control policy. The collection of algorithms which take this approach are broadly referred to as “model-based” as opposed to the “model-free” algorithms such as TD-learning and Q-learning described previously. This distinction echoes the difference (Spence, 1950) between the development and use of “cognitive maps” in decision-making (Tolman, 1948) versus the reliance on conditioned stimulus-response behaviors (Pavlov, 1927). Since the early days of instrumental conditioning, evidence has accumulated that animals make decisions based on both of these systems depending on many environmental factors. A primary factor appears to be time, or more specifically, the reduction in uncertainty over time. The more experience an animal has with an action-outcome contingency, the more automated or habitual the selection of this action becomes. In contrast, in early learning periods, decision-making remains “goal-directed” in the sense that animals incorporate predictions regarding the expected outcome of an action during action evaluation. Two behavioral assays are used (Bernard W. Balleine & Anthony Dickinson, 1998) to distinguish between these modes of decision-making: (i) outcome devaluation and (ii) contingency degradation which test for sensitivity to rewards and transition probabilities, respectively. Neurobiologically, the neural substrates subserving these two methods of action selection have been localized to dorsolateral striatum and infralimbic or orbitofrontal cortex in rodents using a range of techniques from lesions to optogenetics (Jones et al., 2012). In humans, homologous regions such as the posterior putamen for habits (E. Tricomi, B. Balleine, & J. P. O'Doherty, 2009) and ventromedial prefrontal cortex for goal-directed action selection (Glascher, Daw, Dayan, & O'Doherty, 2010; Alan N. Hampton, Peter Bossaerts, & John P. O'Doherty, 2006; John P. O'Doherty, 2004) have been implicated. The essential distinction between these systems is the manner in which decision value is computed, whereby habits are based on cached or pre-computed values that can be retrieved in the manner of a look-up table while goal-directed decision values are generated online. A theoretic study mapped habits and goal-directed decisions onto model-free and model-based reinforcement learning systems, respectively (Nathaniel D. Daw, Yael Niv, & Peter Dayan, 2005).

Questions of Representation

Note that these two systems require different environment inputs in order to learn and control. For example, since the habit system is not outcome-sensitive, one expects that the neural system that generates a habitual action response would not receive or contain a representation of the predicted outcome(s). In contrast, the goal-directed system requires both action and outcome information in order to generate a decision. In Chapter 4, I describe an fMRI-MVPA study in which we decoded actions and outcomes at the time of the presentation of an initial stimulus (i.e., before an action is performed or an outcome received). We localized action (but not outcome) representations to posterior putamen in agreement with previous GLM-based analyses (Elizabeth Tricomi et al., 2009) and showed that dlPFC contained overlapping representations of actions and outcomes during the putative decision phase. We also show that the action decoding accuracy in putamen and dlPFC (but not other regions such as vmPFC) correlated with reaction times on a per-subject basis, thus providing evidence that these regions are causally involved in generating responses.

Hierarchical Reinforcement Learning[1]

Many RL algorithms fail “in the real world” due to a problem with dimensionality, in which there are too many states over which to integrate information to make decisions let alone learn (Barto & Mahadevan, 2003). It has been proposed that state-space structures be compressed in order to make calculations tractable. In particular, multiple actions (and their interceding states) might be concatenated into “meta-actions” or, more generally, “options” (R. S. Sutton, Precup, & Singh, 1999). Decision policies would be developed over these options rather than the individual actions, thus reducing the computational complexity of any policy-learning algorithm. An example of an option might be “go to the lab”, which would be composed of more basic actions “leave home”, “catch bus”, and “enter building”. Of

[1] The next three subsections are adapted with permission from (O’Doherty, Lee, & McNamee, 2014).

course, these basic actions can themselves be composed of even more elemental actions reflecting a nested hierarchy of action complexity. It is has been suggested that the brain might implement such a hierarchical scheme, with different levels of a hierarchy tasked with selecting actions at different levels of abstraction. (Botvinick, Niv, & Barto, 2009). The notion of a hierarchy in RL appeals to a long literature in cognitive neuroscience suggesting the existence of a cognitive hierarchy within prefrontal cortex, with certain brain systems sitting higher up in the hierarchy (possibly located more anteriorly within prefrontal cortex) and thereby exerting control over systems lower down in the hierarchy (Badre & D'Esposito, 2009; Koechlin, Ody, & Kouneiher, 2003). Consistent with the hierarchical RL notion, a recent neuroimaging study has found that neural activity in ACC and insula correlated with prediction errors based on "pseudo-rewards" (representing the completion of an elemental action forming part of a rewarding option) in a temporally extended, multi-step decision-making task (Ribas-Fernandes et al., 2011).

Bayesian Approaches to Reinforcement Learning

Another important trend in the literature has been to use Bayesian inference to learn about reward distributions, or any other task-related decision variable, instead of using an RL approach involving reward prediction errors (Behrens, Woolrich, Walton, & Rushworth, 2007; Friston et al., 2013; A. N. Hampton, P. Bossaerts, & J. P. O'Doherty, 2006; O'Reilly, Jbabdi, & Behrens, 2012). One advantage of the Bayesian approach is that this method provides a natural way to resolve the issue of how the rate at which a belief about the world is updated in the face of new information is set as a function of the environment (Yu & Dayan, 2003) . In particular, among other factors, the amount of volatility present in the environment (the extent to which reinforcement contingencies are subject to change), should influence the rate at which new information is incorporated into one's beliefs, and this can be modeled in a very straightforward way in a Bayesian framework (Behrens et al., 2007). Another advantage of Bayesian inference is that because these models encode representations of full probability distributions (or approximations thereof), a natural consequence is that it is easy to extract a measure of the degree of uncertainty (or, conversely, precision) one has in a particular belief. Such uncertainty or precision signals can be used not only to inform that rate at which one should update learning rates (see (Payzan-LeNestour & Bossaerts, 2011), but can also be used to inform decision-strategies such as when to explore or when to exploit a given decision option (i.e., one might want to explore an option about which one is maximally uncertainty) (Badre, Doll, Long, & Frank, 2012; Donoso, Collins, & Koechlin, 2014; Payzan-Lenestour & Bossaerts, 2012; Schwartenbeck, Fitzgerald, Dolan, & Friston, 2013). Supporting the relevance of a Bayesian framework for understanding the neural implementation of RL, uncertainty and precision signals have been reported in a number of brain structures, including the midbrain, amygdala, and prefrontal and parietal cortices (Payzan-LeNestour, Dunne, Bossaerts, & O'Doherty, 2013; Prevost, McCabe, Jessup, Bossaerts, & O'Doherty, 2011; Prevost, McNamee, Jessup, Bossaerts, & O'Doherty, 2013; Schwartenbeck, FitzGerald, Mathys, Dolan, & Friston, 2014; K. Wunderlich, U. R. Beierholm, P. Bossaerts, & J. P. O'Doherty, 2011). However, it is important to note that while Bayesian approaches have considerable appeal due to their elegance and appeal to optimality, it remains challenging to definitively ascertain whether or not the brain is literally implementing Bayesian inference, and it is often possible to capture many of the same features of Bayesian models, such as flexible adjustments of learning rate, or a representation of variance or uncertainty in a learned belief with non-Bayesian (Li, Schiller, Schoenbaum, Phelps, & Daw, 2011; John M. Pearce & Geoffrey Hall, 1980; Preuschoff & Bossaerts, 2007; K. Wunderlich et al., 2011) or hybrid Bayesian-RL approaches (Dearden, Friedman, & Andre).

Learning and Inference over State-Space

In many typical RL applications, the states and actions available in those states are defined from the outset, i.e., used as input into the algorithm and not considered further. However, perhaps the biggest single challenge for RL agents is how to determine the relevant states and actions in the first place: when faced with noisy sensory information from the world, how does the agent determine the relevant features that constitute a state, and then identify what are the relevant actions in that state? (Ahmad & Yu, 2013; Gershman & Niv, 2010; Kwok & Fox, 2004). This problem is essentially already being worked on by neuroscientists studying sensory perception and sensorimotor learning, as it depends on the capacity to segment and identify relevant objects and contexts and determine actions (Poggio & Ullman, 2013; Ridderinkhof, 2014; Shadlen & Kiani, 2013; Wolpert, Diedrichsen, & Flanagan, 2011). One approach to this problem in the neural RL community has involved setting up an experimental situation in which a given stimulus has multiple dimensional attributes (e.g., shape, color, motion). Inspired by earlier cognitive set-shifting tasks (Milner, 1963; Robbins, 1996), one of these dimensions is unbeknownst to the participant, selected to be "relevant" in terms of being associated with a reward, and the goal of the agent is to work out which attribute is relevant, as well as to work out which exemplar within an attribute (e.g., a green color vs a red color) is actually reinforced (R. C. Wilson & Y. Niv, 2011; K. Wunderlich et al., 2011). Bayesian inference or RL can then be used to establish the probability that a particular dimension is relevant, which can then be used to guide further learning about the value of individual exemplars within a dimension.

The ability to construct a simplified representation of the environment focused only on essential details reduces the complexity of the state-space encoding problem. One way to accomplish this is to represent states by their degree of similarity to other states, either via relational logic (Kumaran & Maguire, 2009), transition statistics (Schapiro, Rogers, Cordova, Turk-Browne, & Botvinick, 2013), or feature-based metrics (Konidaris, Scheidwasser, & Barto, 2012; Mnih et al., 2013). Furthermore, the construction of generalized state-space representations can speed up state-space learning considerably by avoiding the time cost of re-learning repeated environmental motifs (e.g., if I learn how to open my first door, I can generalize this to all doors).

Latent Learning

Part of the process of identifying the current state of an environment with respect to decision-relevant variables is to accurately estimate the state of unobservable variables based on observable, conditionally related signals using hierarchical inference. This problem becomes particularly challenging when multiple distinct states must be estimated over as opposed to a binary state-space where the likelihood of each possible state is anti-correlated (K. Wunderlich, U. Beierholm, P. Bossaerts, & J. P. O'Doherty, 2011; Yang & Shadlen, 2007). Due to cognitive limitations, it is possible that humans use a serial hypothesis testing strategy when performing Bayesian inference over many states in order to simplify the inference process. More specifically, a person might eliminate a possible state from their “belief space” if its posterior probability falls below a fixed threshold and continue to perform inference over a reduced state-space. We investigated whether such manipulations of internal models are performed in a computational fMRI study, as reported in Chapter 5.

Chapter 2

STIMULUS-DEPENDENCY OF VALUE ENCODING IN VMPFC[2]

In order to choose between manifestly distinct options it is suggested that the brain assigns values to goals using a common currency. While previous studies have reported activity in ventromedial prefrontal cortex (vmPFC) correlating with the value of different goal-stimuli, it remains unclear whether such goal-value representations are independent of the associated stimulus categories as required by a common currency. Using multi-voxel pattern analyses on fMRI data, a region of medial prefrontal cortex was found to contain a distributed goal-value code that is independent of stimulus category. More ventrally in vmPFC, spatially distinct areas of the medial orbitofrontal cortex were found to contain unique category-dependent distributed value codes for food and consumer items. These results implicate the medial prefrontal cortex in the implementation of a common currency and suggest a ventral versus dorsal topographical organization of value signals within vmPFC.

[2] Adapted with permission from (McNamee et al., 2013).

Introduction

There is a considerable body of research demonstrating value signals in the brain while participants engage in a variety of decision-making tasks, particularly within the medial orbitofrontal and adjacent medial prefrontal cortices, collectively known as the ventromedial prefrontal cortex or vmPFC (T. Hare, O'Doherty, Camerer, Schultz, & Rangel, 2008; Kable & Glimcher, 2007; Padoa-Schioppa & Assad, 2006; H. Plassmann, J. P. O'Doherty, & A. Rangel, 2007; Rushworth & Behrens, 2008; Tom, Fox, Trepel, & Poldrack, 2007; L. Tremblay & Schultz, 1999). In order to enable decisions to be made between stimuli with fundamentally different qualities, it has been suggested that the brain uses a "common currency" in which values are assigned to different stimuli on a common neural scale (P Read Montague & Berns, 2002; John P O'Doherty, 2007; Rangel, Camerer, & Montague, 2008). Consistent with this hypothesis, several fMRI studies have reported overlapping univariate value signals in vmPFC while human subjects evaluate different types of goods such as food, money, books, DVDs, clothes, and social rewards (Vikram S. Chib et al., 2009; FitzGerald, Seymour, & Dolan, 2009; D. J. Levy & Glimcher, 2011; Lin, Adolphs, & Rangel, 2011).

However, the finding of overlapping neural activations representing goal-value for distinct stimuli in a univariate manner does not provide sufficient evidence for the existence of a stimulus-independent goal-value code, as required by the common currency hypothesis. There remains the possibility that an area exhibiting scaling with goal-values in a similar manner across different categories could in fact be composed of distributed and distinct yet spatially overlapping goal-value codes for different item categories. The first aim of this study was to determine whether distributed value signals within the vmPFC are unique for each category of good, even if such value signals overlap spatially, or whether by contrast there exists a truly generic common value signal in which the value of each category of good is encoded using the same distributed code.

Even if there is a common currency to facilitate comparisons across goals of different types at choice time, it is also necessary to represent unique goal-specific value codes. This is because in order to compute the current incentive value of particular goals, the specific sensory properties of a goal-outcome must be integrated together with the organism's current motivational state. For instance, the goal-value of salted peanuts and a soda will differ markedly depending on whether an individual is salt-deprived or thirsty. Moreover, according to attribute integration theories of value computation, the summary value of a complex good is computed online by summing the value of component attributes of a good at the time of decision-making (Padoa-Schioppa, 2011; Rangel & Hare, 2010). This type of mechanism would also involve the encoding of a goal-value signal that depends on the sensory features of the goal-stimulus being valued as an intermediate step toward the computation of a generic value code. This motivates the second aim of this study: to test for distributed patterns of activity in which goal-values are encoded in a manner that is specific to particular categories of stimuli.

Results

To address these aims, we modified a previously deployed paradigm (Vikram S. Chib et al., 2009), in which we optimized the design for multi-voxel pattern analysis techniques (MVPA). MPVA has been successfully applied in many decision-making paradigms: economic value (Krajbich, Camerer, Ledyard, & Rangel, 2009), associative value (Kahnt, Heinzle, Park, & Haynes, 2010), reward modality (Clithero, Smith, Carter, & Huettel, 2011), value-based decisions (Alan N. Hampton & O'Doherty, 2007), and consumer choices (I. Levy, Lazzaro, Rutledge, & Glimcher, 2011; Tusche, Bode, & Haynes, 2010) have all been decoded from fMRI data. In this study, participants were scanned with fMRI while they reported their “willingness to pay” (a proxy measure of their stimulus valuation obtained via a Becker-DeGroot-Marschack auction process (Becker, DeGroot, & Marschak, 1964)) for three different classes of goods: food items, monetary lotteries, and non-comestible consumer items (Figure 1a). We trained a pattern classifier on distributed voxel activity to categorize stimuli at the time of decision-making as being either high or low in subjective value based on each participant’s ratings. Although each category was composed of different stimuli, many value-relevant features are common to all stimuli in each category and there is little to no overlap across categories. Thus we hypothesized that a classifier would be able to decode stimulus-independent value patterns across categories, while stimulus-dependent value representations should only be decodable within categories.

This motivated the following classifier training procedures: first, to test for the presence of category-independent value signals, we trained a classifier to decode value from samples drawn from one of the categories, and tested its performance in recognizing the value of exemplars drawn from a different category. Second, to test for category-dependent value codes, we trained the classifier on one stimulus category only, and determined if this classifier could decode the value of independent exemplars drawn from that same category but not exemplars from other categories. Third, we tested for regions representing stimulus identity (particularly the category from which the items were drawn) independent of stimulus value.

Notably, all multivariate analyses were performed on data in which the regularly observed univariate value signals had been removed (see Online Methods), thus ensuring that the MVPA could not classify based on this smoothed “global” activity. On account of prior findings in which stimulus value signals and other decision-making variables have been localized to vmPFC (Elliott, Dolan, & Frith, 2000; John P. O'Doherty, Kringelbach, Rolls, Hornak, & Andrews, 2001; Padoa-Schioppa & Assad, 2006; L. Tremblay & Schultz, 1999; Wallis, 2007, 2011), we focused our analysis on this area. To elucidate the spatial organization of various value coding strategies in vmPFC, we correlated voxel t-scores from the group-level multivariate value analysis with those from the univariate to determine how these qualitatively different value signals relate to each other. Moreover, we correlated the multivariate value voxel t-scores with voxel location to assess spatial variation in value signals across vmPFC. These correlation analyses suggest a topographic map of value signals within vmPFC with respect to stimulus-dependency and coding complexity (the distributed or univariate nature of the neural activity).

All reported value-related effects (both univariate and multivariate) are significant at a voxelwise FDR-adjusted threshold of $p < 0.05$ corrected for multiple comparisons within vmPFC. Effects which are unrelated to value representation are corrected across the whole brain at the same threshold (see Online Methods). A cluster extent threshold of 10 voxels was applied in all analyses. All conjunctions are performed using the “conjunction null” hypothesis (Nichols, Brett, Andersson, Wager, & Poline, 2005).

Univariate stimulus value signals.

To replicate previously reported univariate results (Vikram S. Chib et al., 2009) in which an overlapping area of vmPFC was found to correlate with the stimulus value of goods from all three categories, we performed the same univariate analysis (Vikram S. Chib et al., 2009) , testing for overlapping correlations with WTP for the goods from each category. Consistent with our previous findings, an area of vmPFC showed a significant effect in a conjunction contrast (peak $[x = 0, y = 35, z = -7], t = 3.14$, Figure 1b). We then searched for a brain region expressing univariate value uniquely for a particular class of items by examining linear contrasts between the WTP regressor parameter estimates for each category. No part of vmPFC showed a significant correlation between smoothed BOLD activity and WTP for only one of the categories (even at $p < 0.005$ uncorrected). In a whole-brain analysis, some activations were observed in parts of the visual and premotor cortices for the trinkets category only (only at $p < 0.005$ uncorrected); however, these clusters did not survive a corrected threshold and are thus not interpreted further. This lack of category-dependent univariate value coding is in agreement with previous results (Vikram S. Chib et al., 2009).

Distributed category-dependent stimulus value signals.

Our multivariate analyses showed that regions of the medial orbitofrontal cortex (mOFC) encode the value for food and trinkets in a category-dependent manner (Figure 2a). A posterior region of mOFC exhibited food-dependent value coding (peak $[x = -9, y = 17, z = -22], t = 3.05$), while a more anterior region of mOFC exhibited trinket-dependent value coding (peak $[x = -3, y = 41, z = -11]$, $t = 3.86$). We did not find evidence for a unique category-dependent value-coding region for monetary gambles in prefrontal cortex.

To replicate these results independently, we repeated our procedures on a previously acquired dataset (Vikram S. Chib et al., 2009), which used a similar task, but was not optimized for MVPA. This additional analysis revealed the same pattern of category-dependent stimulus value signals in mOFC, with an anterior locus encoding the value of trinkets and a posterior locus encoding the value of food goals (Supplementary Figure 2).

Spatial organization of the category-dependent value codes.

Figure 2b plots MVPA t-values for the within-category analyses for all voxels in mOFC as a function of the MNI y-coordinate. Taking this t-score as an indication of the strength of the distributed value representation, we found that the strength of food value representation declined ($\mathrm{r} = -0.52$) along the posterior-anterior axis, while the value representation strength of trinket items increased ($\mathrm{r} = 0.54$). Linear regressions of these voxel accuracy t-scores against their MNI y-coordinates, performed separately for each category, were highly significant according to parametric tests ($\mathrm{p} < 10^{-21}$). To control for correlation inflation caused by the spatial smoothing of the classification results, we ran a simulation analysis (see Online Methods). In this non-parametric test, no correlation drawn from the simulated null distribution exceeded the empirically observed correlations for either food or trinkets, thereby ruling out a spatial smoothing confound. These results therefore show an interaction between item category and the directionality of the encoding gradient.

We also performed an analogous test using a leave-one-participant-out procedure to alleviate concerns about the possibility of a non-independence bias contributing to this result. This supplementary approach also yielded a significant interaction between decoding accuracies for food and trinket values as a function of posterior vs. anterior location within mOFC (Supplementary Figure 3). A similar analysis in mPFC showed that this category-dependent encoding gradient was not present in mPFC, and thus is specific to mOFC.

Another potential concern is that our anterior/posterior gradient results are due to differences in generic properties (i.e., independent of the category definitions) of the goal-stimuli across categories such as, for instance, the familiarity of the stimuli or their availability to the participant. To address this, we obtained behavioral ratings for the stimuli from a subset of the original participants (8 out of 13) on five attribute scales (“valence”, “intensity”, “liking”, “access”, and “familiarity”), and tested for a difference in average ratings between the food and trinket stimulus categories. At the group-level there was no significant difference with respect to any attribute ($\mathrm{p} > 0.05$). There were few significant differences in some of these attributes at the level individual subjects, and none of those differences were consistent across individuals (Supplementary Figure 4, Supplementary Table 2, Supplementary Table 3, Supplementary Table 4). This analysis therefore suggests that potential generic stimulus attribute differences do not explain the anterior/posterior gradient results.

Category-independent value signals.

In contrast to the category-dependent value representation results in mOFC, a more dorsal region of vmPFC (overlapping with that from the univariate analysis, Figure 3a) was found to contain category-independent value signals. A classifier trained in this area using data from one item category was able to predict the value class (high vs. low) in either of the other stimulus categories as well as in its own category. At $\mathrm{p} < 0.05$ SVFDR, all six cross-category training/testing combinations were significant in a conjunction test (peak $[\mathrm{x} = -3, \mathrm{y} = 41, \mathrm{z} = 3], \mathrm{t} = 2.40$).

A potential confound is the fact that for zero bids (which make up a large proportion of the low value items), no motor response had be performed, while high value items always require a button press. Thus the neural processes involved in generating the motor response may be contributing to the significant category-independent value classification signals in vmPFC. To account for this possibility, we performed a category-independent value searchlight analysis with zero bid trials omitted and tested whether there was a significant classification signal within a 20mm radius sphere surrounding the peak coordinates of the category-independent value signal identified previously. We again found evidence for a category-independent value signal, albeit at an uncorrected threshold due to the fewer number of trials and smaller value variance (peak $[\mathrm{x} = 9, \mathrm{y} = 53, \mathrm{z} = 7], \mathrm{t} = 1.94$). We also tested whether there were any clusters within the regions identified as representing value in a category-dependent manner (using small volumes around the relevant coordinates), and found no significant classification signals. Thus, only the dorsal portion of vmPFC represented value in a category-independent manner regardless of whether zero bid trials were included. In addition, we replicated the category-independent result in mPFC in the independent dataset from Chib et al. (Vikram S. Chib et al., 2009) (Supplementary Figure 2a). This provides further evidence against a possible motor confound since in that paradigm a motor response was required on all trials, yet the same result was found.

Another issue is that the information on the bid feedback screen (Figure 1a) is correlated with our measurement of goal value, and thus activity could be driven by a signal elicited by the bid feedback as opposed to the goal-value per se. Evidence that this effect does not explain our results also comes from the additional analysis of the Chib et al. dataset (Vikram S. Chib et al., 2009). In that task, no feedback was given to the subjects at the end of each trial, yet we still find evidence of a category-independent value code in mPFC.

Comparison between univariate and multivariate value codes.

Our finding of both univariate and multivariate value signals within vmPFC raises the question of how these different value encoding mechanisms relate to each other. It is possible that a set of voxels might encode both a univariate code and a multivariate code simultaneously. Alternatively, a set of voxels might exclusively encode a univariate value signal but no multivariate value signal, or vice versa. In order to establish whether value signals within the vmPFC are either uniquely multivariate or univariate, or show multiplexed univariate and multivariate value coding, we computed correlations between multivariate decoding accuracy and univariate signal strengths separately for our two main areas of interest: the mOFC and the mPFC. A multiplexed signaling strategy would manifest as a relatively high correlation between multivariate decoding accuracy and univariate signal strength. Alternatively, a low correlation implies that either a univariate or multivariate signal is present but not both. These distinct possibilities have implications for the computational nature of value encoding processes occurring in a given region.

On the basis of the findings for category-dependent multivariate value codes in mOFC, and univariate value signals more dorsally in mPFC, we hypothesized that the complexity of value coding in vmPFC might follow a ventral/dorsal gradient such that value codes distributed along the orbital surface tend not to contain any univariate encoding, but that as one moves superiorly up the medial wall, value codes could come to increasingly reflect a univariate code in conjunction with multivariate signals, while at the same time shedding category-dependency in the value code.

This hypothesis makes several predictions: (1) univariate value coding strength should increase along the z-axis while multi-voxel encoding should be more evenly divided between mOFC and mPFC, (2) the univariate signal should be relatively stronger than the multivariate signal in mPFC on average across voxels, and (3) univariate and multivariate coding should be more highly correlated dorsally in mPFC (such that both of these encoding strategies are present in the same voxels). We investigated this coding gradient hypothesis by testing each of these predictions in analyses which compare the univariate and within-category multivariate value coding results within mOFC and mPFC (Figure 3): (1) we correlated second-level voxel t-scores against voxel MNI z-coordinates across vmPFC (that is, mOFC and mPFC together) for the univariate and multivariate signals separately and examined whether these correlations were significantly different, (2) paired t-tests on a per-voxel basis were used to study how the relative strengths of these encoding strategies change across vmPFC, and (3) a correlation study was implemented to investigate whether the predictive relationship between univariate and multivariate signaling is different in these two subregions.

To implement the first test, we correlated the multivariate and univariate value t-scores from the second-level analyses with the z-axis coordinate of the associated voxel (Figure 4a). This was done for the food (univariate $r = 0.89$, distributed $r = 0.4$) and trinket stimulus categories (univariate $r = 0.72$, distributed $r = 0.68$). For each combination of item class and coding strategy, value signal strength increased along a ventral-dorsal gradient ($p < 0.05$, in both parametric and non-parametric tests). By bootstrapping the empirically observed results, we estimated sampling distributions for these correlation strengths. Non-parametric confidence bounds on the correlation strengths were established and indicated that although the strength of both signals increased along an ventral-dorsal gradient, univariate coding increased significantly more steeply ($p < 0.05$). In addition, we implemented a similar analysis investigating differences between peak value representation scores in ventral and dorsal regions of vmPFC for multivariate and univariate encoding strategies while utilizing data on a leave-one-participant-out basis. This analysis again confirmed these results (Supplementary Figure 5).

Our second test examined the relative prevalence of univariate and multivariate codes in these regions. We found a significant difference (paired t-tests, $p < 0.05$) in the relative strengths of the multivariate and univariate value signals between mOFC and mPFC for both the food and trinkets categories by comparing t-scores on a per-voxel basis. An important caveat here is that univariate and multivariate analyses have different intrinsic sensitivities (Jimura & Poldrack, 2012), thereby complicating the interpretation of absolute differences. However, this result does show that the multivariate signal is relatively stronger in mOFC as compared to mPFC (Figure 4b).

The third test aimed to determine how the univariate and distributed codes interact in mOFC and mPFC. The second-level t-scores from the univariate and multivariate analyses were correlated on a per-voxel basis in each of these two subregions separately. This revealed a strong difference between the subregions, whereby the univariate and multivariate t-scores are significantly more correlated in mPFC for both food ($r = 0.24$) and trinkets ($r = 0.61$) than in mOFC (food $r = -0.28$, trinkets $r = 0.34$, Figure 3b). This indicates that the distributed goal-value signals found in mOFC are largely independent from univariate goal-value codes, whereas this is not the case in mPFC.

Distributed coding of stimulus category.

Finally we looked for regions showing distributed coding of stimulus category, independent of its value. We found category discriminating activity in several areas of the brain (Figure

5). Areas in the frontal lobe included medial PFC (peak $[x = -3, y = 20, z = -22]$, $t = 6.12$), central OFC (peak $[x = -21, y = 38, z = -11]$, $t = 11.14$), dorsolateral PFC (right hemisphere peak $[x = 45, y = 32, z = 21]$, $t = 5.84$, left hemisphere peak $[x = -60, y = 17, z = 14]$, $t = 11.34$), and frontopolar cortex (peak $[x = 6, y = 65, z = -11]$, $t = 6.89$). Fusiform (peak $[x = 24, y = -43, z = -29]$, $t = 6.90$), parahippocampal (peak $[x = 36, y = -10, z = -33]$, $t = 6.56$), and inferior temporal gyri (right hemisphere peak $[x = 30, y = -73, z = -15]$, $t = 7.36$, left hemisphere peak $[x = -45, y = -64, z = -22]$, $t = 7.64$) were observed in the temporal lobes. More posteriorly, the intraparietal sulcus (right hemisphere peak $[x = 33, y = -70, z = 42]$, $t = 7.94$, left hemisphere peak $[x = -48, y = -31, z = 42]$, $t = 11.65$), precuneus (peak $[x = -6, y = -64, z = 14]$, $t = 5.54$), posterior cingulate cortex (peak $[x = 3, y = -43, z = 42]$, $t= 7.43$), and visual cortex (peak $[x = 9, y = -79, z = 32]$, $t = 11.34$) were implicated.

Discussion

It has been argued that to make decisions involving different types of goods the brain needs to encode item values on a comparable scale, often called a “common currency” (Alan N. Hampton & O'Doherty, 2007; P Read Montague & Berns, 2002; Rangel et al., 2008). While a number of studies have found that BOLD responses in an overlapping area of vmPFC correlate with the value of stimuli at the time of making decisions (Vikram S. Chib et al., 2009; FitzGerald et al., 2009; T. A. Hare, Camerer, Knoepfle, & Rangel, 2010; Hackjin Kim, Shimojo, & O'Doherty, 2011; D. J. Levy & Glimcher, 2011), there are many open questions regarding the nature of the code used in these computations. In particular, previous tests cannot rule out the possibility that the results were generated by category-dependent (e.g., foods vs. social vs. objects) value codes that are implemented in distinct yet spatially intermingled networks, and which are inconsistent with the common currency hypothesis. In addition, previous studies have been unable to find a spatial topography in the organization of goal-value signals in vmPFC.

Here, by using a paradigm optimized for multivariate analyses, we found evidence for the existence of both category-dependent value signals (which only reflect the value of particular types of stimuli), and category-independent value signals (which reflect the value of all stimuli, regardless of their category). The category-independent value signals were located in a region of vmPFC along the medial wall but above the orbital surface, and coincided substantially with the areas found in previous univariate analyses (Vikram S. Chib et al., 2009), as well as in a univariate analysis of the present dataset. Our results provide evidence, up to the fidelity provided by multi-voxel fMRI (Formisano & Kriegeskorte, 2012; Misaki, Kim, Bandettini, & Kriegeskorte, 2010), for the existence of a truly generic value code in the mPFC in which goal-values are represented independently of the category from which the stimuli are drawn. They also point towards a ventral-dorsal gradient within the vmPFC, as one transitions from the category-dependent value regions of the orbital surface to the more dorsal category-independent regions of mPFC. This suggestion is consistent with the fact that in many fMRI studies which have identified value representations in vmPFC for different classes of reinforcers using standard univariate techniques, decision-value and goal-value signals tend to appear superior to the orbital surface (Vikram S. Chib et al., 2009; T. Hare et al., 2008; Hilke Plassmann et al., 2007). In contrast, two distinct voxel clusters in mOFC were found at the group level to encode category-dependent goal-values for food and trinkets; a more posterior region was found to contain food-dependent value signals, while a more anterior region of mOFC was found to encode a trinket-dependent value signal.

Furthermore, a correlational analysis of the classifier's local sensitivity vs. spatial location revealed an anterior-posterior gradient in the mOFC, with category-dependent values of increased abstractness (i.e., trinkets) encoded more anteriorly. Although a similar effect could be caused by two separate food and trinkets peaks with Gaussian noise, visual inspection of the t-score plots and the strength of the linear dependence suggest an actual gradient effect in the nature of the value code. These findings resonate with the results of a meta-analysis (Kringelbach & Rolls, 2004) in which an anterior vs. posterior gradient was reported within mOFC in response to reward outcomes according to the "complexity" or degree of abstractness of the reinforcer. A previous univariate fMRI study reported dissociated posterior and anterior clusters of activation within OFC for reward expectation representations for sexual vs. money reinforcers (Sescousse, Redouté, & Dreher, 2010), though this effect was located more laterally where we observed stronger distributed encoding of stimulus category rather than stimulus value. However, unlike these studies, the results of the present study correspond specifically to goal-value representations where values are used as an input to the choice process as opposed to pure expectancy signals or the value computed at the time of the consumption experience (often called outcome value). These results support the proposal that there is indeed a gradient within mOFC whereby value signals corresponding to the processing of more biologically basic stimulus attributes, such as food or sexual stimuli, are encoded more posteriorly, while the value of more abstract stimulus attributes are encoded more anteriorly.

The findings obtained here implicating vmPFC in the encoding of a common currency for goal-values are consistent with evidence from lesion studies in both human and non-human primates implicating this region in value-based decision-making (Bechara et al., 1994; Fellows & Farah, 2005; Walton, Behrens, Buckley, Rudebeck, & Rushworth, 2010). The present results suggest that a lesion to the vmPFC would alter or disrupt the encoding of goal-values that are in turn used to guide behavior, thereby resulting in a decision-making impairment. In particular, an implication of the present findings is that a selective lesion to either anterior or posterior mOFC might result in a very specific impairment at decision-making over only certain classes of goods. While it is unlikely that lesions studied in human patients would ever have the anatomical specificity to enable such a possibility to be tested, this is something that could be potentially tested in an animal model.

It is notable that we did not find evidence for a category-dependent value code for monetary gambles, while both a univariate value signal for these gambles and category-independent value signals (training or testing on neural samples from the money category) were robustly encoded more dorsally within the mPFC areas involved in implementing category-independent value codes. One possible interpretation of this result is that because money is by definition a generalized reinforcer that has acquired value by virtue of its exchangeability for other reinforcers, money might only be represented according to a generic (category-independent) as opposed to a category-dependent value code. Furthermore, money is not tied to a specific sensory modality, and is therefore not dependent on specific sensory coding mechanisms (such as taste, olfaction, or vision). Moreover, within the attribute integration account of valuation, given that money does not have any component attributes, it could be argued that money cannot be encoded in a category-dependent manner. Another more mundane possibility is that, unlike items drawn from the food and trinkets categories, the actual values of the monetary sums are presented explicitly and do not require a complex stimulus-to-value transformation as would be the case if, for example, piles of coins had been displayed whose composition and size were indicative of monetary value.

Beyond goal-value signals, we also found evidence for value independent category identity codes within a region of central OFC, but also extending more medially to overlap with some of the value coding areas. These findings suggest the existence within parts of OFC of stimulus-identity codes. Perhaps unsurprisingly, such stimulus-identity coding was also found to be widespread in other parts of the brain outside of the OFC, including dorsal frontal cortex, parietal cortex, and visual cortical areas. Many of these areas were previously implicated in a EEG study of the time course of value computation (Harris, Adolphs, Camerer, & Rangel, 2011). Nevertheless, the presence of such signals within the OFC provides insight into the possible mechanisms by which value codes might get computed within the vmPFC during the choice process. In order to compute a category-dependent value code, it is clearly necessary to first have access to information about the identity of the stimulus, so that the incentive value of the goal state can be retrieved with respect to prior associations between the identity of the goal state and motivational states acquired through incentive learning (Bernard W. Balleine & Anthony Dickinson, 1998). Such goal-value codes are also likely necessary in order to facilitate choices over goals to be computed, because when comparing between the values of different goods, it is also necessary to be able to bind the results of the comparison process with the identity of the specific goods in question. Furthermore, according to the attribute integration view of value computations, it is necessary to encode information about various attributes associated with each stimulus in order to pass such information to the areas involved in category-dependent valuation. Further work will need to be performed to determine how these distinct value and identity representations within vmPFC get integrated and used during the decision process.

The loci of the value coding and category coding results in vmPFC can be interpreted in terms of the neuroanatomical structure of the brain. Based on cytoarchitectonic heterogeneities in OFC (Mackey & Petrides, 2010; Ongur & Price, 2000), a broad distinction has been made between a lateral prefrontal network (Brodmann areas 11, 13, and 47/12) covering central and lateral OFC and a medial prefrontal network (Brodmann areas 11m, 13 medially, along with 14 and extending up the medial wall to areas 10m, 24, 25, and 32) corresponding to ventromedial prefrontal cortex. Recently, a resting-state connectivity study (Kahnt, Chang, Park, Heinzle, & Haynes, 2012) has provided functional evidence in support of this parcellation scheme in human OFC. The sensory efferents of central OFC and the visceromotor afferents of the medial network (Croxson et al., 2005; Ongur & Price, 2000) suggest that the sensory-visceromotor pathway from central OFC to mOFC to mPFC could support a high-level stimulus-to-value transformation during decision-making. In this study, we found that central OFC coded stimulus category bilaterally, with these areas partially overlapping value-coding regions in vmPFC. This part of OFC has been shown to receive sensory input in all sensory modalities (both unimodal and multimodal), association cortices, and memory-related regions, and in particular is connected to several of the posterior regions which were found to encode stimulus category in a distributed manner. Moreover, adjacent to this central OFC result, category-dependent value signals were located in medial OFC, which has strong reciprocal connections to limbic areas involved in the emotional and hedonic processing of stimuli along with other parts of prefrontal cortex, which may contribute to an evaluation of the stimulus in the context of the current internal state of the subject and external state of the world (Louie & Glimcher, 2012). These effects could include inhibiting desires to consume food (T. Hare et al., 2009), or retrieving goal-related episodic memories (Duarte, Henson, Knight, Emery, & Graham, 2010) such as remembering whether or not a book has been read or not. Finally, these attribute-dependent value signals would be passed to the more dorsal areas of mPFC involved in category-independent value representations where a summary goal value is transmitted to action control circuits via mPFC (T. Hare, J. Malmaud, & A. Rangel, 2011; John P. O'Doherty, 2011; Rangel & Hare, 2010). Further support for this value processing pathway model can be found in a recent electrophysiological study (Cai & Padoa-Schioppa, 2012) in monkeys which found that neurons in anterior cingulate cortex (ACC) encoded the value of a chosen outcome only after a decision had been made and in particular after the same variables had been signaled by, presumably upstream, neurons in OFC. In addition, neurons in the dorsal bank of ACC but not the ventral bank were sensitive to the action required to make the choice.

It is important to note that while the present conclusions are supported by the particular set of stimuli we have used (consumer items, food, and money gambles), we cannot rule out the possibility that if we had used an entirely different class of goods (such as luxury goods, or social stimuli, etc.), the results could have turned out differently. Future studies will need to further establish the generality of the common coding area in the more dorsal part of vmPFC identified here, as well as establish whether other classes of items are coded uniquely within the medial orbital surface.

Finally, the importance of the multivariate methodology used in this paper is worth highlighting. As described above, a large number of previous studies have found that neural activity in an overlapping area of vmPFC, which encompasses the area where we have found category-independent goal-value signals, correlates with the value of a wide class of stimuli and stimuli at the time of choice. However, none of these previous univariate studies found the category-dependent value codes identified here. The reason might be due to the nature of the category-dependent signals. If, as conjectured above, they reflect the computation of value for stimulus specific attributes, then the category-dependent value signals are likely to be distributed across multiple voxels, which makes them difficult to localize using univariate approaches.

Figures

Figure 1. Task, univariate value signals, and behavioral results.

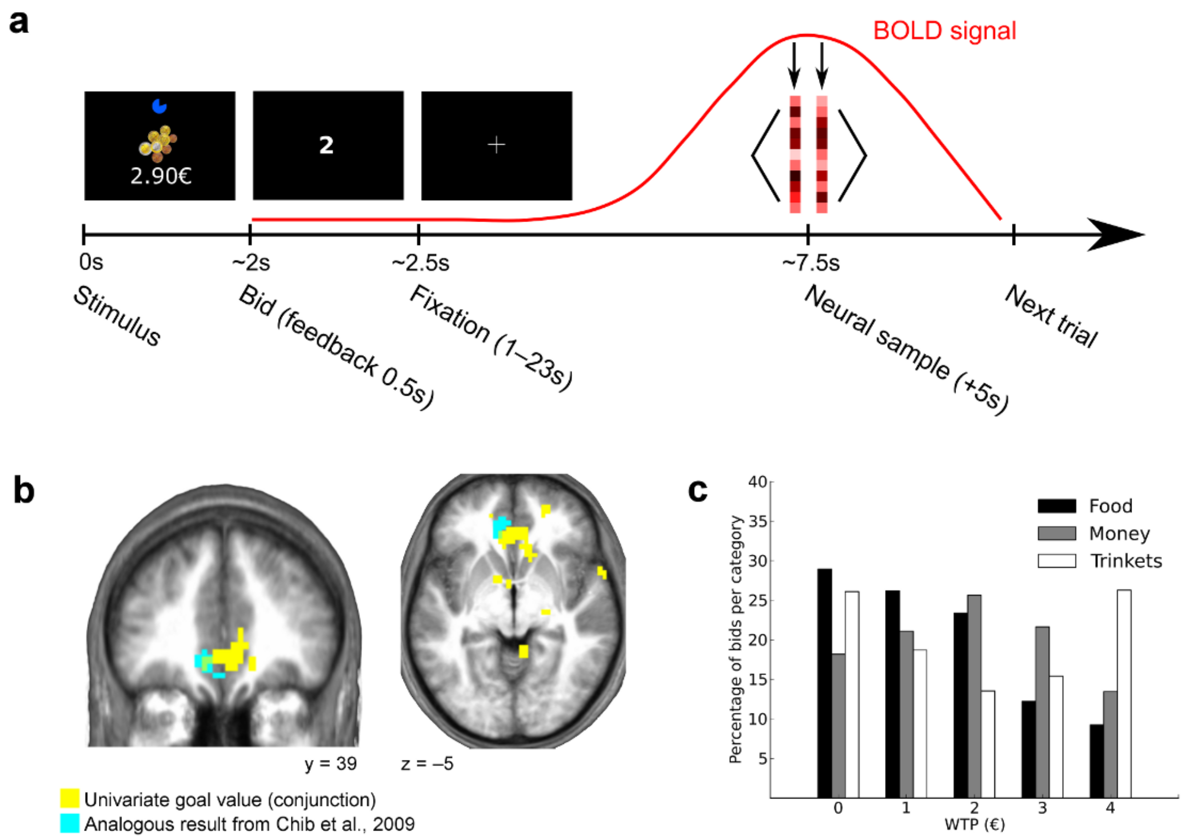


Figure 1. a.) Illustration of experiment time course and data extraction. Subjects were presented with an 80% chance of obtaining a stimulus drawn from a pool of 120 stimuli evenly divided into three categories (food, money, and “trinkets”) and they responded with an integer willingness-to-pay value between zero and four euros inclusive (see Online Methods). In preparation for the multivariate analyses, we extracted a sample of neural data at the bid time-point on each trial (with a shift of five seconds to account for haemodynamic delay). For a given bid, the two volumes closest in time (one before, one after) to the shifted time-point were averaged to create a single sample (Clithero et al., 2011). **b.**) A region of

vmPFC, overlapping with a previous similar result (Vikram S. Chib et al., 2009), was found to be parametrically modulated by the chosen bid value at the time of decision, peak coordinates $[x = 0, y = 35, z = -7], t = 3.14$, $p < 0.05$ SVFDR (results presented at $p < 0.005$ uncorrected in figure). **c.**) Distribution of WTP bids across all subjects for each category of items. The distribution of bids is shown in c, and is similar to those obtained previously (Vikram S. Chib et al., 2009). The average bid was €1.47 (SD, €1.28) for food items, €1.91 (SD, €1.3) for monetary sums, and €1.97 (SD, €1.56) for trinkets. There was a difference between the mean bids of the three categories (ANOVA, $p < 0.001$). The average bids were significantly greater than zero for all three classes ($p < 0.001$). The majority of bids were non-zero (71% for food, 82% for money, 74% for trinkets).

Figure 2. Distributed category-dependent value codes in mOFC for food and trinkets.

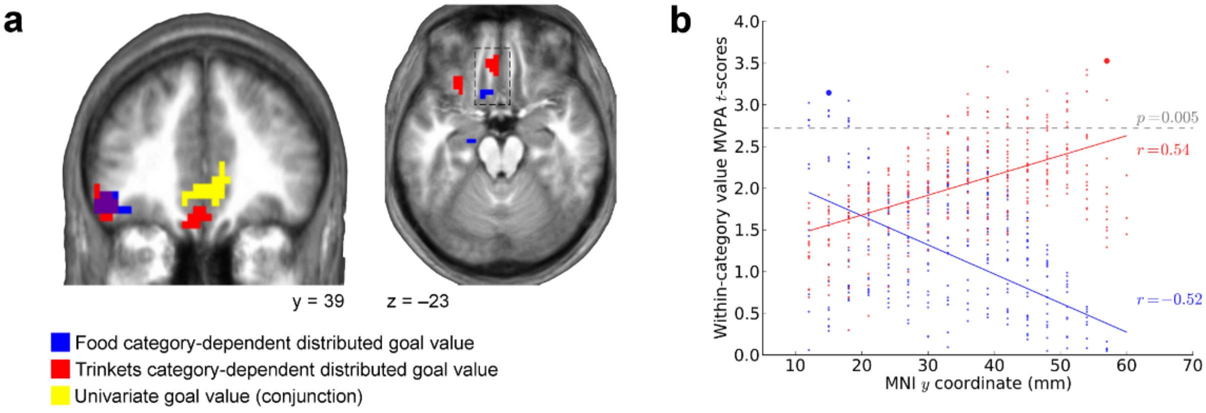


Figure 2. a.) Stimulus value was found to be represented in distributed codes in mOFC for the food (blue) and trinkets (red) categories. The peak classification accuracy t-scores are at the following coordinates; food $[x = -9, y = 17, z = -22], t = 3.05$; trinkets $[x = -3, y = 41, z = -11], t = 3.86$, p<0.005 SVFDR (results presented at $p < 0.005$ uncorrected in figure). No evidence for a multi-voxel money category value representation was found. **b.**) Food and trinket category-dependent value encoding regions in mOFC are organized along an anterior to posterior axis across subjects, with the most significant voxels for food values located significantly more posteriorly within mOFC, while the most significant voxels for trinket values are located significantly more anteriorly, as shown in a correlation analysis of second-level voxel t-scores vs. y-axis location. The large dots indicate peak t-scores. No such effect was found more dorsally in mPFC.

Figure 3. Organization of univariate and distributed value signals in vmPFC distinguished by coding mechanism and stimulus information content.

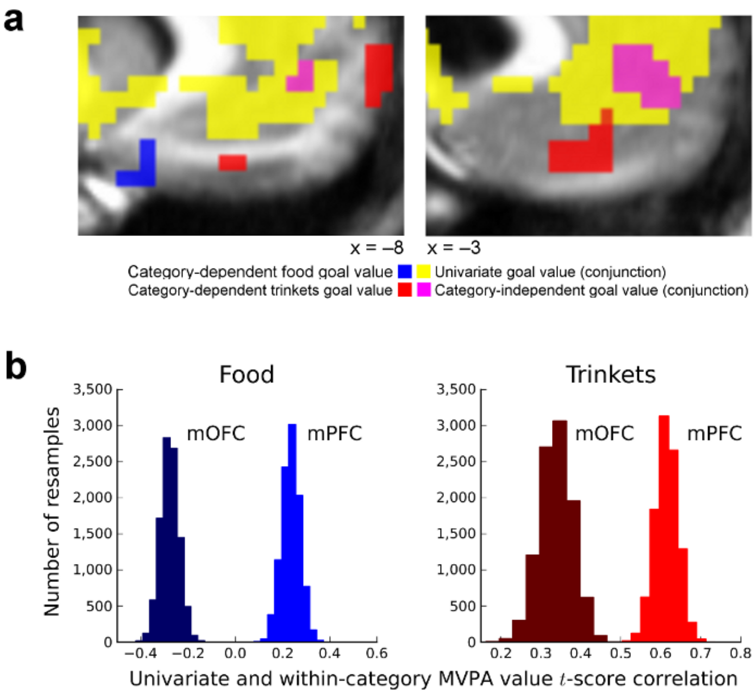


Figure 3. a.) A sagittal view of vmPFC shows that univariate and multivariate category-independent value representations are concentrated in mPFC while category-dependent value signals (for the food and trinkets categories) are located more ventrally in OFC. The peak of the category-independent value decoding conjunction was found at $[x = -3, y = 41, z = 3], t = 2.40, p < 0.05\ SVFDR$ (results presented at p<0.005 uncorrected in figure). **b.**) Histograms reflect bootstrapping results for univariate/multivariate value correlation analyses performed for each combination of category and vmPFC subregion. Correlations were significantly stronger in mPFC compared to mOFC for both food and trinkets. For food, the univariate and multivariate value t-scores were significantly anti-correlated.

Figure 4. Comparisons of univariate and multivariate value signal strengths across vmPFC subregions.

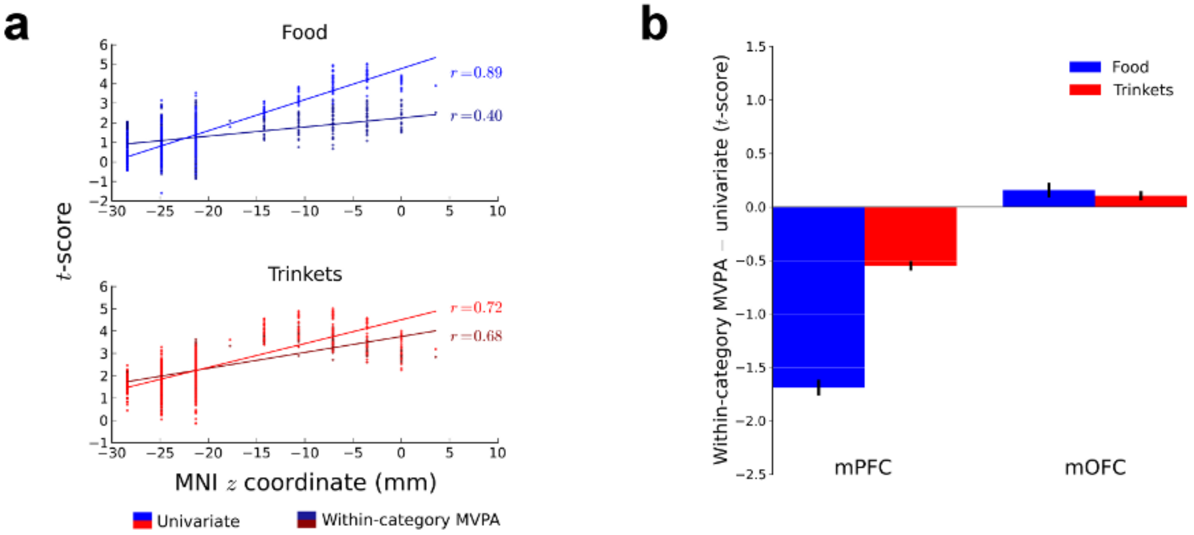


Figure 4. a.) For the food and trinket categories, univariate (brighter colors) and within-category MVPA (darker colors) second-level voxel t-scores are plotted as a function of the voxel's z-coordinate. This plot shows that the t-scores in the univariate brain maps show a significantly greater tendency to increase along the z-axis ($p < 0.05$). **b.**) Bars indicate the difference between the within-category MVPA and univariate value t-scores across voxels for the food and trinkets item categories within mPFC and mOFC. Error bars indicate standard error of the mean (+/- SEM).

Figure 5. Stimulus category coding.

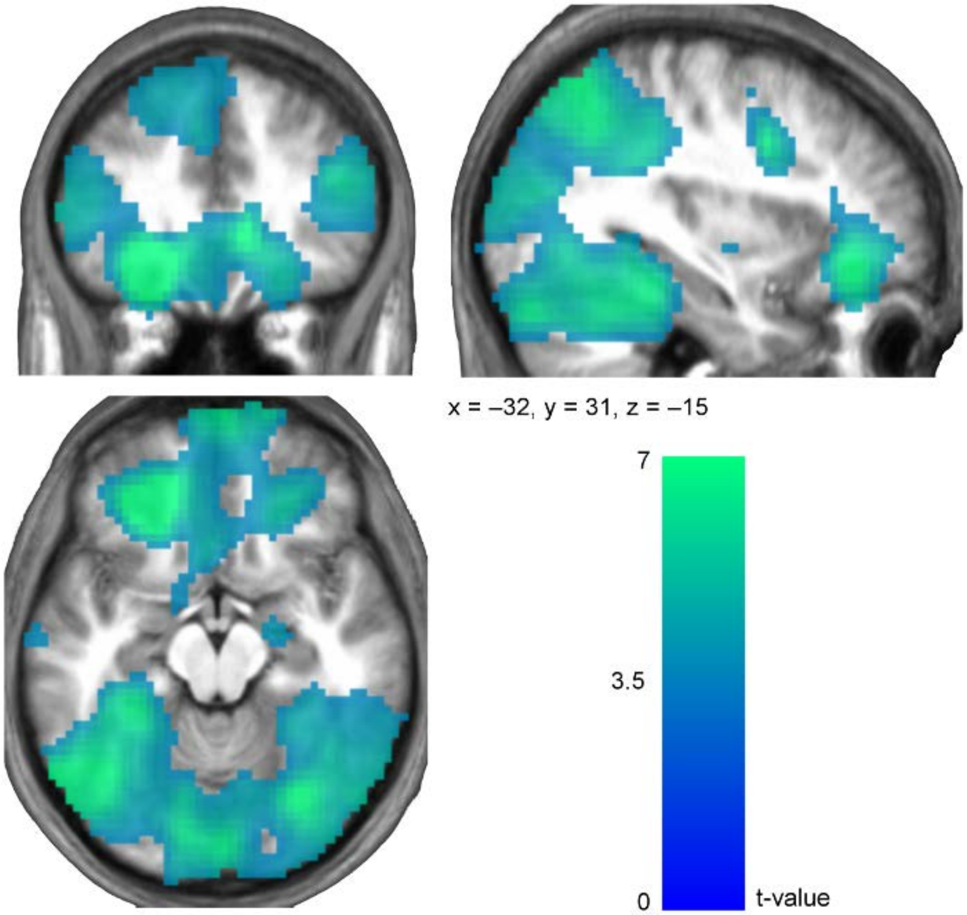
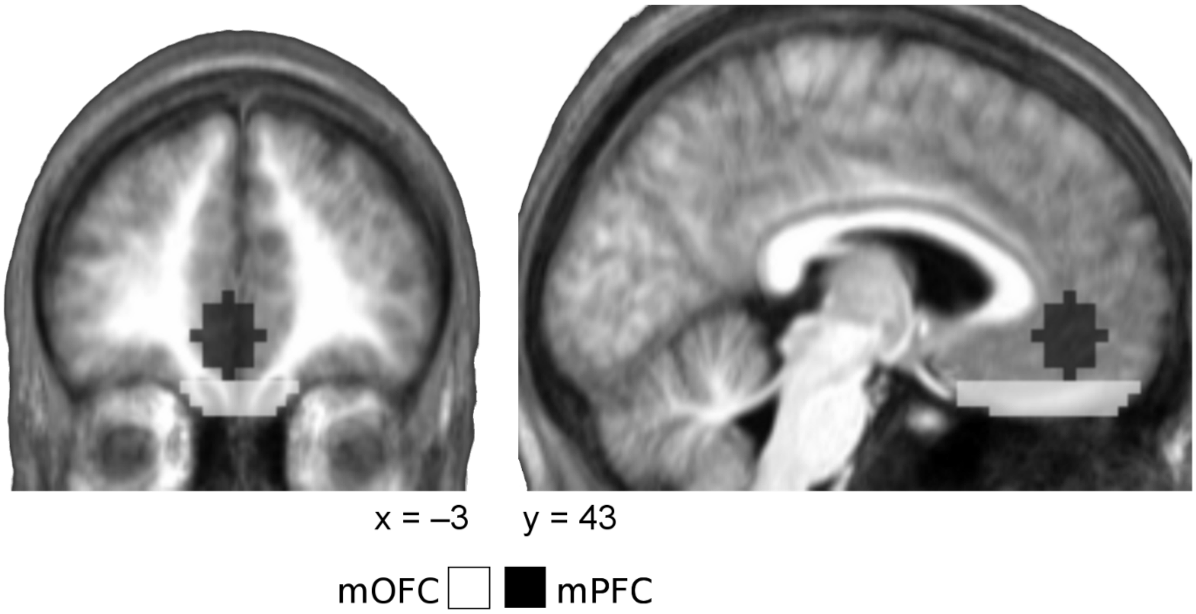


Figure 5. In the frontal lobe, central OFC (peak $[x = -21, y = 38, z = -11]$, $t = 11.14$), mPFC (peak $[x = -3, y = 20, z = -22]$, $Z = 6.12$), mFPC (peak $[x = 6, y = 65, z = -11]$, $t = 6.89$), and dlPFC (peak $[x = -60, y = 17, z = 14]$, $t = 11.34$) contain distributed neural patterns pertaining to the identity of the stimulus under consideration. More posteriorly, regions of the temporal lobes including the fusiform, inferior temporal, and parahippocampal gyri and areas around the intraparietal sulci also reflect category discriminating activity (see Supplementary Table 1). Results presented at $p < 0.005\ FDR$.

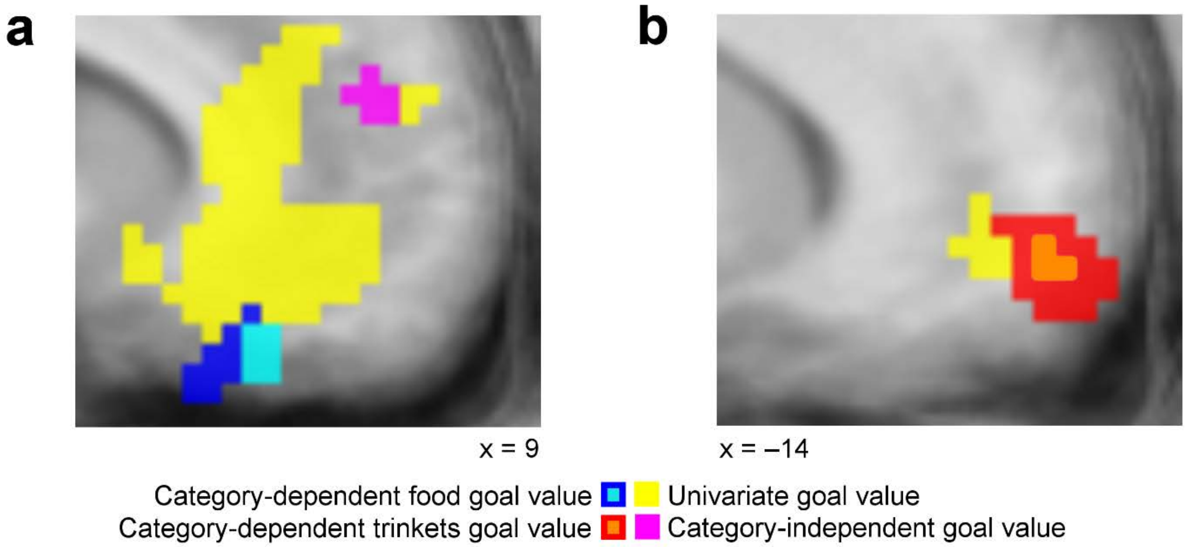
Supplementary Figures

Supplementary Figure 1. Masks covering distinct subregions of vmPFC.



Supplementary Figure 1. Based on previous functional and anatomical results, our *a priori* hypothesis was that distributed and univariate value encoding signals would be found in vmPFC extending from the orbital surface to more dorsal regions up to and including parts of Brodmann areas 10 and 32. Due to the similarity of the experimental design, we used univariate peak coordinates from a related study (Chib et al., 2009) to construct a medial prefrontal cortex (mPFC) mask as a sphere with a radius of 9mm surrounding these peak coordinates (corresponding to the size of the multivariate searchlight sphere). A similar functional mask did not exist for the medial orbitofrontal cortex (mOFC), most likely due to the distributed nature of the value codes found there and the relative scarcity of MVPA studies in value-based decision-making, and thus the mOFC mask was constructed according to anatomical descriptions used previously in the literature (Beckmann et al., 2009). This mask encompassed the medial orbital and olfactory sulci bilaterally with the anterior and posterior limits defined by the extents of these sulci. The vmPFC mask was defined as a union of these two masks.

Supplementary Figure 2. Independent replication of main results.


a
b
x = 9
x = –14
Category-dependent food goal value
Univariate goal value
Category-dependent trinkets goal value
Category-independent goal value

Supplementary Figure 2. For an independent replication of our results, we applied our analysis procedures to the data acquired for a previous study (Chib et al., 2009) with a similar task paradigm but with some important differences. This study also used a BDM auction process to elicit the participants' willingness-to-pay (WTP) on an integer scale from $0 to $3 for a variety of items drawn from three categories (food, monetary sums, and “trinkets”). However, the WTP bids (that is, the goal values) for all the items were recorded before the participant entered the scanner. Subsequently, on each trial in the scanner, participants were required to make binary choices between an item and a fixed reference sum of money (the median bid over all items placed during the pre-scan behavioral experiment). The motor response performed was a left or right button press and was completely uncorrelated with both the choice and the value of the item since the item and the reference sum of money were randomly presented to the left and right of a fixation cross. Choosing the item meant that the participant paid the reference price in exchange for an 80% chance of receiving the item. If they chose the reference amount of money, they would neither pay anything nor have the opportunity to play the lottery.

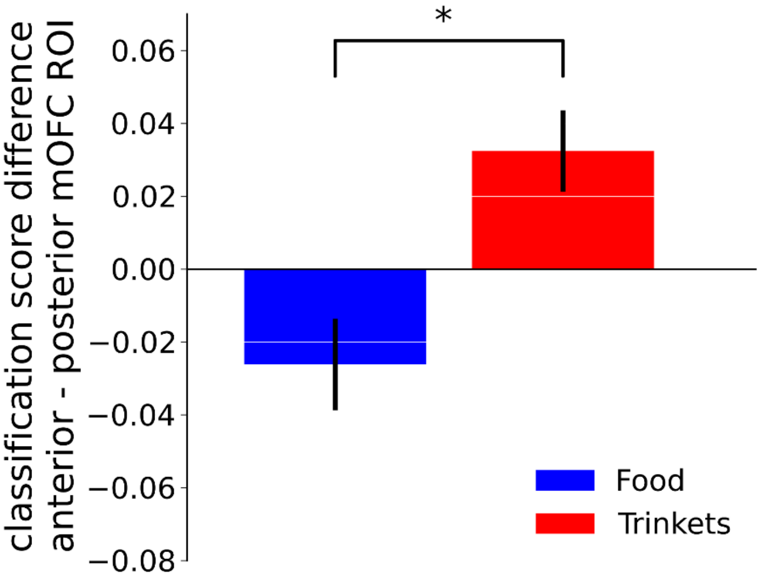
The analyses in the original study indicated that the value of the lottery item on each trial was commonly represented (as a smoothed univariate BOLD response) in a dorsal portion of vmPFC for all three item categories. This value representation was interpreted as a "decision value" signal (as opposed to a goal value in the paradigm used in the current study), since it is being computed in order to make a binary decision choice. In light of our results, we hypothesized that distributed value signals, both category-dependent and category-independent, would accompany this smoothed value signal in the ventral and dorsal portions of vmPFC, respectively. More specifically, we expected to see an anterior/posterior dissociation in category-dependent value signals along the medial orbital surface, whereby food value would be located more posteriorly and trinkets more anteriorly. We performed the same value decoding analyses as described in the main text on this dataset (19 participants; 15 male; mean age, 23.7; age range, 18-47). For this dataset, we thresholded all statistics at p < 0.005 uncorrected (unless otherwise specified), since this data was not optimized for MVPA and since we have strong *a priori* hypotheses from the primary analyses in the main text.

a.) At $p < 0.005$ uncorrected, food-category-dependent value representation was located in posterior mOFC (peak $[x = 3, y = 33, z = -24]$, $t = 2.86$). At $p < 0.05$ uncorrected, a category-independent value signal (conjunction across training/testing on food/trinkets and trinkets/food respectively) was located in mPFC (peak $[x = 6, y = 57, z = 12]$, $t = 1.98$).

b.) At $p < 0.005$ uncorrected, a trinket-category-dependent value representation was located in anterior mOFC (peak $[x = -15, y = 57, z = -9]$, $t = 2.94$). No clusters were present in any unanticipated ROI (e.g., a trinket category-dependent value signal where food category-dependent signals were found in the primary dataset). Not unexpectedly (since this dataset was not optimized for MVPA), none of these clusters reached significance under

SVFDR correction (though the category-dependent results survive small volume familywise error correction, $p < 0.05\,SVFWE$). Results are thresholded at $p < 0.005$ and $p < 0.05$ and overlaid on an averaged structural image. These results provide a completely independent replication of the ventral/dorsal and anterior/posterior vmPFC value coding dissociations observed in the main study.

Supplementary Figure 3. Leave-one-participant-out anterior/posterior mOFC gradient analysis.



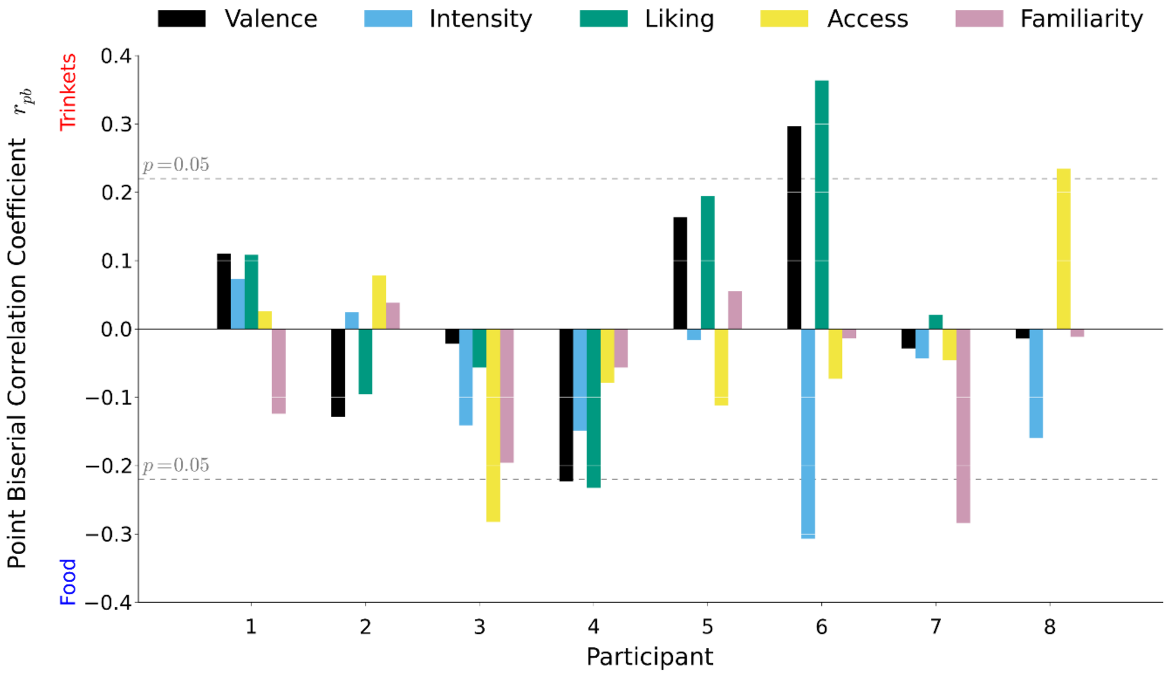
Supplementary Figure 3. Here we replicate the anterior/posterior mOFC gradients identified in the main text in a completely non-circular manner using ANOVA interaction tests applied to per-subject classification scores derived using a leave-one-participant-out approach.

For each subject and item category, we first performed second-level mass-univariate t-tests on the classification maps for 12 participants only (leaving one participant out). The peak t-score coordinate within the mOFC ROI was identified and the classification score for the left-out participant at the peak coordinate was recorded. In addition, the peak t-score from the alternative item category analysis within a searchlight sphere of voxels (restricted to the mOFC ROI) surrounding that peak coordinate was also taken. For example, for each subject we recorded two food value classification scores: (1) one based on the peak coordinate in mOFC and (2) the other based on the peak coordinate within a searchlight sphere of the peak coordinate from the trinket value decoding. Similarly, two trinket value classification scores were also acquired for each subject. In this way, for each item category and subject, we independently derived a classification score and then also recorded a classification score for the alternative item category within the same locality. This process was repeated for every subject in both analyses being contrasted. The end result was a dataset composed of four classification scores for each subject derived in a completely independent manner.

The data was entered into a repeated-measures 2x2 ANOVA design (spatial location x item category) and there was a significant interaction between the two factors ($p < 0.05$) whereby the trinket-category-dependent value encoding signal was stronger in the region identified more anteriorly but not posteriorly and vice versa for the analogous food-related signal. This replicates the corresponding result in the main text (Figure 2b) in a completely independent manner.

In this figure, the simple main effect of spatial location on classification score is plotted across item category, i.e., the distribution of the relative differences in t-scores between the anterior and posterior ROIs (food items in blue, trinkets in red). Error bars reflect standard error of the mean.

Supplementary Figure 4. Item ratings.



Supplementary Figure 4. We acquired post-hoc behavioral ratings of the food and trinket items used from 9/13 of the original participants. One participant did not complete the questionnaire, leaving 8/13 to be analyzed. The items were rated on five scales from a score of 1 to 7: “valence”, “intensity”, “liking”, “accessibility”, and “familiarity”. Items were presented in a random fashion across categories. Specifically, the questions were:

LIKING – How much do you like this item? A score of 1 means “I do not like this item at all”, a score of 4 means “I neither like nor dislike this item”, while a score of 7 means “I really like this item a lot”.

FAMILIARITY – How familiar are you with this item? A score of 1 means “This item is unknown to me”, a score of 4 means that “I am somewhat familiar with this item”, while a score of 7 means “I’m completely familiar with this item”.

INTENSITY – How intense are the feelings evoked by this item? A score of 1 indicates "This item evokes no feelings or emotion for me", a score of 4 "I have some feelings towards this item", while a score of 7 means "I have very intense feelings towards this item". Note that for this question, it is irrelevant whether the feelings/emotions you have are positive or negative.

ACCESSIBILITY – How easy do you feel it is for you to obtain this item? A score of 1 means "It is almost impossible for me to get this item", a score of 4 means "I can get this item without much difficulty", while a score of 7 means "I would have no problem getting this item".

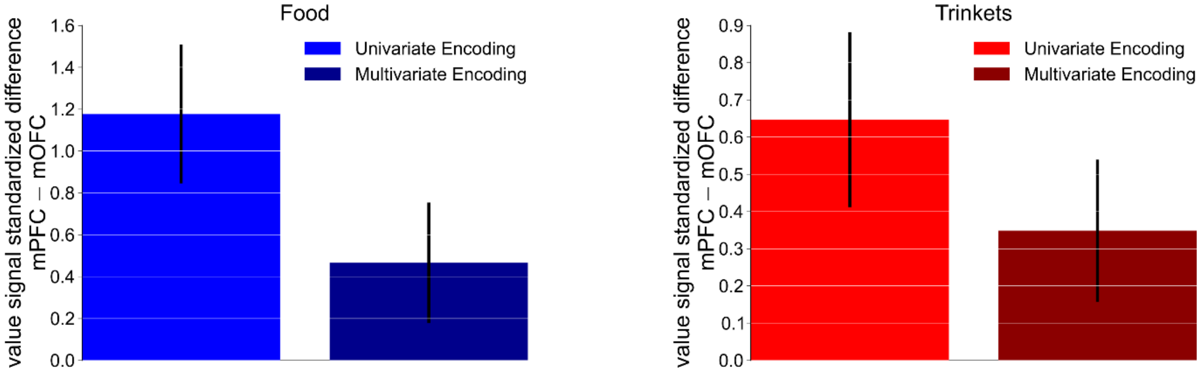
VALENCE – How pleasant or unpleasant is this item? A score of 1 means "It is a very unpleasant item", a score of 4 means "This item is neither pleasant nor unpleasant", while a score of 7 means "This is a very pleasant item".

The point-biserial correlation r_{pb} is the Pearson correlation between item ratings and the dichotomous variable indicating whether the item is a food item or a trinket. It describes to what extent higher or lower ratings are correlated with trinkets or food items. Positive correlations indicate that higher ratings correlate with trinkets; negative correlations indicate that higher ratings correlate with food items. A zero correlation implies that the ratings are evenly matched across items.

Results of statistical analyses can be seen in Supplementary Tables 2-4. At $p > 0.05$, there was no significant difference between food and trinket items with respect to any rating (across subjects or across items). In two ratings ("intensity" and "familiarity"), there was a trend towards higher ratings in the food category. The subject-level point-biserial correlation showed that this was a weak effect within individual subjects, with only one subject reaching a $p < 0.05$ significance threshold for each rating. Though these ratings were taken post-hoc, it is unlikely that the time interval since the scanning took place caused a systematic change to the between-category differences in ratings.

The bar chart in this figure reflects the point-biserial correlation coefficients r_{pb} for each subject between item ratings and a dichotomous variable which indicated whether the item was drawn from the food or trinket category. Repeated-measure statistical tests of any ratings difference between the food and trinkets category were not significant ($p > 0.05$). As can be seen from this figure, there is a high degree of variability within and across subjects in these ratings indicating that they are unlikely to account for the gradient effects reported in the main analyses.

Supplementary Figure 5. Leave-one-participant-out ventral/dorsal vmPFC gradient analysis.

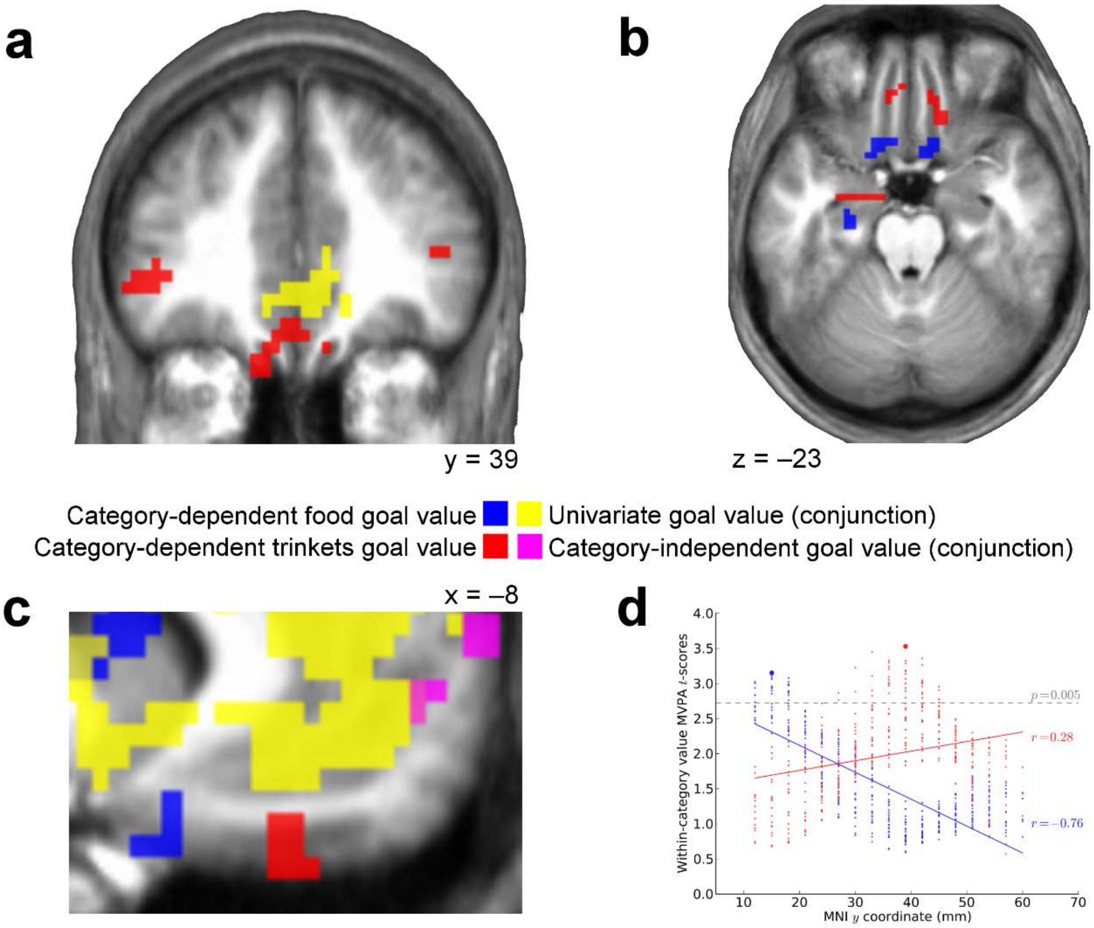


Supplementary Figure 5. Here we replicate the ventral/dorsal vmPFC gradients identified in the main text in a completely non-circular manner using ANOVA interaction tests applied to per-subject classification scores derived using a leave-one-participant-out approach.

For each analysis, we first performed second-level mass-univariate t-tests on the multivariate classification maps and general linear modeling beta maps for 12 participants only. The peak t-score coordinate within each vmPFC ROI was identified and a value representation score (classification score for the multivariate analyses or first-level GLM t-score for the univariate analyses) for the left-out participant at the peak coordinate was recorded. This process was repeated for every subject for both the food and trinket item categories. The end result was a dataset composed of four classification scores for each subject derived in a completely independent manner.

Since we seek to compare results across encoding strategies, we standardized these results by computing the distribution of standardized value signal differences between the ventral and dorsal ROIs for each item category and encoding strategy. That is, we subtracted the mPFC scores from the mOFC scores and then divided by the standard deviation across both ROIs. This data is plotted in this figure. The data was then entered into a repeated-measures 2x2 ANOVA design (spatial location x encoding strategy) and there was a significant interaction between the two factors ($p < 0.05$) across item categories, whereby there was a greater drop in signal strength in mOFC compared to mPFC for univariate encoding as opposed to multivariate encoding. This replicates the corresponding result in the main text (Figure 4a) in a completely independent manner.

Supplementary Figure 6. Value decoding based on "mean-subtraction" searchlight.


a
b
y = 39
z = –23
Category-dependent food goal value 🟦
🟨 Univariate goal value (conjunction)
Category-dependent trinkets goal value 🟥
🟪 Category-independent goal value (conjunction)
c
x = –8
d
Within-category value MVPA t-scores
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
p=0.005
r=0.28
r=−0.76
10
20
30
40
50
60
70
MNI y coordinate (mm)

Supplementary Figure 6. We have used the terms "univariate" and "multivariate" to refer to signals identified using mass-univariate general linear modeling and MVPA (after orthogonalization with respect to the univariate signals), respectively. An alternative interpretation of "univariate" and "multivariate" signals in the context of a multivariate searchlight algorithm would be the signal identified using the mean and "mean-subtracted" activity, respectively, for each sample in the searchlight. The mean-subtracted activity is the voxel responses in a searchlight after subtracting the mean voxel response across the searchlight. We repeated the value decoding analyses using this alternative approach. This involved applying the same classification procedures as in the main text except with two crucial differences: (1) the smoothed GLM-estimated goal value signal was not projected out and (2) the mean activity was subtracted and the variance across voxels normalized on a per-sample basis in every searchlight sphere (in effect, the neural pattern was standardized for every sample/sphere).

We repeated both the category-dependent and category-independent value decoding analyses in this manner. To ensure that this different methodology identified the same signals as previously in the main text, we tested for a significant activation (at $\mathrm{p} < 0.05\ \mathrm{SVFDR}$, 10 voxel extent threshold) within ROIs defined as 20mm-radius spheres (Chib et al., 2009) surrounding peaks defined by the equivalent analyses. We also checked that no activations were unexpectedly present in an alternative ROI.

Significant clusters of voxels overlayed on an averaged brain template are presented in this figure. Food-category-dependent goal value coding was observed in posterior mOFC (peak $[\mathrm{x} = 9, \mathrm{y} = 14, \mathrm{z} = -22]$, $\mathrm{t} = 3.15$, $p < 0.005\ SVFDR$) and trinket-category-dependent goal value coding in anterior mOFC (peak $[\mathrm{x} = -3, \mathrm{y} = 41, \mathrm{z} = -11]$, $\mathrm{t} = 4.20$, $p < 0.005\ SVFDR$). Cross-category value representations (conjunction across pairwise-category analyses) were located more dorsally in mPFC (peak $[\mathrm{x} = -3, \mathrm{y} = 47, \mathrm{z} = -4]$, $\mathrm{t} = 2.74$). No results were “mismatched” between the two analysis methodologies occurred. That is, no trinket-category-dependent value representation was found in the food ROI and vice versa, and no category-dependent value decoding was present in the category-independent ROI and vice versa.

We also implemented an average signal searchlight whereby we attempted to decode cross-category value signals based on the mean activity within a searchlight sphere only. Out of six training/testing data combinations (e.g., train to decode value on monetary sums, test on food items), four resulted in a significant cluster in dorsal vmPFC at $\mathrm{p} < 0.005$ uncorrected (10 voxel extent threshold), though they did not reach the corrected threshold $\mathrm{p} <$

0.05 SVFDR. The ROI was defined as a 20mm-radius sphere surrounding the peak coordinates $[x = -6, y = 53, z = -4]$ from the cross-category value decoding conjunction in the main text.

Each panel refers to an equivalent panel in the main text: Figure 2a **(**a,b**)**, Figure 3a (c**)**, and Figure 2b **(**d**)**. Results are thresholded at $p < 0.005$ uncorrected and overlaid on an averaged structural image.

Supplementary Tables

Supplementary Table 1. fMRI results.

Category	Region	Hemi	x	y	z	t	p
Univariate Value Representation							
Food*	Medial prefrontal cortex	L	-3	38	-4	4.35	<0.001
Money	Medial prefrontal cortex	L	-3	32	-7	3.24	0.001
Trinkets*	Medial prefrontal cortex		0	41	-7	4.35	<0.001
Conjunction	Medial prefrontal cortex		0	35	-7	3.14	0.001
Distributed Category-Dependent Value Representation							
Food*	Medial orbitofrontal cortex	L	-9	17	-22	3.05	0.002
Trinkets*	Medial orbitofrontal cortex	L	-3	41	-11	3.86	<0.001
Distributed Category-Independent Goal Value Representation							
Conjunction*[†]	Medial prefrontal cortex	L	-6	53	-4	2.88	0.002
Conjunction	Medial prefrontal cortex	L	-3	41	3	2.40	0.008
*Distributed Goal Category Representation**[‡]							
Conjunction	Medial orbitofrontal cortex	L	-3	20	-22	6.12	<0.001
Conjunction	Medial prefrontal cortex	L	9	29	0	7.56	<0.001
Conjunction	Lateral orbitofrontal cortex	L	-21	38	-11	11.14	<0.001
Conjunction	Frontopolar cortex	R	6	65	-11	6.89	<0.001
Conjunction	Frontopolar cortex	L	-12	71	14	6.78	<0.001
Conjunction	Dorsolateral prefrontal cortex	L	-60	17	14	11.34	<0.001

Conjunction	Dorsolateral prefrontal cortex	R	45	32	21	5.84	<0.001
Conjunction	Insula	R	45	5	3	5.27	<0.001
Conjunction	Middle frontal gyrus	L	-33	5	35	6.84	<0.001
Conjunction	Middle frontal gyrus	R	30	2	28	6.29	<0.001
Conjunction	Middle frontal gyrus	L	-18	2	60	5.68	<0.001
Conjunction	Anterior cingulate cortex	R	15	8	32	6.90	<0.001
Conjunction	Intraparietal sulcus	L	-48	-31	42	11.65	<0.001
Conjunction	Intraparietal sulcus	R	33	-70	42	7.94	<0.001
Conjunction	Precuneus	L	-6	-64	14	5.54	<0.001
Conjunction	Posterior cingulate cortex	R	3	-43	42	7.43	<0.001
Conjunction	Parahippocampal gyrus	R	36	-10	-33	6.56	<0.001
Conjunction	Inferior temporal gyrus	L	-45	-64	-22	7.64	<0.001
Conjunction	Inferior temporal gyrus	R	30	-73	-15	7.36	<0.001
Category	Region	Hemi	x	y	z	t	p

Distributed Goal Category Representation‡ (continued)*

Conjunction	Fusiform	R	24	-43	-29	6.90	<0.001
Conjunction	Fusiform	L	-27	-43	-22	7.18	<0.001
Conjunction	Extrastriate cortex	L	-12	-79	32	11.34	<0.001
Conjunction	Extrastriate cortex	R	63	-61	10	6.99	<0.001
Conjunction	Extrastriate cortex	R	9	-70	-7	5.15	<0.001
Conjunction	Striate cortex	L	-21	-79	14	9.67	<0.001

Supplementary Table 1. Results thresholded at $p < 0.05\ FDR$. Voxelwise FDR correction was performed within a ventromedial prefrontal ROI for all value-related results (i.e., SVFDR).

* Results which survive at $p < 0.005$ FDR or SVFDR.

[†] Conjunction across five binary category permutations (all except training value on money and decoding value on trinkets).

[‡] Conjunction across all three binary category combinations.

Supplementary Table 2. Item ratings, subject-level analysis.

	P1	P2	P3	P4	P5	P6	P7	P8
Valence ($\mathbf{r}_{\mathbf{pb}}$)	0.110	-0.129	-0.022	-0.223	0.163	0.296	-0.028	-0.014
Intensity ($\mathbf{r}_{\mathbf{pb}}$)	0.073	0.025	-0.141	0.149	-0.016	-0.307	-0.043	-0.160
Liking ($\mathbf{r}_{\mathbf{pb}}$)	0.108	-0.095	-0.056	-0.232	0.195	0.364	0.021	0.000
Access ($\mathbf{r}_{\mathbf{pb}}$)	0.026	0.079	-0.282	-0.079	-0.112	-0.073	-0.046	0.235
Familiarity ($\mathbf{r}_{\mathbf{pb}}$)	-0.124	0.038	-0.196	-0.056	0.056	-0.014	-0.284	-0.011

	Mean	SEM	p
Valence ($\mathbf{r}_{\mathbf{pb}}$)	0.019	0.058	0.750
Intensity ($\mathbf{r}_{\mathbf{pb}}$)	-0.090	0.043	0.077
Liking ($\mathbf{r}_{\mathbf{pb}}$)	0.038	0.065	0.579
Access ($\mathbf{r}_{\mathbf{pb}}$)	-0.032	0.053	0.572
Familiarity ($\mathbf{r}_{\mathbf{pb}}$)	-0.074	0.042	0.122

Supplementary Table 2. Trinkets (+), Food (–), grey background indicates significance at $\mathrm{p} = 0.05$ ($\mathrm{r}_{\mathrm{pb}} = \pm0\ .2199$)

Supplementary Table 3. Item ratings, distribution per category across subjects.

	Food		Trinkets		Food > Trinkets
	Mean	SEM	Mean	SEM	Repeated-measures t-test
Valence	4.450	0.199	4.456	0.107	p = 0.973, t = -0.035
Intensity	4.028	0.166	3.712	0.146	p = 0.059, t = 2.250
Liking	4.400	0.242	4.494	0.126	p = 0.692, t = -0.412
Access	5.309	0.265	5.225	0.270	p = 0.640, t = 0.489
Familiarity	5.534	0.214	5.278	0.207	p = 0.111, t = 1.822

Supplementary Table 3.

Supplementary Table 4. Item ratings, distribution per category across items.

	Food		Trinkets		Food > Trinkets
	Mean	SEM	Mean	SEM	Independent t-test
Valence	4.450	0.126	4.456	0.109	p = 0.970, t = -0.037
Intensity	4.028	0.100	3.712	0.162	p = 0.101, t = 1.661
Liking	4.400	0.133	4.494	0.129	p = 0.614, t = -0.507
Access	5.309	0.141	5.225	0.134	p = 0.666, t = 0.440
Familiarity	5.534	0.144	5.278	0.190	p = 0.285, t = 1.077

Supplementary Table 4.

Supplementary Table 5. Items used organized by category.

Food Items	Money Items	"Trinket" Items
Ambrosia	20c	1984, George Orwell (book)
Apple Pies	30c	A Brief History of Time, Stephen Hawkings (book)
Bombay Mix	40c	A Portrait of the Artist as a Young Man, J. Joyce (book)
Cashews	60c	Abbey Road, The Beatles (music CD)
Choco Chip Cookies	70c	Alarm Clock
Coco Pops	90c	Batteries
Cornflakes	1.2EUR	Blade Runner (movie DVD)
Cream Crackers	1.3EUR	Blank DVDs
Crunchies	1.5EUR	Bourne Ultimatum (movie DVD)
Digestives	1.6EUR	Calendar
Doritos	1.8EUR	Combination Lock
Elevenses	1.9EUR	Dracula (book)
Fig Rolls	2EUR	Family Guy Season 7 (TV series DVD)
Fingers	2.1EUR	Golden Compass, Philip Pullman (book)
Frosties	2.2EUR	Harry Potter (book)
Fruit Pastilles	2.3EUR	Indiana Jones (movie DVD)
Gherkins	2.4EUR	James Bond, Quantum of Solace (movie DVD)
Granola Bar	2.5EUR	Joshua Tree, U2 (music)
Spam	2.6EUR	Kings of Leon Live (music/movie DVD)
Jaffa Cakes	2.7EUR	Lord of the Rings (movie DVD)
Bacon Fries	2.8EUR	Monopoly (boardgame)
Liquorice All Sorts	2.9EUR	OK Computer, Radiohead (music CD)

Mikado	3EUR	Playing Cards
Mini Rolls	3.1EUR	Shampoo
Beetroot	3.2EUR	Sherlock Holmes (book)
Pineapple Rings	3.3EUR	Slumdog Millionaire (movie DVD)
French Fancies	3.4EUR	Socks
Pickled Onions	3.5EUR	The Departed
Green & Black Chocolate	3.6EUR	The Hitchhiker's Guide to the Galaxy (book)
Rice Krispie Squares	3.7EUR	Stapler
Riesen	3.8EUR	T-Shirt
Salted Peanuts	3.9EUR	The Dark Knight (movie DVD)
Sesame Sticks	4.1EUR	The Wire Season 4 (TV series DVD)
Pringle's Original	4.2EUR	Travel Plug Adaptor
Fox's Shortcakes	4.3EUR	Trinity College Key Chain
Tea Cakes	4.4EUR	Trinity College Mug
Terry's Orange	4.6EUR	Trinity College Sweatshirt
Pickled Eggs	4.7EUR	Umbrella
Walkers Crisps	4.8EUR	USB Key 2GB
Werthers Sweets	4.9EUR	Water Bottle

Supplementary Table 5. The items used were similar to those used in Chib et al, 2009, although they were customized to be familiar to participants in Ireland, where the study was performed. Our motivation for using these specific goods is that we wanted to include large varied groups of items that were approximately similar in their average values to participants so as to control for the effects of value per se when doing between category comparisons. In addition, we required that the items be “everyday” items to ensure that all the subjects would be similarly exposed to the items and thus would be able to reasonably evaluate them. All subjects reported a high degree of familiarity with each of the items in a post-scan verbal report. Monetary amounts were selected in the range 10c to 5 euros, in 10c increments. Item order was randomly determined at the beginning of each experiment. Forty items were used in each class.

Subjects were only allowed to bid using a discrete set of values (Chib et al., 2009, Plassmann et al., 2007), thus the bids recorded are an approximation to the true values for the items. A true WTP of €1.20 is measured as €1, and if the subject values an item at €4.50, we record a value of €4. However, a correlation analysis reported in Plassmann et al., 2007 showed that this discretized WTP distribution strongly reflects how much a subject likes the items, and thus can be taken as a good approximation to their true subjective goal-valuations.

Supplementary Table 6. Items used in Chib et al., 2009 organized by category.

Food Items	Money Items	Trinket Items
Chocolate Chip Cookies	20c	300 (movie DVD)
Chocolate Pudding	40c	A Brief History of Time, Stephen Hawkings (book)
Cookies	60c	Batman Begins (movie DVD)
Doritos Chips	80c	Blade Runner (movie DVD)
Fig Rolls	$1	Bourne Ultimatum (movie DVD)
Ghiradelli Chocolate Bar	$1.2	Caltech Backpack
Hershey's Milk Chocolate Bar	$1.4	Caltech Cap
Ho-Ho Chocolate Cake Rolls	$1.6	Caltech Flag
Kit Kat Chocolate Bar	$1.8	Caltech Key Chain
Lindt Chocolate Bar	$2	Caltech Mug
Mrs Fields Cookies	$2.2	Caltech Travel Mug
Oreo Cookies	$2.4	Caltech Sack
M & Ms	$2.6	Caltech Straw Hat
Powdered Donuts	$2.8	Freakonomics, S. Levitt & S. Dubner (book)
Pringles Chips	$3	Indiana Jones Boxed Set (movie DVDs)
Reeses Peanut Butter Cups	$3.2	Stapler
Rice Krispie Squares	$3.4	Spiderman (movie DVD)
Skittles Sweets	$3.6	The Big Lebowski (movie DVD)
Sweetarts Hard Candy	$3.8	The World is Flat, T. Friedman (book)
Twix Chocolate Bar	$4	Transformers (movie DVD)

Supplementary Table 6.

Methods

Task.

Subjects were presented with high-resolution images of three classes of goods: snacks, consumer goods (e.g., DVDs, books), and monetary prizes (see Table S5 for details). On each trial, participants bid for the right to the prospect of obtaining a displayed item with 80% probability and nothing otherwise. We introduced the probabilistic element to ensure that valuations for monetary sums would be nontrivial. Bids were elicited using a Becker-DeGroot-Marschack (BDM) auction process. On a given trial, the participant bids €0, €1, €2, €3, or €4 for an item. At the end of the experiment one trial is selected at random from each of the categories. For each trial selected, a random number is drawn with equal probability from the categories of €0, €1, €2, €3, or €4. If the bid equals or exceeds the amount drawn in the lottery, then participants pay the bid amount and receive the corresponding item prospect. Otherwise, they pay nothing. These rules favor an optimal strategy of bidding the amount closest to one's subjective valuation. The BDM rules were fully explained to participants. Subjects were asked to refrain from eating for four hours before arrival for testing. Compliance was confirmed through self-reports. Participants were requested to remain in the laboratory for one hour post-scan to consume items obtained during the experiment. This helped maximize participants' valuation for food items during testing. On each trial subjects were endowed with €4 for bidding (since one trial from each category is ultimately played out, this corresponds to a €12 endowment across all three categories). Any remaining money from the initial endowment is retained by the subject.

Each trial began with a stimulus presentation (Figure 1a). Subjects generated a bid within 5s by pressing one of four buttons or not responding for a zero bid, followed by a presentation of the bid amount (500ms). The inter-trial interval was uniformly drawn from 1-23 seconds. Four sessions of length 16 minutes each were completed. The hand used for responding was switched after two sessions and the correspondence between the buttons and bids was alternated for the second and fourth sessions. The button configurations were practiced at the beginning of each session.

fMRI Data Acquisition.

Fifteen healthy right-handed subjects participated in this study. The data from two subjects were excluded because of technical problems with the MRI scanner, leaving thirteen subjects (eight male; mean age 22.1; SD 3.6 years). All subjects gave informed consent and the experiment was approved by the School of Psychology Research Ethics Committee, Trinity College Dublin. Functional imaging was performed on a 3T Philips scanner with an 8-channel SENSE head coil at Trinity College Institute of Neuroscience, Dublin, Ireland. Thirty-five contiguous sequential ascending echo-planar T2*-weighted slices were acquired for each volume, giving whole brain coverage with a slice thickness of 3.55mm and no slice-gap (in-plane resolution: 3.00 x 3.00 mm; repetition time (TR): 2000ms; echo time (TE): 30ms; field of view: 240 x 240 mm^2; matrix: 80 x 80). A whole-brain high-resolution T1-weighted structural scan (voxel size: 0.9 x 0.9 x 0.9mm^3) was also acquired for each subject. Slice orientation was tilted -30° from a line connecting the anterior and posterior commissure to alleviate signal loss in the OFC (Vikram S. Chib et al., 2009).

Data Preprocessing and Filtering.

Slice timing correction, motion correction, and spatial normalization were applied to the data. For the general linear model (GLM), the data was high-pass filtered (120s cut-off), and serial autocorrelations were estimated using a first-order autoregressive model.

To minimize analysis differences between the univariate and multivariate approaches, we carried out the following: prior to multi-voxel sample extraction, low frequency components (below 1/120Hz), serial autocorrelations, and head motion were subtracted from the data. In addition, smoothed univariate value signals for all three categories identified in the GLM analysis were removed from the data to ensure that the multi-voxel patterns identified in the MVPA analyses do not reflect overlying univariate signals. This was accomplished by multiplying the convolved parametric value regressor by the beta estimated in the GLM and subtracting the resulting time series from the data on a per-voxel basis. To correct for session-related mean and scaling effects, we applied second-order detrending and z-scoring on a per-voxel per-session basis (Alan N. Hampton & O'Doherty, 2007; Kahnt et al., 2010). Here we use the terms "univariate" and "multivariate" to refer to signals identified using mass-univariate general linear modeling and MVPA (after orthogonalization with respect to the univariate signals), respectively. An alternative interpretation of "univariate" and "multivariate" is the signal identified using the mean and "mean-subtracted" activity, respectively, within the searchlight. We repeated the value decoding analyses using this alternative approach, which yielded similar results (Supplementary Figure 6). We applied spatial smoothing (8mm full-width-half-maximum) to the data used for the univariate GLM, but not in the multi-voxel pattern analysis, in order to preserve local variance (Kahnt et al., 2010; Pereira et al., 2009). Pre-processing and filtering were performed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/), except detrending and z-scoring for which the PyMVPA package was used (Hanke et al., 2009).

General Linear Model.

We used a GLM to identify activity at decision time correlating with goal-values (as measured by WTP). The GLM included regressors for image presentation and bid defined for each item category (0s duration). Subject-specific WTPs were used as a parametric modulator for each regressor. To minimize head motion confounds, motion parameters were included as nuisance regressors. For the second-level analysis, beta maps corresponding to the WTP regressors for each subject for each item category were included in a 3x1 factorial design (each category being a factor). To test for regions representing stimulus value for all item categories in a univariate manner, we performed a conjunction analysis across all three categories using the “conjunction null” hypothesis (Nichols et al., 2005).

Classification Algorithm.

We used a Gaussian Naive Bayes (GNB) classification algorithm(Mitchell, 1997) with an assumption of zero covariance across voxels . To perform binary classification, our algorithm first estimates mean activity vectors and covariance matrices from training data for the Gaussian distributions $\mathbf{p}\ (\mathbf{x}|\mathbf{A})$ and $\mathbf{p}\ (\mathbf{x}|\mathbf{B})$. Then, the algorithm assigns a test sample $\mathbf{x}_{\mathbf{test}}$ to the condition with the maximum posterior probability at $\mathbf{x}_{\mathbf{test}}$ based on the estimated distributions: if $\mathbf{p}\ (\mathbf{x}_{\mathbf{test}}|\mathbf{A})\ >\ \mathbf{p}\ (\mathbf{x}_{\mathbf{test}}|\mathbf{B})$ the algorithm infers that $\mathbf{x}_{\mathbf{test}}$ was sampled under condition $\mathbf{A}$. Generalization accuracy is estimated using cross-validation (Mitchell, 1997). This involves training and testing on mutually exclusive subsets of samples and repeating with a different partitioning on each “fold”. Cross-validation was done on a leave-one-session-out basis. On every fold, the classifier was trained on three sessions and tested on the remaining session, thereby avoiding session-related dependencies between training and testing samples (N. Kriegeskorte, W. K. Simmons, P. S. F. Bellgowan, & C. I. Baker, 2009; Mitchell, 1997; Pereira et al., 2009). Accuracy scores were averaged to give the generalization accuracy. All preprocessing and filtering was performed on a per-session basis.

Multivariate Pattern Analysis.

A searchlight procedure (Kriegeskorte et al., 2006) provided a spatially unbiased estimator of distributed activity across the brain. Each fMRI data sample had two task-related characteristics, stimulus category and value. A potential concern is that significant correlation between stimulus category and stimulus values could bias the classification results, since the classifier might leverage variance which distinguishes between categories when attempting to decode value, and vice versa. WTP for food was lower on average compared to the money or trinket categories (Figure 1c). To address this concern the set of samples for each category was median split into ‘high’ and ‘low’ value classes on a cross-session basis for each subject. This relabeling eliminates correlations between value and category labels for every subject (Spearman correlation, $p > 0.2$ for all subjects), resulting in six classes of samples, one for each value/category combination. To avoid class imbalance bias, all analyses were balanced on a per-session basis (i.e., the number of samples in each class was equalized for each session and therefore cross-validation folded) by randomly removing some samples. Analyses were run multiple times to confirm that the outcome of the analysis was not dependent on the balancing procedure.

Category-independent value.

We identified category-independent value signals as those whose representations enabled decoding of value level across stimulus categories. We ran all six binary cross-category value classification analyses by training to decode high vs. low value on samples drawn from one category (e.g., food) and testing on samples drawn from another (e.g., money).

Within-category value.

We searched for areas encoding value that were able to predict within the same category. Note that the value representations pinpointed in this analysis may or may not be category-dependent, but the results of this exercise are necessary to carry out the category-dependent analysis described next.

Category-dependent value.

We identified regions involved in category-dependent valuation as those that allowed us to decode values only within particular categories. These value representations would be coded in voxel response distributions which differ across categories.

For this, we compared results of the cross-category and within-category value decoding analyses. We first identified voxels that could significantly decode (p $< 0.005\ SVFDR$) between high and low values within each category. Next we tested if these areas could predict value across categories. Any voxel that survived the cross-category analysis even at p < 0.05 (corrected for two comparisons at each voxel) was deemed to exhibit properties of category-independent value encoding. Clusters which survived the within-category analysis, but did not survive the cross-category analysis, were deemed to involve category-dependent valuation.

Stimulus category identity.

Finally, we looked for regions exhibiting multivariate encoding of stimulus category. We implemented three binary classification analyses – food vs. money, money vs. trinkets, and food vs. trinkets. The searchlight accuracy maps were entered into a conjunction analysis (Nichols et al., 2005) to identify regions whose activity discriminated between all category pairs. This ensured that areas of the brain identified by this analysis contained distributed codes pertaining to the identity of each stimulus category individually.

Significance Testing.

For the searchlight analyses, the percentage of correctly identified samples averaged across folds in the cross-validation was used as the classification score in each searchlight, and this score was assigned to the voxel at the center. This defined q classification accuracy map for each subject, which was then smoothed with an 8mm FWHM kernel. A second-level analysis was implemented by performing voxel-wise t-tests, comparing the distribution of accuracies across subjects against 50%, which is the expected performance of an algorithm randomly labeling samples. Since multivariate classification is susceptible to optimistic classification biases, we carried out permutation tests to validate our decoding procedure (Mukherjee, Golland, & Panchenko, 2003) (see Permutation Testing For Multivariate Analyses).

All univariate and multivariate results were significant at FDR-adjusted $\mathrm{p} < 0.05$ corrected for multiple comparisons by controlling the voxelwise false discovery rate (FDR) with a 10 voxel extent threshold. We had a strong prior hypothesis regarding value signals in medial prefrontal regions (Elliott et al., 2000; John P. O'Doherty et al., 2001; Padoa-Schioppa & Assad, 2006; L. Tremblay & Schultz, 1999; Wallis, 2007, 2011), thus, for value-based analyses, correction was performed within a vmPFC mask defined *a priori* from related functional (Vikram S. Chib et al., 2009) and anatomical (Beckmann, Johansen-Berg, & Rushworth, 2009) studies (see Supplementary Figure 1). This correction threshold is denoted $\mathrm{p} < 0.05$ SVFDR. For other analyses, unrelated to value, whole-brain correction was used (denoted $\mathrm{p} < 0.05$ FDR). For display purposes, we present all results at $\mathrm{p} < 0.005$. Results corrected within a small volume are displayed uncorrected. All results are overlaid on a normalized T1-weighted image averaged across subjects. Our main results are based on the $\mathrm{p} < 0.05$ SVFDR threshold (and displayed at p<0.005 uncorrected) since (a) it was used previously in a similar paradigm (Vikram S. Chib et al., 2009), thus allowing a direct signal power comparison, and (b) controlling the false discovery rate rather than the familywise error rate has been shown to have greater sensitivity with minimal risk of false positives (Chumbley, Worsley, Flandin, & Friston, 2010).

Permutation Testing For Multivariate Analyses.

For each multivariate analysis, the searchlight procedure was repeated 200 times with permuted labeling (Krajbich et al., 2009; Kriegeskorte et al., 2006; Pereira et al., 2009). To satisfy exchangeability criteria (Pereira & Botvinick, 2011) and to prevent label imbalances in the cross-validation, labels were permuted along with their positions in the dataset partitions. The resulting accuracy maps were entered into mass univariate t-tests to determine if the accuracy distributions over the permuted datasets were significantly different from chance. At $\mathbf{p < 0.1}$, for all analyses, no voxel's accuracy distribution significantly deviated from random chance in any subject. This indicates that the classification algorithm used for the data analysis across all conditions is fair and unbiased, i.e., the significant results reported for the non-permuted labels are not due to an optimistic classification bias.

ROI Gradient Analyses.

The t-score maps computed at the second-level in our univariate and within-category multivariate value analyses are indicative of the relative strengths of distinct types of value coding within vmPFC. We used these maps to investigate how the structure of stimulus value representation varies along (a) an anterior-posterior gradient in mOFC in relation to the abstractness of the stimulus being valued and (b) a ventral-dorsal gradient in vmPFC as a whole with respect to the relationship between the univariate and multivariate representation of value.

Anterior-posterior gradient of stimulus abstractness.

For voxels in mOFC, the t-scores obtained from the within-category value decoding analyses were tested for a correlation with the position of the voxels along the y-axis (Figure 2b). This was done for the food and trinkets categories separately. Since the smoothing applied to classification accuracy maps prior to the second-level analyses artificially inflates the strength of any spatial correlation, we generated a more reasonable correlation distribution under the null hypothesis by randomly generating noise within mOFC using the same mean and variance as in the empirically observed unsmoothed t-scores. We then smoothed this noise and computed the t-score/y-axis correlation, repeating this process 10,000 times. A non-parametric p-value was derived by determining the fraction of randomly generated correlations which exceeded the actual correlation.

Ventral-dorsal gradient of value processing complexity.

Three analyses were performed to compare univariate and multivariate value signals in mOFC and mPFC: first, we correlated each voxel's univariate and within-category MVPA t-scores with its position along the z-axis (Figure 4a). This was done for all voxels in the mOFC and mPFC masks together. We generated a null correlation distribution for each combination of category and value coding strategy by randomly generating correlations from simulated data generated using the same process described above. The null correlation distribution defines a non-parametric p-value as the proportion of randomly generated correlations which exceed the empirically observed correlation scores. Since we sought to determine whether or not the univariate and distributed coding strengths were differentially correlated with the z-axis, we derived confidence intervals around the respective correlation estimations via bootstrapping. That is, 10,000 samples were randomly generated with replacement and a sampling distribution estimated for each category and value coding strategy. From this sampling distribution, we can establish the range of values that the actual correlations might take (within an error probability thresholded at $p < 0.05$).

Second, we examined how voxel preference for multivariate or univariate coding of value changes along an inferior-superior axis. To do this, we extracted the t-scores obtained in the second-level analyses for the univariate and within-category MVPA analyses for all voxels in each mask, and then for each voxel we subtracted the univariate t-score from the MVPA t-score, resulting in a single value indicative of that voxel's relative preference for the multivariate or univariate encoding value. This was done for all voxels in mPFC and mOFC separately (Figure 4b). The resulting samples were tested using two-sided paired t-tests.

In our third test, we correlated the second-level t-scores from the univariate and within-category MVPA analyses on a voxel-by-voxel basis in each region. Again, this procedure was implemented for the food and trinkets categories separately. Since the number of voxels in each vmPFC subdivision is different, we tested differences in correlations using a bootstrap procedure (Hastie, Tibshirani, & Friedman, 2008). For each combination of stimulus category and vmPFC subdivision, we resampled 348 data points of interest with replacement (corresponding to the number of voxels in the larger mOFC mask) and computed the correlation. In this way, 10,000 correlation coefficients were generated (Figure 3b), giving an estimate of the empirical distribution.

Chapter 3

PROSPECTIVE AND RETROSPECTIVE CORTICAL AND STRIATAL REPRESENTATIONS OF DECISION VARIABLES

While there is accumulating evidence for the existence of distinct neural systems supporting goal-directed and habitual action selection in the mammalian brain, much less is known about the nature of the information being processed in these different brain regions. Associative learning theory predicts that brain systems involved in habitual control, such as the dorsolateral striatum, should contain stimulus and response information only, but not outcome information, while regions involved in goal-directed action, such as ventromedial and dorsolateral prefrontal cortex and dorsomedial striatum, should be involved in processing information about outcomes as well as stimuli and responses. To test this prediction, human participants underwent fMRI while engaging in a binary choice task designed to enable the separate identification of these different representations with a multivariate classification analysis approach. Consistent with our predictions, the dorsolateral striatum contained information about responses but not outcomes at the time of an initial stimulus, while the regions implicated in goal-directed action selection contained information about both responses and outcomes. These findings suggest that differential contributions of these regions to habitual and goal-directed behavioral control may depend in part on basic differences in the type of information that these regions have access to at the time of decision-making.

Introduction

Two distinct strategies support behavioral control: a goal-directed strategy that flexibly generates decisions based on deliberate evaluation of the consequences of actions, and a habitual strategy that relies on a reflexive, automatic, elicitation of actions (B.W. Balleine, Daw, & O'Doherty, 2008; B. W. Balleine & A. Dickinson, 1998; Dickinson, 1985). These distinct mechanisms depend on at least partly dissociable brain systems, with the posterior dorsolateral striatum (DLS) being implicated in habits (E. Tricomi, B. W. Balleine, & J. P. O'Doherty, 2009; Yin, Knowlton, & Balleine, 2004) and the ventromedial prefrontal cortex (vmPFC) (or homologous regions in the rodent brain) and the dorsomedial striatum (DMS) contributing to goal-directed control (B. W. Balleine & A. Dickinson, 1998; Killcross & Coutureau, 2003; S. C. Tanaka, B. W. Balleine, & J. P. O'Doherty, 2008; Yin, Ostlund, Knowlton, & Balleine, 2005). However, the nature of the information encoding in these regions is much less understood.

According to associative learning theory, in habits, associations are formed between stimuli (S) and responses (R), without any encoding of the goal or outcome (O). By contrast, in goal-directed learning, associations are formed between stimuli, responses, and outcomes (B. W. Balleine & A. Dickinson, 1998). Specifically, goal representations are suggested to be elicited via S-O associations, which in turn retrieve response representations via an O-R association (Bernard W Balleine & Ostlund, 2007; Sanne de Wit & Dickinson, 2009).

Several neurophysiology studies have explored whether brain regions implicated in habitual and goal-directed action contain different types of associative information. The results of such studies are equivocal, with most reporting similar information encoding in both DMS and DLS (Gremel & Costa, 2013; H. Kim, Lee, & Jung, 2013; Hoseok Kim, Sul, Huh, Lee, & Jung, 2009; Kimchi, Torregrossa, Taylor, & Laubach, 2009; Thomas A Stalnaker, Gwendolyn G Calhoon, Masaaki Ogawa, Matthew R Roesch, & Geoffrey Schoenbaum, 2010). Another approach has been to adopt formal computational models of learning and correlate these to data acquired with fMRI or neurophysiology (J. O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003; Samejima, Ueda, Doya, & Kimura, 2005). Some studies report that different types of reinforcement-learning models (model-free vs model-based) correlate with activity in DLS compared to DMS, vmPFC, and dorsolateral prefrontal cortex (dlPFC) (Glascher et al., 2010; A. N. Hampton et al., 2006; Lee, Shimojo, & O'Doherty, 2014; K. Wunderlich, Dayan, & Dolan, 2012) while others have found evidence for more mixed representations (N. D. Daw, Gershman, Seymour, Dayan, & Dolan, 2011; Simon & Daw, 2011). However, while such analyses reveal the computations that might be operating in a given area, they do not illuminate the type of information being encoded in those regions upon which a particular computational process may act.

In the present study, human participants underwent fMRI while performing a binary decision task, in which we carefully manipulated response and outcome identity across experimental sessions. Using a multivariate pattern analysis classification method, we tested for the presence of response information and outcome information at the time of stimulus and action performance in different brain regions. We hypothesized that brain regions implicated in habitual control, such as the DLS, would encode response, but not outcome, information at the time of stimulus presentation, indicating a role for this region in supporting stimulus-response associations, while other brain regions such as the vmPFC, dlPFC, and anterior DMS would contain representations of both responses and outcomes, indicative of a role for those regions in goal-directed learning and control.

Materials and Methods

Participants

Nineteen healthy right-handed volunteers participated in this study (11 male; mean age 22.9; SD 4.1 years). The volunteers were pre-assessed to exclude those with a history of neurological or psychiatric illness. All participants gave informed consent and the study was approved by the Institutional Review Board of the California Institute of Technology.

Task

Participants performed a simple binary decision task (Figure 1A). At the start of each trial an initial stimulus environment was indicated by one of two Sanskrit characters. The participant then performed one of two possible actions and subsequently entered an outcome state in which they received an associated reward. There were two distinct initial stimuli, two possible outcome states (respectively represented by a blue circle and a red square), and two reward distributions—high (equal probability of $8, $10, $12) and low (equal probability of $2, $4, $6). Participants interacted with the environment using two qualitatively different actions: a double button press and a trackball roll. These actions were performed on the same device using the right hand. Action outcomes were anti-correlated and deterministic, i.e., actions always led to outcomes and these outcomes (and associated reward distributions) were always distinct. The initial stimulus determined the subsequent action-outcome contingencies and thus indicated which of the two actions was highly rewarded on a given trial. Crucially, the relationships between initial stimuli, actions, outcomes, and reward distributions were permuted across four conditions, ensuring that representations of one decision variable (e.g., actions) could not be confounded with those of another (e.g., outcomes) (phi correlation coefficient < |0.005|). A description of the full permutation order is provided in Table 1. Participants received training prior to each condition to ensure that they were fully aware of the relevant action-outcome contingencies. Twelve sessions were run in total (three for each condition) with sixteen trials in each session (eight for each initial stimulus). In addition, after training, participants nearly always selected actions leading to the high reward distribution and thus could not confound our decoding of initial stimulus, actions, and outcomes with the anticipated level of reward.

fMRI Data Acquisition

Functional imaging was performed with a 3 T Siemens Trio scanner. Forty-five contiguous interleaved transversal slices of echo-planar T2*-weighted images were acquired in each volume, with a slice thickness of 3mm and no gap (repetition time, 2650ms; echo time, 30ms; flip angle, 90°; field of view, 192mm2; matrix, 64x64). Slice orientation was tilted 30° from a line connecting the anterior and posterior commissure. This slice tilt alleviates the signal drop in the OFC (Deichmann, Gottfried, Hutton, & Turner, 2003). We discarded the first three images before data processing and statistical analysis, to compensate for the T1 saturation effects. A whole-brain high-resolution T1-weighted structural scan (voxel size: 1 x 1 x 1mm^3) was also acquired for each subject.

Data Preprocessing and Filtering

Slice timing correction, motion correction, and spatial normalization were applied to the data. Prior to multi-voxel sample extraction, low frequency components (below 1/120Hz), serial autocorrelations, and head motion were subtracted from the data. To correct for session-related mean and scaling effects, we applied second-order detrending and z-scoring on a per-voxel per-session basis (Pereira et al., 2009). Pre-processing and filtering was performed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/), except detrending and z-scoring, for which the PyMVPA package was used (Hanke et al., 2009).

General Linear Model

Eight regressors of interest were included in the general linear model (GLM). Each regressor corresponded to the identity of a particular decision variable (i.e., one regressor for each initial stimulus, each action, each outcome, and each reward distribution). In addition, parametric modulators reflecting the actual reward delivered on a given trial were added to the reward distribution regressors. Time series of head motion estimated during realignment were included as covariates of no interest.

Classification Algorithm

We used a Gaussian Naive Bayes (GNB) classification algorithm (Mitchell, 1997) with an assumption of zero covariance across voxels. To perform binary classification our algorithm first estimates mean activity vectors and covariance matrices from training data for the Gaussian distributions $p(x|A)$ and $p(x|B)$. Then, the algorithm assigns a test sample x_{test} to the condition with the maximum posterior probability at x_{test} based on the estimated distributions: if $p(x_{test}|A) > p(x_{test}|B)$ the algorithm infers that x_{test} was sampled under condition A. Generalization accuracy is estimated using cross-validation. This involves training and testing on mutually exclusive subsets of samples and repeating with a different partitioning on each “fold”. Cross-validation was done on a leave-one-session-out basis. On every fold, the classifier was trained on three sessions and tested on the remaining session, thereby avoiding session-related dependencies between training and testing samples (Nikolaus Kriegeskorte et al., 2009; Mitchell, 1997; Pereira et al., 2009) . Accuracy scores were averaged to give the generalization accuracy. All preprocessing and filtering was performed on a per-session basis. Importantly, the average Spearman correlation between combinations of decision variables across subjects was $2x10^{-5}$, $1x10^{-3}$, and $2x10^{-4}$, respectively, indicating that the classifier could not erroneously decode one decision variable based on correlated representations of another.

Multivariate Pattern Analysis

A searchlight procedure (Kahnt et al., 2010; Kriegeskorte et al., 2006) provided a spatially unbiased estimator of distributed activity across the brain. This involved the performance of GNB classification on the fMRI data in spheres of voxels of radius 3 throughout the brain. We extracted a sample of neural data corresponding to the initial stimulus, action, and outcome timepoints in each trial (with a shift of 5 s to account for hemodynamic delay) by averaging the two volumes closest in time (one before and one after) to the relevant timepoint (Clithero et al., 2011; McNamee et al., 2013). Each fMRI data sample had two task-related characteristics: timepoint and identity. Our hypotheses required us to decode based on a variety of interactions between these two characteristics, which we detail below. In particular, we performed the following analyses in order to determine whether neural representations of decision variables are present at timepoints in a trial other than the moment of perception or action. For all analyses below, cross-fold validation was used, in which training was done on the data from 11 sessions, and testing was done on the data from the 12th session. This was then repeated 12 times, using a different test session on each occasion.

Time-span Decoding

In “time-span” decoding analyses, we trained the classifier at one time point in the trial to discriminate activity patterns elicited at another time point in the trial (Figure 1B).

Time-shift Decoding

Here we train and test at the same timepoint in the trial but label the samples according to an alternate timepoint. For example, we decode preceding action at the time of outcome state presentation. The important distinction between these “time-shift” analyses and the previous “time-span” analyses is that, in our example, the “time-shift” analyses do not require that the action representation at the outcome timepoint be the same as that at the action timepoint.

Action at Stimulus Time

To detect regions involved in encoding action representations at the time of initial stimulus presentation, we trained our classifier to discriminate between different action representations (double button press vs. trackerball roll) at the time of action selection. We then tested the classifier at the time of initial stimulus presentation to assess whether activity in a given brain area during the time of initial stimulus presentation reflected the action that would subsequently be selected on that trial. A successful classification in a given brain region would indicate that information about the to-be-performed action is being represented during the initial decision period, suggesting the presence of stimulus-response associations in that region.

Outcome at Stimulus Time

To detect regions involved in encoding outcome representations at the time of initial stimulus presentation in the trial, we trained our classifier to discriminate different target outcomes, e.g., blue circle vs red square, as they were presented at the time of outcome delivery. We then tested the classifier at the time of initial stimulus presentation (i.e., at the onset of the trial), to assess whether activity in a given brain area during the time of initial stimulus presentation reflected the outcome that would ultimately be delivered on that trial (contingent on the subsequent action). A successful classification in a given brain area would indicate that information about the goal of an action is represented in that area during the initial decision period.

Decoding of Integrated Representations

We were also interested in testing for “integrated representations”, in which distinct combinations of stimuli and actions (e.g., S1-A1 and S2-A2) might be encoded as unique configurations (S1A1 vs S2A2), as opposed to being encoded as elemental action representations (A1 vs A2). The key distinction is that an integrated representation of an S1-

A1 combination would successfully decode only on trials in which A1 is selected in the presence of S1 but not otherwise; in contrast, a unitary representation of action A1 would successfully decode on any trial in which A1 was selected, irrespective of whether S1 or S2 was present. To detect such integrated representations we performed the following steps:

(1) Establishing potential ROIs: First we trained the classifier to decode S1-A1 vs S2-A2 configurations at the time of initial stimulus presentation and tested for those representations at the same time-point. A significant signal in this analysis is indicative of the encoding of unitary stimulus representations, unitary action representations, or integrated stimulus-action representations.

(2) Ruling out unitary stimulus representation: Secondly, we used the classifier weights trained up in stage (1) in order to also test for discrimination between S1-A2 vs S2-A1. If the classifier performs significantly above-chance, this would indicate that unitary stimulus information is being decoded (since the only consistent labels between the training and testing data are S1 and S2).

(3) Ruling out unitary action representation: In our third analysis, we again used the classifier weights from stage (1), and tested if the classifier could decode S2-A1 vs S1-A2. Similar logic implies that significant decoding in this analysis is consistent with unitary action representations.

It is also possible that both integrated and unitary representations are present in a region simultaneously. Significant classification in stage (1) but not in stages (2) and (3) is indicative of integrated stimulus-action representations only. Thus, in order to attribute decoding signals specifically to integrated representations, we consider the conjunction between two statistical maps obtained from per-voxel paired t-tests between (i) accuracy scores in stage

(1) vs. stage (2) and (ii) accuracy scores in stage (1) vs. stage (3). The only explanation for a signal that survives this stringent criterion is that it is generated by an integrated stimulus-action representation, since the first paired t-test rules out stimulus-only decoding and the second rules out action-only decoding.

Significance Testing

For the searchlight analyses, the percentage of correctly identified samples, averaged across folds in the cross-validation, was used as the classification score in each searchlight and this score was assigned to the voxel at the center of the searchlight sphere. This defined a classification accuracy map for each subject, which was then smoothed with an 8mm FWHM kernel. A second-level analysis was implemented by performing voxel-wise t-tests comparing the distribution of accuracies across participants against 50%, which is the expected performance of an algorithm randomly labeling samples. Since multivariate classification is susceptible to optimistic classification biases, we carried out permutation tests to validate our decoding procedure (McNamee et al., 2013).

All results were significant at FWE-adjusted $p < 0.05$ corrected for multiple comparisons by controlling the familywise error rate (FWE) with a 10 voxel extent threshold. We had strong prior hypotheses regarding action and outcome representations in posteriolateral and anterior medial striatum, and in ventromedial prefrontal cortex. Thus, in these areas, corrections were performed within small volumes defined *a priori* based on relevant functional imaging studies (see Table 2). Small volume corrections are denoted throughout by SVFWE and whole-brain corrections by FWE. For display purposes, we present overlays thresholded at $p < 0.005$ uncorrected.

Psychophysiological Interactions

BOLD time courses were extracted from regions of interest (ROIs) using SPM's "Volume of Interest" functionality correcting for an F-contrast composed of all effects of interest in the GLM. ROIs were defined as the set of voxels within a 6mm radius of seed coordinates which were independently defined based on related functional imaging studies (see Supplementary Table S2). A GLM was then constructed with three regressors in the following order: the BOLD time course from the seed region (the physiological term), an indicator regressor encoding the initial stimulus onset on each trial (the psychological term), and the corresponding interaction regressor. Once the GLMs were estimated for all participants, a second-level contrast (i.e., across participants) was specified for the interaction regressor. The resulting statistical map details the degree of coupling, modulated by the psychological regressor, between the seed region and voxels throughout the brain. It does this by measuring how much BOLD activity in the target location is accounted for by the interaction term in the GLM.

Results

Behavioral Performance

Due to the relatively simple nature of the task, and the training conducted before each experimental condition, participants were expected to perform close to optimally (defined as choosing actions associated with the high reward distribution in each condition). Consistent with this prediction, mean percentage of optimal action-selection was 90% across participants. Performance ranged from (85.4% to 97.4%), except in one participant who was an outlier in terms of having a performance level of 60.1%. The individual with the outlier performance level was nonetheless included in the fMRI analysis, as a sufficient number of trials were still available for the classification performance.

Neuroimaging Results

Goal-directed associative encoding: S-O, R-O and S-R associations.

We first tested for areas surviving individual tests for outcome information or action information during the initial stimulus period, and then finally report a conjunction across those tests, which is the key criterion for a region involved in goal-directed associative encoding (Balleine & Ostlund, 2007).

Outcome at stimulus time.

We first tested for brain regions involved in encoding outcome identity at the time of the initial stimulus, as such representations would be indicative of regions having access to the goal or outcome at the time of decision-making. Prospective representations of the predicted outcome state at the time of stimulus were identified in right dlPFC ($p < 0.05$ SVFWE, t(18)=4.55, x=60, y=20, z=34) and in central OFC ($p < 0.05$ SVFWE, t(18)=3.11, x=18, y=32, z=-20).

Outcome at action time.

We also tested for regions encoding outcome information at the time of action performance. For this we used a time-span analysis to train the classifier on outcome representations at the time of outcome delivery, and then tested at the time of action execution. We found significant signals in dlPFC (*p* < 0.05 SVFWE, t(18)=5.15, x=51, y=17, z=37), vmPFC (*p* < 0.05 SVFWE, t(18)=6.02, x=0, y=53, z=-20), central OFC (*p* < 0.05 SVFWE, t(18)=5.15, x=30, y=38, z=-11), and caudate (*p* < 0.05 SVFWE, t(18)=4.02, x=9, y=20, z=16).

Action at stimulus time.

We also expected regions involved in goal-directed control to encode action information at the time of initial stimulus presentation. Out of the regions identified above as containing outcome information at the time of either initial stimulus presentation or action execution, two regions also contained action information at the initial stimulus time, the dlPFC (*p* < 0.05 SVFWE, x=57, y=8, z=34, t(18)=3.85) and vmPFC (*p* < 0.05 SVFWE, x=0, y=53, z=-20, t(18)=6.02).

Regions containing both action and outcome information at stimulus time.

In order to formally identify voxels containing *both* outcome and action information at the time of the initial stimulus, we performed a conjunction analysis on the results of the "outcome at stimulus time" and "action at stimulus time" statistical maps. This analysis yielded significant effects only in right dlPFC (conjunction, *p* < 0.05 SVFWE, x=60, y=17, z=34, t(18)=3.47). (Figure 7a*).* Although we did not specify the dorsomedial prefrontal cortex as an a-priori region of interest, activity was also found in this region at an uncorrected threshold. Given that this region was identified as being involved in model-based RL in a previous study by Lee et al, 2014, we performed a post-hoc small volume correction using dmPFC coordinates identified in that study (x=12, y=32, z=37), which revealed a significant cluster ($p < 0.05$ SVFWE, x=21, y=35, z=40, t(18)=3.36). As this was a post-hoc inference, we refrain from discussing it further, but report it for completeness.

Regions containing BOTH action at stimulus time and outcome at action execution.

We also performed a conjunction analysis in order to pinpoint regions in which action information is available at the initial stimulus time, while outcome information is represented during action execution. This contrast revealed significant effects in the vmPFC ($p < 0.05$ SVFWE, x=3, y=53, z=-20, t(18)=5.54), as well as the left ($p < 0.05$ SVFWE, x=-42, y=26, z=49, t(18)=4.78) and right ($p < 0.05$ SVFWE, x=51, y=29, z=43, t(18)=3.97) dlPFC.

Habitual encoding of stimulus-response associations.

In order to identify brain regions that could potentially be involved in habitual action-selection we tested for areas that encoded action information at the initial stimulus time but that were *not* encoding outcome information at either the stimulus time or during action performance. Of the areas identified in the analysis testing for significant decoding of actions at the time of stimulus, two regions in particular were identified as containing action representations that did not also contain outcome representations: the posterior lateral putamen ($p < 0.05$ SVFWE, x=-27, y=-22, z=7, t(18)=3.24) (peak within same cluster x=-21, y=-19, z=7, t(18)=4.79), and the supplementary motor cortex ($p<0.05$ FWE, x=15, y=32, z=61, t(18)=7.95). In an independent follow up analysis using anatomically defined regions of interest centered on the posterior putamen and supplementary motor cortex, we tested whether these regions contained on average significantly better predictions of actions compared to outcomes at the time of stimulus. In a paired t-test we found that action representations were significantly more strongly represented than outcome representations in both these regions (see Figure 8; putamen $p=0.001$, t(18)=3.9; SMA $p=0.005$, t(18)=2.86). In addition to these paired t-tests, we performed one-sample t-tests against a random-chance accuracy score, which indicated that only action information, but not outcome information, was present in the putamen and SMA at the time of the initial stimulus presentation.

If putamen is driving motor activity during the performance of habitual actions, one would expect this area to be functionally connected to the thalamus, and the thalamus in turn to the motor cortex in the contralateral (left) hemisphere, as dictated by the anatomy of corticostriatal loops. We tested for psycho-physiological interactions (PPI) between an indicator variable for the onset of the initial stimulus and neural activity seeded at the putamen and thalamus. The putamen-based PPI resulted in a significant correlation with activity in the left thalamus (p<0.05SVFWE, x=-12, y=-19, z=-2, t(18)=4.07) and the thalamus PPI correlated significantly with activity in left premotor cortex (p<0.05SVFWE, x=-39, y=-7, z=46, t(18)=7.39). A weaker effect was also found in the ipsilateral (right) premotor cortex (p<0.05SVFWE, x=42, y=-10, z=58, t(18)=5.45). All seed coordinates used in the PPI were defined independently of results of the other analyses in this study (see Table 2).

Integrated Stimulus-Action Representations.

We also tested for integrated stimulus-action representations – encoding specific stimulus-action pairs as unique configurations (e.g., S1A1) – at the time of the initial stimulus (see Methods). Integrated stimulus-action representations were identified in the anterior dorsomedial striatum (caudate nucleus; (p < 0.05 SVFWE, x=15, y=11, z=22, t(18)=4.16) and hippocampus (p < 0.05 SVFWE, x=24, y=-1, z=-20, t(18)=4.35).

Integrated Action-outcome Representation.

We also tested for evidence of integrated action-outcome representations at the time of stimulus, using a similar approach. No significant decoding of integrated action-outcome representations was found.

Ruling out response time confounds.

There was a significant difference in response times for the two actions (two-sided paired t-test, t(18)=3.415, p=0.003); in contrast, no difference was found in response times as function of the identity of the initial stimulus (t(18)=0.561, p=0.582), or outcome state (two-sided paired t-test, t(18)=0.577, p=0.571). To ensure that response times were not confounding our results, we ran additional analyses assessing action-dependent decoding. Specifically, in these analyses, we included individual trial reaction times as a covariate of no-interest in the fMRI design matrix and re-ran all of the classification analyses involving actions as described above. We filtered out any voxel activity variance explained by trial-to-trial response time at the INITSTIM and ACTION timepoints. This was accomplished by estimating a GLM which included trial-to-trial reaction times as parametric modulators time-locked to the INITSTIM and ACTION trial events. Following GLM estimation, beta values for these parametric modulators were multiplied by the corresponding regressors and linearly subtracted from the data. All of our results remained significant after inclusion of the reaction time covariate, indicating that our classifier is not relying on differences in reaction times in order to decode action information.

Discussion

Contemporary associative theory distinguishes between habitual S-R associations and a combination of S-O, O-R, and R-O associations thought to mediate goal-directed performance (Bernard W Balleine & Ostlund, 2007). In this study, we used multivariate pattern analysis to assess whether dissociable regions of the human brain encode these distinct associative structures. Unlike previous work in humans, contrasting qualitatively different experimental conditions designed to encourage different action selection strategies, or comparing largely parameter driven value signals generated by RL algorithms, our approach sought to identify a neural implementation of the *associative content* of goal-directed versus habitual behavioral control. We found evidence for stimulus-elicited response representations but no outcome representations, indicative of habits, in the DLS (posterior putamen). Conversely, in the vmPFC, dlPFC, and anterior caudate nucleus, both response and outcome representations were present, indicative of goal-directed decision-making.

Our finding of stimulus-elicited response, but not outcome, representations in the DLS suggests that this area is especially involved in encoding S-R associations. While previous studies have found evidence that activity in this area increases over time as habits come to control behavior (E. Tricomi et al., 2009), and that activity in this region correlates with model-free value signals (Lee et al., 2014; K. Wunderlich et al., 2012), the present study illuminates the associations being encoded in the region. A previous report found that the degree of structural connectivity between the posterior putamen and the premotor cortex predicts susceptibility to habit-like “slips-of-action” (S. de Wit et al., 2012). Our connectivity analysis suggests a potential mechanism by which stimulus-response related activity in the putamen is ultimately transferred to the motor cortex via the thalamus, in order to implement habitual motor control.

Whereas habits depend on a reflexive retrieval of a previously reinforced response, goal-directed behavior involves selecting, evaluating, and initiating an action based on the probability and utility of its consequences. The “associative cybernetics theory” (Bernard W Balleine & Ostlund, 2007; Sanne de Wit & Dickinson, 2009) postulates that the retrieval of potential outcomes, of the actions that produce them, and of the values of those actions is mediated respectively by S-O, O-R and R-O associations. Critically, to allow for sensitivity to sensory-specific outcome devaluation and contingency degradation, defining features of goal-directed performance, the associations relating the probabilities and utilities of potential outcomes to the stimuli and actions that produce them must be flexible and current, suggesting a dynamic binding of features.

One area well suited for the dynamic binding of stimuli, actions, and outcomes is the dlPFC, given prior evidence for a role of this structure in working-memory and goal-directed behavior more generally (Patricia S Goldman-Rakic, 1996; Earl K Miller & Jonathan D Cohen, 2001). We found that activity in this region reflected representations of both action and outcome identities at the time of initial stimulus presentation, indicative of a key role for this region in encoding the information necessary to guide goal-directed actions at the time of decision-making. Specifically, the finding that dlPFC activity reflects information about action and outcome identities, necessary for computing goal-directed action values, is consistent with a contribution of this area to encoding the model component of a model-based RL algorithm. Previous findings reported state-prediction errors in this region that could underpin the learning of the underlying associations needed to form such a model (Glascher et al., 2010). The present findings suggest that not only is dlPFC involved in learning or updating such a model, but also in encoding (or at least retrieving) the model itself.

The contribution of dlPFC to the encoding of associative information necessary for computing goal-directed actions at the time of initial stimulus presentation can be contrasted with our findings in the vmPFC. Whereas vmPFC did encode information about the action at the time of initial stimulus presentation, information about the outcome identity was not present until later in the trial, during action execution. However, in the central orbitofrontal cortex (cOFC), an area adjacent to and highly connected to the vmPFC (Carmichael & Price, 1996), outcome identity information was represented at the time of initial stimulus presentation. One possibility, therefore, is that the cOFC encodes the identity of a goal at the time of initial decision-making and that this outcome-identity representation is then used to retrieve outcome value signals in the vmPFC. Consistent with this interpretation, a previous study by our group reported activity in central OFC extending to vmPFC correlating with the categorical identity of the goal at the time of decision-making (McNamee et al., 2013). In that previous study, information about the value of the goal was most prominently represented in vmPFC. An important feature of our experiment is that we have controlled for value (i.e., kept value constant throughout, with high and low value outcomes assigned equally often to every possible combination of stimuli and actions), to ensure that outcome identity information is not confounded with the outcome-value. Thus, we cannot test in the present design when value information about outcomes emerges in vmPFC. However, previous studies have reported such information to be present in both the vmPFC and in the dlPFC at the time of decision-making (V. S. Chib, A. Rangel, S. Shimojo, & J. P. O'Doherty, 2009; H. Plassmann, J. O'Doherty, & A. Rangel, 2007).

Our findings provide new insight into the differential functions of DLS vs DMS. While the posterior DLS (posterior putamen) was found to encode representations of responses elicited by discriminative stimuli, but not of outcomes, the anterior dorsomedial striatum (anterior caudate) was likewise found to encode response representations at the time of initial stimulus presentation but also to contain significantly decodable information about outcome identities at the time of action performance. However, there was also a difference in the type of stimulus-response coding present in the DMS compared to the DLS. In the DMS, the encoding of response associations was integrated with stimulus identity; a unique distributed representation was present in the DMS for each stimulus-response pair. In contrast, in the DLS, each response was coded independently of the stimulus that elicited it. The binding of stimulus-response associations into a single representation found in the DMS could underpin a form of abstraction of stimulus-response codes, which could potentially be part of a mechanism for chunking stimulus-response chains. Our finding of a difference in the type of encoding present in the dorsomedial vs dorsolateral striatum is important, given that a number of previous neurophysiology studies have not found clear differences in information encoding between these regions (Hoseok Kim et al., 2009; Kimchi et al., 2009; Thomas A Stalnaker et al., 2010).

Some neurophysiology studies have reported outcome representations at the single-neuron level in both DLS and DMS (Hikosaka, Sakamoto, & Usui, 1989; H. Kim et al., 2013; Hoseok Kim et al., 2009; Lau & Glimcher, 2007; Thomas A Stalnaker et al., 2010), whereas here we found such representations only in DMS. An important feature of our experimental design is that differences found in outcome identity representations could not be accounted for by potential differences in the value of the outcomes. While we did have actions leading to high vs low rewards in our experiment, we trained the classifier to distinguish between different outcome states leading to the SAME high valued reward. This is necessary because differences in outcome value, i.e., between high and low valued goal states, could drive differences in outcome-related neural activity in a brain region even if that area is not explicitly representing outcome identity; indeed, even a pure S-R learning system would discriminate high and low valued states as the high valued state would be associated with stronger S-R associations through trial-by-trial reinforcement.

Naturally, the absence of significant decoding from the BOLD signal in a given brain area does not imply the absence of that information at the level of single neurons. fMRI and single unit data may capture different aspects of neural activity in any event, with the BOLD signal suggested to be correlated more closely with input into a region and intrinsic processing therein as opposed to output (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001). Nevertheless, it is striking that our current findings about information content do accord very well with previous evidence about the differential role of dorsolateral striatum in habitual control, and a corresponding role for dorsomedial striatum and prefrontal cortex regions in goal-directed actions (Glascher et al., 2010; Valentin, Dickinson, & O'Doherty, 2007; Yin et al., 2004; Yin et al., 2005).

To conclude, our present results suggest that different brain areas are involved in encoding different kinds of information about responses and outcomes, consistent with a differential role for these regions in goal-directed and habitual learning and control. Whereas cortical areas, including the dorsolateral prefrontal cortex and the ventromedial prefrontal cortex alongside the anterior dorsomedial striatum, contained associative information about the identities of both responses and outcomes necessary for goal-directed control, the dorsolateral striatum contained only information about stimuli and responses, which would be sufficient for habitual but not goal-directed control.

Figures

Figure 6, Task structure.

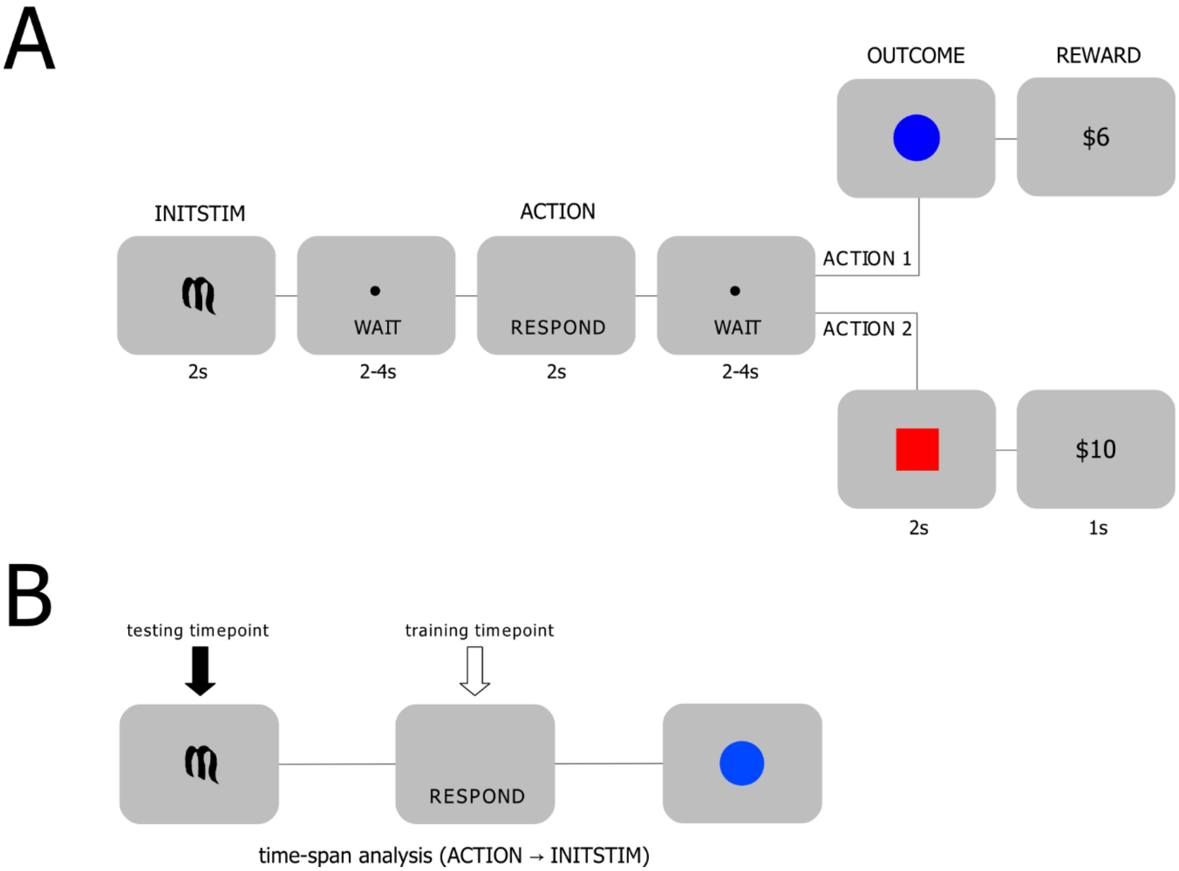


Figure 6. **a**) Task structure. Subjects performed a binary decision task. One of two possible initial stimuli (INITSTIM) was presented which determined the subsequent deterministic action-outcome contingencies between two possible actions and two possible outcome states. Outcome states were denoted by either a blue circle or a red square and were followed after a short delay by one of two distributions of monetary rewards (large or small). Crucially, each possible combination of stimulus, action, and outcome was permuted across sessions, thereby ensuring that the identity of the stimulus, or the value of the monetary reward per se, does not get conflated with action identity or outcome identity, which are the critical variables being examined in the present study. **b**) Time-span analysis. The aim of this

analysis was to identify ROIs containing “preplayed” action and outcome state representations which may be contributing to action control. In the example presented (ACTION->INITSTIM), this would require an action representation which is present at the time of action performance to be encoded at the time of the initial stimulus presentation.

Figure 7, Goal-directed, S-O, R-O, and S-R representations.

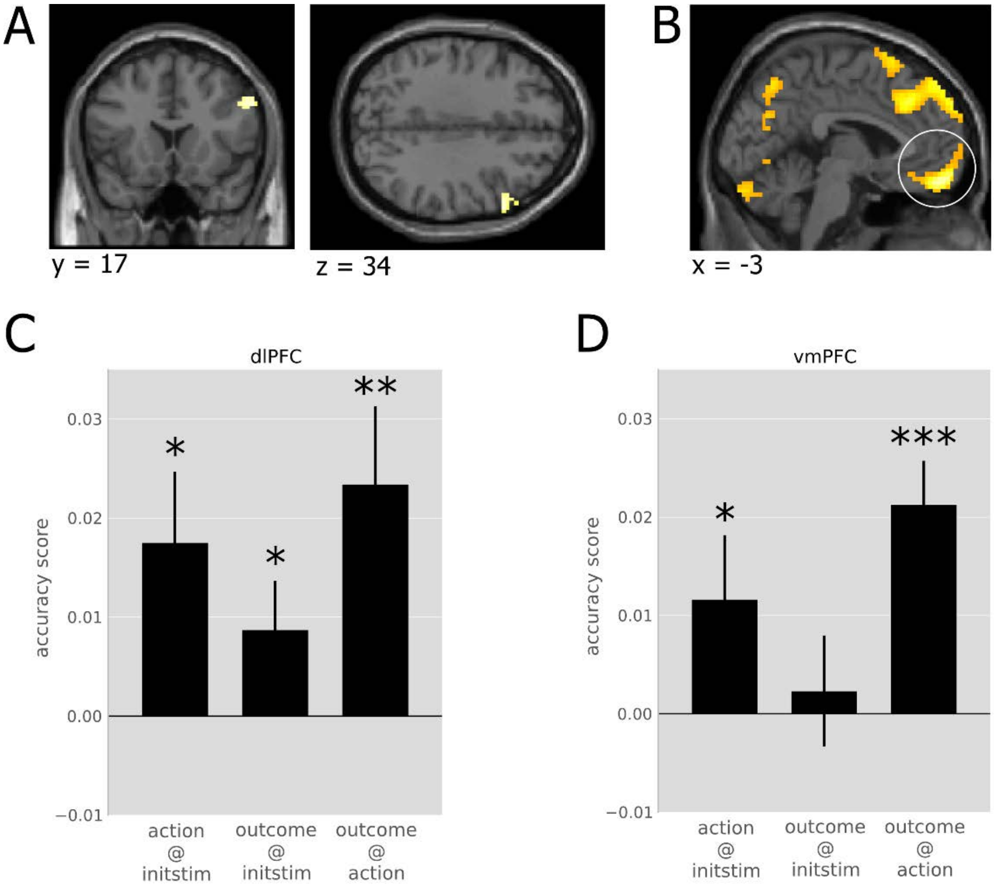


Figure 7. **a**) Right dlPFC encoded both action and outcome representations at the time of the initial stimulus presentation (conjunction analysis, x=60, y=17, z=34, t(18)=3.47). **b**) vmPFC encoded action at the time of initial stimulus presentation and outcome information at the time of action performance (conjunction, x=3, y=53, z=-20, t(18)=5.54). **c**) Bar plot depicts accuracy score distributions in an independently defined dlPFC ROI. This score is the decoding accuracy minus 0.5, which is the expected accuracy of a random algorithm. * indicates significance at p<0.05, **p<0.005. **d**) Bar plot depicts accuracy score distributions for vmPFC, ***p<0.0005.

Figure 8, Habitual, S-R associations.

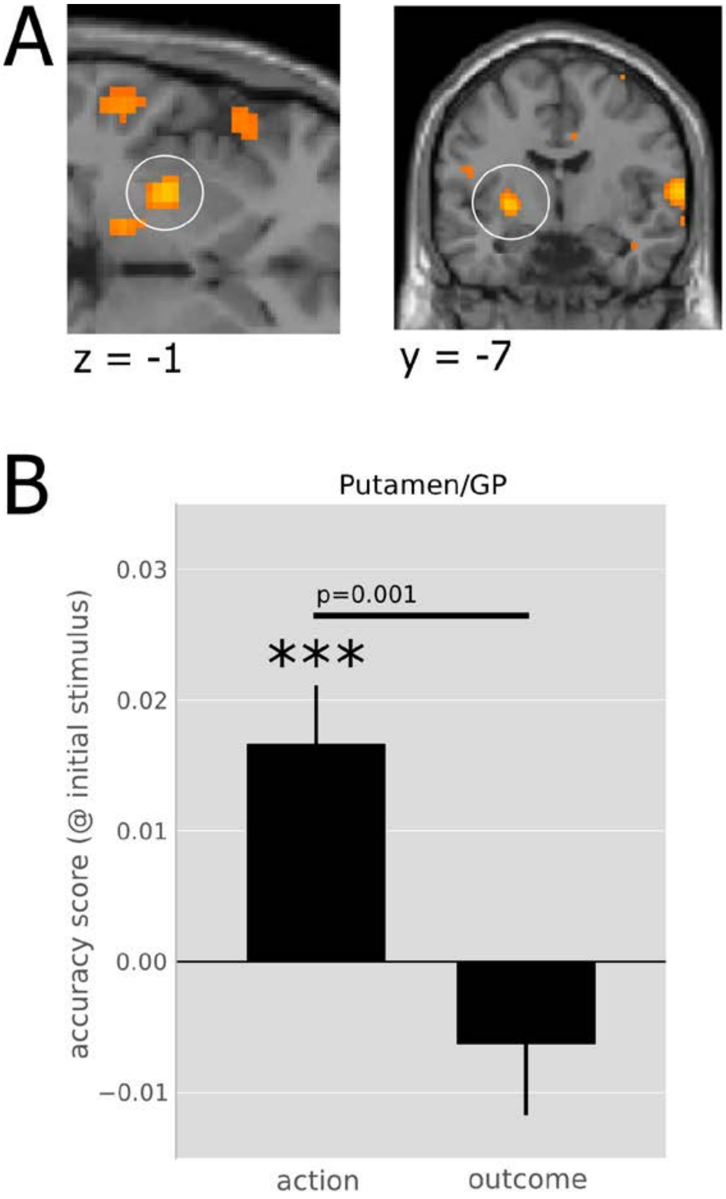


Figure 8. **a**) A region of DLS (posterior putamen), extending into the globus pallidus (GP) was found to encode information about the action to be performed at the time of initial stimulus presentation (p < 0.05 SVFWE, x=-27, y=-22, z=7, t(18)=3.24), but critically no

significant information about outcome. **b**) The distribution of accuracy scores for actions and outcomes at the time of initial stimulus in an independently defined putamen/GP region-of-interest, ***p<0.0005.

Figure 9. Integrated stimulus-action representations.

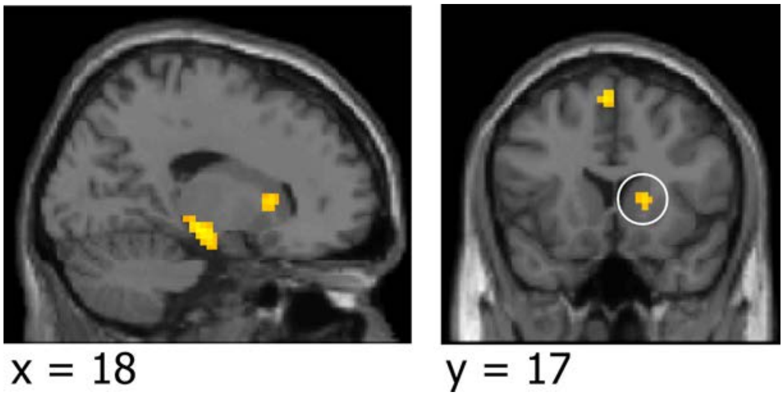


Figure 9. Integrated stimulus-action representations were localized in DMS (anterior caudate nucleus) (p<0.05SVFWE, x=15, y=11, z=22, t(18)=4.16) and hippocampus (p<0.05SVFWE, x=24, y=-1, z=-20, t(18)=4.35).

Figure 10. Behavioral performance, decision variable orthogonalization, RT/MVPA correlations.

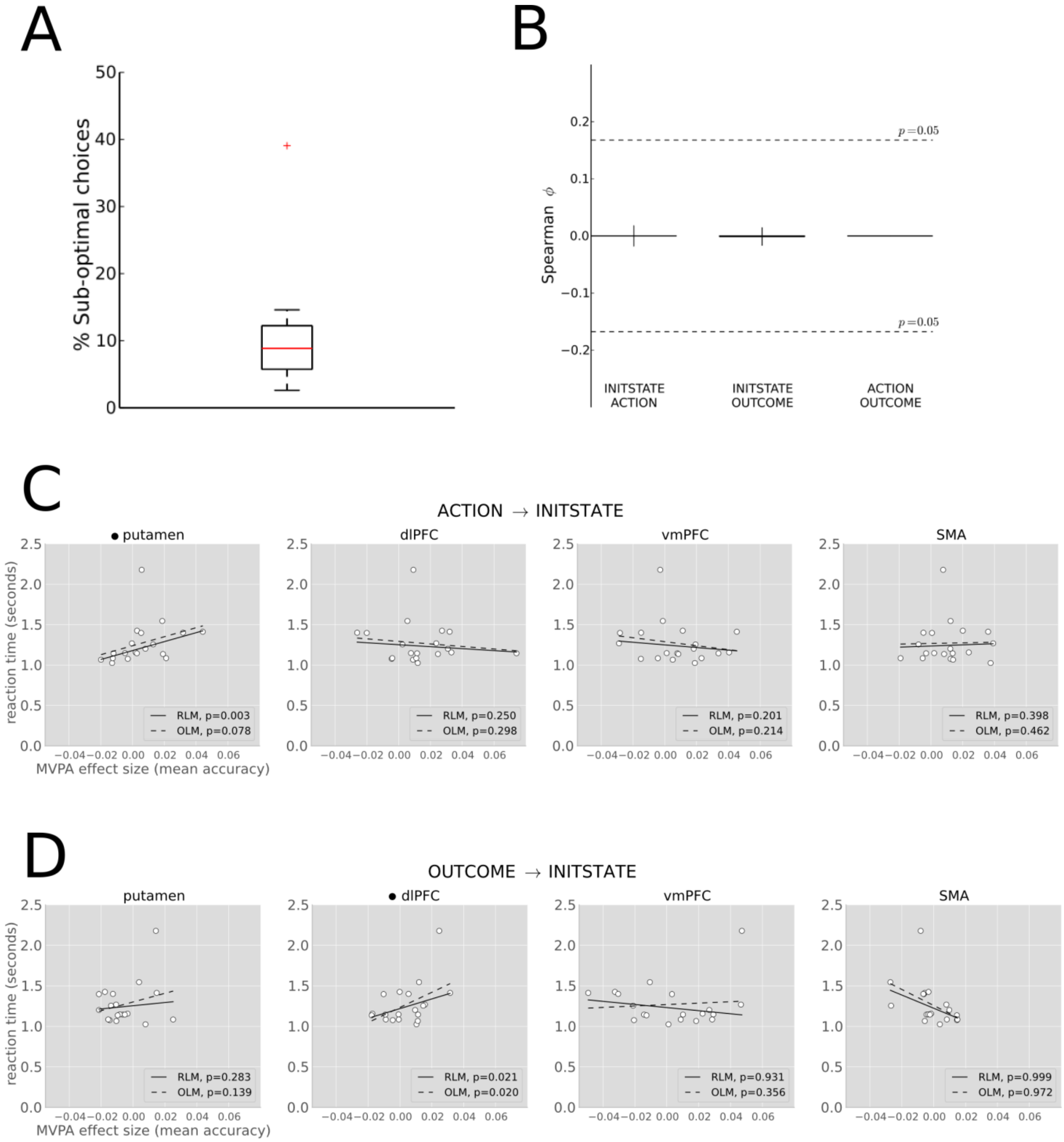


Figure 10. **a**) Percentage of sub-optimal choices, the median percentage of sub-optimal choices across subjects was 8.9% (mean 10.03, S.E.M. 1.83) corresponding to 1.42 incorrect choices per session on average (16 trials per session). One outlier subject did not respond optimally on 39.06% of trials. **b**) The average Spearman correlation between combinations of decision variables across subjects was $2x10^{-5}$, $1x10^{-3}$, and $2x10^{-4}$, respectively, indicating that the classifier could not erroneously decode one decision variable based on correlated representations of another. **c**) We performed linear regressions based on the average decoding accuracies in ROIs with significant time-span decoding of ACTION at INITSTIM. We hypothesized that the strength of the multivoxel representation would correlate with response times in ROIs which causally contribute to response selection (Picard, Matsuzaka, & Strick, 2013). RLM denotes “robust linear regression”, OLM denotes “ordinary linear regression”. Due to the discrepancy between the regression p-values for putamen we ran a bootstrap analysis in order to empirically generate regression slope distributions for significant ROIs. This analysis strongly indicated that decoding accuracies in the putamen significantly explained reaction times and confirmed the results for the other ROIs. **d**) Linear regressions between reaction times and OUTCOME time-span decoding at INITSTIM. dlPFC was the only region with a significant correlation between decoding accuracy and response time. There was also a significant anti-correlation between SMA decoding accuracy and response time. Note that all accuracy scores are averaged within an independently defined ROI and subtracted from expected performance of a randomized algorithm (0.5).

Tables

Table 1. Experimental conditions.

Condition	INITSTIM	ACTION	OUTCOME	REWARD
1	S1	A1	Red	High
	S1	A2	Blue	Low
	S2	A1	Red	Low
	S2	A2	Blue	High
2	S1	A1	Blue	Low
	S1	A2	Red	High
	S2	A1	Blue	High
	S2	A2	Red	Low
3	S1	A1	Blue	High
	S1	A2	Red	Low
	S2	A1	Blue	Low
	S2	A2	Red	High
4	S1	A1	Red	Low
	S1	A2	Blue	High
	S2	A1	Red	High
	S2	A2	Blue	Low

Table 1. Illustration of the experimental conditions describing the permutation across sessions. (S1,S2) stimuli indicating initial stimulus; (A1,A2) button press and tracker ball actions. Note that the order of presentation of the experimental conditions was also permuted across participants.

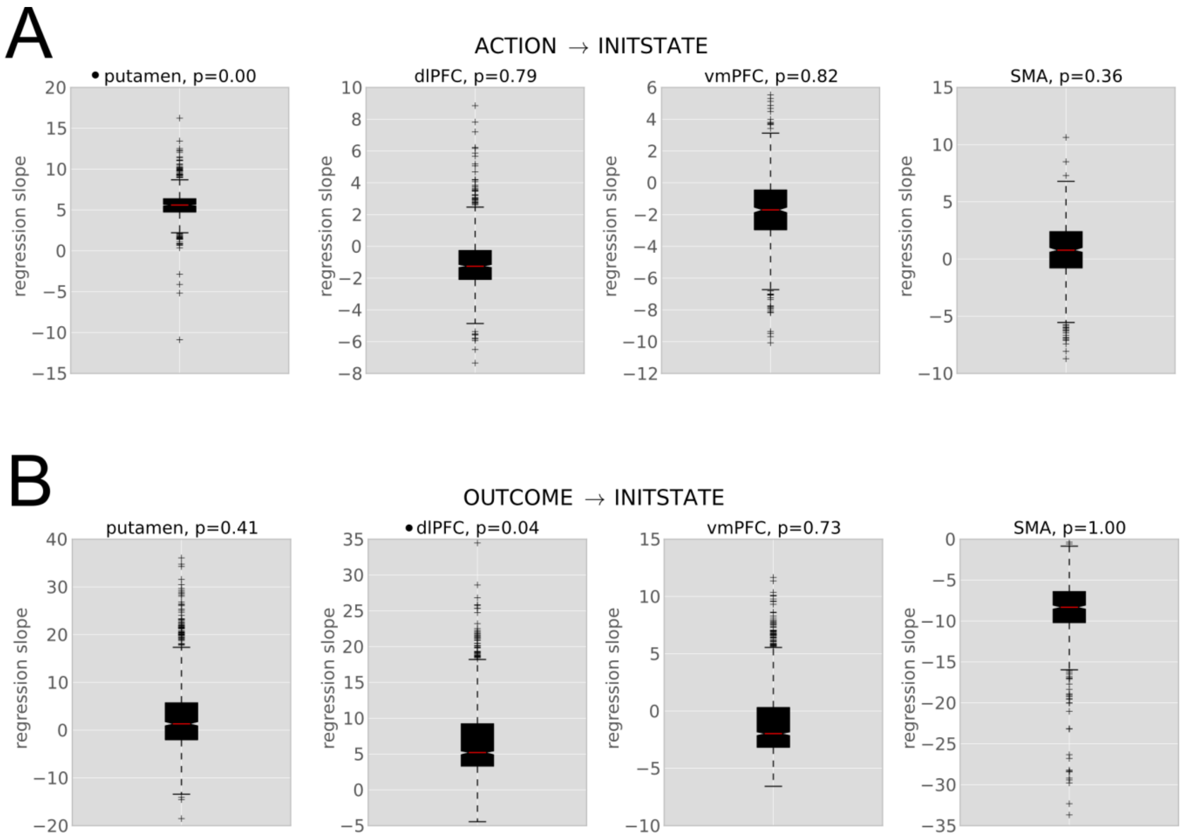
Table 2. Regions of interest (ROIs).

Region	Coordinates	Source
dlPFC	(48,9,36)	(Glascher et al., 2010)
vmPFC	(-3,41,-11)	(McNamee et al., 2013)
cOFC	(21,38,-11)	(McNamee et al., 2013)
Caudate (anterior)	(6,10,20)	(Tanaka et al., 2008)
Putamen/GP (posterior)	(-33,-24,0)	(E. Tricomi et al., 2009)
Hippocampus	(18,-6,-20) (-34,-14,-18)	(Simon & Daw, 2011)
PPI Analysis Only		
Thalamus	WFU PickAtlas mask	(Maldjian, Laurienti, Kraft, & Burdette, 2003)
Motor Cortex	WFU PickAtlas mask	(Maldjian et al., 2003)

Table 2. dlPFC, dorsolateral prefrontal cortex; vmPFC, ventromedial prefrontal cortex; cOFC, central orbitofrontal cortex; GP, globus pallidus. Correction was performed within a 10mm-radius sphere surrounding the corresponding coordinates.

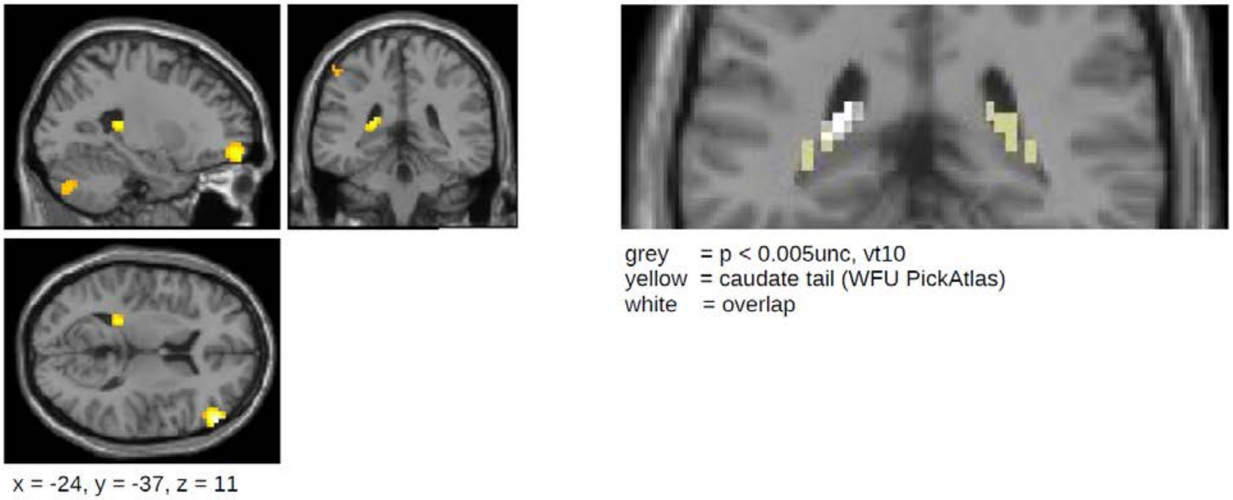
Supplementary Figures

Supplementary Figure 7. Reaction time permutation tests.



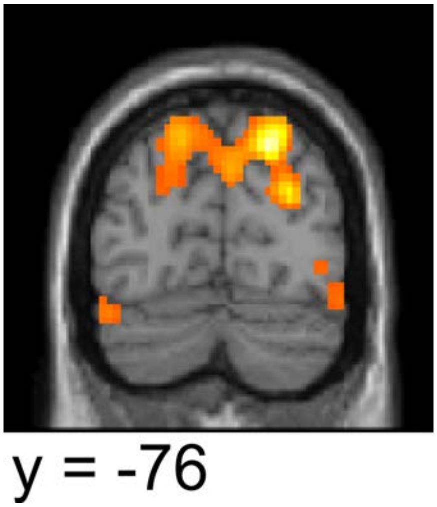
Supplementary Figure 7. In order to clarify the discrepancy between the conclusions of the robust and ordinary linear regression analyses for the putamen ROI *(*Figure 10*)*, we empirically generated a slope distributed via bootstrapping. This involved randomly drawing accuracy/RT pairs with replacement (1000 samples) and computing the corresponding regression slope. The fraction of slope values less than zero provided a one-sided non-parametric one-sided test of the significance of the regression.

Supplementary Figure 8. Decoding action at the time of outcome.


grey = p < 0.005unc, vt10
yellow = caudate tail (WFU PickAtlas)
white = overlap
x = -24, y = -37, z = 11

Supplementary Figure 8. Here, we use time-shift decoding (as opposed to time-span decoding; see Methods). Results of the ACTION@OUTCOME analysis. Action was represented in caudate tail (x=-21, y=-37, z=13, t(18)=4.39) at the time of OUTCOME but not at the ACTION or INITSTIM timepoints.

Supplementary Figure 9. INITSTIM representations in parietal cortex.



Supplementary Figure 9. INITSTIM representation in parietal cortex (x=27, y=-76, z=49, t(18) = 4.95).

Supplementary Tables

Supplementary Table 1. Sub-optimal choice logistic regression.

No. Observations	3648
Df Residuals	3641
Df Model	6
Log-Likelihood	-1150.3
LL-Null	-1188.5
LLR p-value	1.98E-14

Independent Variable	**Coef.**	**Std. Err.**	**z**	**P > |z|**	**[95.0% Conf. Int.]**	
Constant	-1.6365	0.389	-4.211	0	-2.398	-0.875
Timepoint (timepoint within a session)	0.0009	0.003	0.334	0.739	-0.005	0.006
Value (previous outcome value)	-0.0118	0.042	-0.281	0.779	-0.094	0.071
Switch (compared to previous choice)	1.1545	0.379	3.047	0.002	0.412	1.897
Switch x Value	-0.1494	0.045	-3.324	0.001	-0.238	-0.061
Timepoint x Value	-0.0005	0	-1.63	0.103	-0.001	0.000
Timepoint x Switch x Value	9.94E-05	0	0.526	0.599	-0.000	0.000

The binary dependent variable was an indicator of when a sub-optimal choice was made. The only significant coefficients of interest were those corresponding to the “Switch” and “Switch x Value” interaction. The “Switch” variable was a binary indicator of when a different choice was made compared to the previous. The “Value” variable encoded the monetary sum earned on the previous trial. Thus, the negative coefficient of the “Switch x Value” interaction variable indicates that subjects made sub-optimal choices in order to not repeat an action that had resulted in a relatively low outcome on the previous trial.

Chapter 4

MODEL-BASED SIGNALING IN THE HUMAN AMYGDALA DURING PAVLOVIAN CONDITIONING[3]

Contemporary computational accounts of instrumental-conditioning have emphasized a role for a model-based system in which values are computed with reference to a rich model of the structure of the world, and a model-free system in which values are updated without encoding such structure. Much less studied is the possibility of a similar distinction operating at the level of Pavlovian conditioning. In the present study, we scanned human participants with fMRI while they participated in a Pavlovian conditioning task with a simple structure. Fitting a model-based algorithm and a variety of model free algorithms, we found evidence at both behavioral and neural levels to support a role for a model-based as opposed to a model-free learning process in the amygdala. These findings support an important role for model-based algorithms in describing the processes underpinning Pavlovian conditioning, as well as providing evidence of a role for the human amygdala in model-based inference.

[3] Adapted with permission from (Prévost, McNamee, Jessup, Bossaerts, & O'Doherty, 2013). Data acquisition and fMRI data analysis performed by Charlotte Prévost.

Introduction

Neural computations mediating instrumental conditioning are suggested to depend on two distinct mechanisms: a “model-based” reinforcement learning system, in which the value of actions is computed on the basis of a rich knowledge of the states of the world and the nature of the transitions between states, and a “model-free” reinforcement learning system, in which action-values are updated incrementally via a reward prediction error without using a rich representation of the structure of the decision problem (Corrado & Doya, 2007; N. D. Daw, Y. Niv, & P. Dayan, 2005; P. Dayan & Daw, 2008; P. Dayan, Kakade, & Montague, 2000; Doya, Samejima, Katagiri, & Kawato, 2002; Gershman, Blei, & Niv, 2010). Accumulating evidence supports the existence of model-based representations during instrumental conditioning in a number of brain regions, including the ventromedial prefrontal cortex, striatum, and parietal cortex (N. D. Daw et al., 2011; Glascher et al., 2010; A. N. Hampton et al., 2006). However, instrumental conditioning is not the only associative learning mechanism in which model-based computations might play a role.

Pavlovian conditioning can also be framed as a model-based learning process, in which the animal begins with a model of the possible structure of the world: the stimuli within it, and sets of possible contingencies that could exist between conditioned stimuli and unconditioned stimuli, as well as assumptions about how these contingencies might change over time. In essence, learning within such a system corresponds to determining the statistical evidence for which structure out of the set of possible causal structures best describes the environment, as well as determining whether or when the relevant causal processes have changed as a function of time. Model-based approaches to classical conditioning to date have used Bayesian methods to yield inference over structure (A. C. Courville, Daw, & Touretzky, 2006).

Very little is known about the extent to which such model-based algorithms are implemented in the brain during Pavlovian conditioning. The aim of the present study was to address this question using computational fMRI. Human participants were scanned while undergoing a

Pavlovian conditioning procedure with a sufficiently complex structure to enable the predictions of model-based and model-free algorithms to be compared and contrasted (see Figure 11). We then constructed a Bayesian algorithm incorporating a model of the structure of the learning problem and compared the predictions of this algorithm against two widely adopted "model-free" algorithms for Pavlovian conditioning: the Rescorla-Wagner (RW) learning rule (Rescorla, 1972) and the Pearce-Hall (PH) learning rule (J. M. Pearce & G. Hall, 1980).

In order to test for model-based signals in the brain we focused on the amygdala, a structure heavily implicated in Pavlovian conditioning in both animal and human studies (Buchel & Dolan, 2000; Delgado, Olsson, & Phelps, 2006; Fanselow & LeDoux, 1999; Johansen, Cain, Ostroff, & LeDoux, 2011). To obtain signals from this region with sufficient fidelity, we used a high-resolution fMRI protocol in which we acquired images with more than 4 times the resolution of a standard 3mm isotropic scan, alongside an amygdala specific normalization procedure (Prevost et al., 2011). We hypothesized that the model-based algorithm would account better for both behavioral and fMRI data acquired during both the appetitive and aversive conditioning phases than would the models of Pavlovian conditioning which do not contain such structured knowledge.

Results

Behavioral results

Affective ratings for the liquid outcomes

Subjects were asked to give subjective ratings of the pleasant and neutral tasting liquids before and after the appetitive session and of the unpleasant and neutral tasting liquids before and after the aversive session. The pleasant, neutral, and unpleasant tasting liquids (unconditioned stimuli or USs) were reported to be highly pleasant, neutral, and unpleasant by subjects as indicated by their ratings averaged across before and after conditioning (Figure 12a). There was no significant difference in the pleasantness ratings of any of the liquid outcomes before and after conditioning (paired t-tests, all $p>0.05$).

Revealed preference rankings for the cue stimuli

Subjects made binary preferences between the visual cues used in the conditioning protocols before and after the experiment (Figure 12b). Subjects showed increased preference rankings for the cues displayed in the appetitive session (averaging across both CS+ and CS- cues as both were paired with reward and neutral outcomes over the course of the experiment due to the reversal) after as compared to before the experiment ($p<0.001$). Furthermore, the set of cues used in the aversive sessions showed a significant decrease in their relative preference rankings ($p<0.001$). Preference rankings for the control cues (cues not included in either the appetitive or aversive conditioning sessions) showed no significant changes from before to after the experiment. These results indicate that while the cues displayed in the appetitive session have acquired an increased positive value, those displayed in the aversive session have acquired a negative value indicating that subjects showed a modulation in their affective responses to the cue stimuli as a function of the context in which these stimuli had been conditioning (appetitive vs aversive).

Pleasantness ratings for the cue stimuli

We also obtained pleasantness ratings from subjects while in the scanner during the conditioning procedure. In the middle of the appetitive session, a few trials after a new pair of cues was presented, subjects rated the cue paired with the pleasant liquid significantly higher than the cue paired with the neutral liquid (p<0.01) (Figure 12c). Subjective ratings were obtained at the end of the appetitive session, hence following reversal of the last pair of cues, and although they still rated the cue paired with the pleasant liquid higher than the one paired with the neutral liquid, this difference was not significant. Similarly, in the aversive session, the cue paired with the unpleasant liquid was rated significantly higher than the cue paired with the neutral liquid a few trials after a novel pair of cue was presented (p<0.01) but not after a reversal had occurred (Figure 12d).

Heart rate

Participants' pulse rate (an estimation of heart rate) was monitored using a pulse oximeter for the duration of the experiment. Existing research on heart rate responses to significant stimuli has identified an initial bradycardia associated with more aversive stimuli (Libby, Lacey, & Lacey, 1973). This deceleration is thought to express attentional orienting to salient events through parasympathetic activity (Bradley, 2009). Aversive trials were associated with a more pronounced cardiac deceleration (as assessed by the number of beats) compared to appetitive trials during anticipation, in a time window of 1.5-3.5s following stimulus onset, as reported elsewhere (Nicotra, Critchley, Mathias, & Dolan, 2006) (paired t-test, p<0.01). Such physiological changes signal a more aversive emotional state for aversive as compared to appetitive trials, thereby reflecting a differential heart-rate conditioned response in the aversive relative to the appetitive conditioning trials.

Respiration

When analyzing respiration signals, we found that in the aversive condition, subjects learned to inspire before cue offset and expire at the time of the aversive liquid delivery. In contrast, subjects expired before cue offset and inspired at the time of the appetitive liquid delivery in the appetitive condition. The amplitudes between the appetitive and aversive conditions were significantly different both before cue offset (3.5s) and at the time of liquid delivery (4.5s) ($p<0.05$). However, note that these results need to be interpreted with caution because they do not survive multiple comparisons across all time windows tested.

Pupil dilation and blinking

We also recorded pupil diameter, an automatic measure of arousal previously shown to provide a measure of conditioning (Bitsios, Szabadi, & Bradshaw, 2004; Bray, Rangel, Shimojo, Balleine, & O'Doherty, 2008; Seymour, Daw, Dayan, Singer, & Dolan, 2007). We found a significantly smaller amplitude in pupil diameter for trials where the cue was predictive of the pleasant liquid (appetitive condition) as compared to trials where the cue was predictive of the neutral liquid (neutral condition) ($p<0.05$) in a time window of 0.8-1.5s after cue onset where amplitude changes in pupil diameter have previously been reported (Seymour et al., 2007) in the 10 subjects from which we obtained pupil amplitude measures (Figure 12e). A higher degree of arousal (significantly smaller peak amplitude) would have been equally expected when subjects saw cues predictive of the aversive liquid; however, reliable analysis of amplitude in pupil diameter for these trials was prevented by the prolonged blinking elicited by these aversive cues. Given that blinking is also a conditioned response, we looked for evidence of blinking in the aversive condition as opposed to the neutral condition. We found significant differences between the aversive and neutral conditions during the first second after cue onset and the last second before cue offset (paired t-tests, $p<0.05$) as well as at the time of liquid delivery and swallowing (paired t-tests, $p<0.01$).

Model comparison on behavioral data using reaction times.

We used Bayesian information criterion (BIC) to compare the model goodness of the HMM against the baseline model and the model-free algorithms on the basis of trial by trial variation in reaction times. We found that the HMM model fit better than each of the other models including the baseline model, indicating that this model was providing the best account of trial by trial variation in conditioning as reflected in reaction times. On the other hand, neither the RW nor the PH learning rules provided a better fit to the data than did the baseline model, suggesting that these algorithms cannot account for changes in reaction time as a function of conditioning any better than a random actor (Table 3 and Table 4). The normalized RT data is shown plotted against the value signal predictions of the HMM model in Figure 12f,g, indicating that RTs become slower under situations where the cue presented is associated with a stronger prediction of an aversive outcome in the aversive condition, and become faster as cues are associated with a stronger prediction of an appetitive outcome in the appetitive condition.

fMRI results

We report results from our analyses from our model-based learning algorithm (the HMM model) within the amygdala using a height threshold of $p<0.005$, with an extent threshold significant at $p<0.05$ corrected for multiple comparisons. We first report expected value signals because these signals are generated by both our model-based and model-free learning algorithms and can therefore be easily compared. However, note that finding neural evidence for precision signals is an even more critical test for model-based computations in this task, as such signals can only be generated by our model-based learning algorithm.

Expected value signals

We first investigated BOLD activity in the amygdala correlating with expected value (EV) signals at the time of cue presentation (see Figure 13 a for an illustration of EV signals). In the appetitive session, we found significant activity positively correlating with expected value in the medial part of the right amygdala, corresponding to the basolateral complex (Figure 14a in green, MNI [x y z] [10 -10 -18], T = 6.29, k=28 voxels). In the aversive session, activity positively correlating with expected value was found in the centromedial complex of the left amygdala (Figure 14a in red, [x y z] [-27 -2 -9], T = 5.63, k=44 voxels; [x y z] [-17 -15 -14], T = 5.41, k=69 voxels), such that the greater the activity in these areas, the less an aversive outcome is predicted to occur. We also looked for areas correlating negatively with EV in both the appetitive and aversive sessions, that is, areas showing an increase in activity the less a positive outcome was predicted to occur given the cue. We did not find evidence for such activity in the amygdala in either the appetitive or the aversive session at our statistical threshold.

Precision signals

Next, we examined amygdala activity correlating positively with precision or else correlating negatively with precision during both the appetitive and aversive sessions (see Figure 13b for an illustration of precision signals). While no significant negative correlation was found with precision, we did find significant correlations with precision signals during both the appetitive and aversive sessions within our centromedial complex ROI (appetitive session: [x y z] [25 -1 -10], T = 4.12, k=44; aversive session: [x y z] [27 -5 -10], T = 5.31, k=115; [x y z] [18 -2 -16], T = 4.75, k=44) (Figure 15a). To test whether there was a significant overlap between these clusters in the appetitive and aversive sessions, we performed a formal conjunction analysis (at our omnibus threshold of $p<0.005$ with a cluster extent of $p<0.05$). In this contrast we found a common area activated by precision signals in the appetitive and aversive sessions in the centromedial complex of the amygdala ([x y z] [24 -4 -9], T = 3.52, k=23) (Figure 15c).

Model comparison on BOLD data

In order to determine whether BOLD activity in the amygdala is better accounted for by the HMM than by the model-free learning algorithms, we performed a Bayesian Model Selection (BMS) analysis. The expected value contrasts from our “model-based” Hidden state Markov switching model (HMM) and the “model-free” Rescorla-Wagner (RW) and Pearce-Hall (PH) models were used to compare BOLD activity in the amygdala separately for the aversive and appetitive sessions. In this model comparison, we included voxels within a 4mm sphere centered on the peak voxels of amygdala activities correlating with either expected value signals for the ‘model-based’ HMM or expected value signals for the model-free algorithm using the leave-one-out method, thereby avoiding a non-independence bias in the voxel selection. We found that the model-based HMM outperformed both model-free algorithms with an exceedance probability of 0.94 (posterior probability = 0.64) for the aversive session and of 0.93 (posterior probability = 0.55) for the appetitive session.

We also performed a similar BMS to discriminate between our “model-based” HMM and a simpler version of this HMM which uses Bayesian updating but does so in a manner resembling a more “model-free” algorithm. The essential difference between these two HMMs is that the “model-based” HMM does not allow for a reversal without moving from a non-reversal state to a possible reversal state. Note that the expected reward signals generated by these two HMMs are highly correlated, whereas the precision values are not. Hence, we compared neural activity within the contrasts showing activity positively correlating with precision signals including voxels within a 4mm sphere centered on the peak voxels of the amygdalar activities correlating either with precision signals for the “model-based” HMM or the ‘’model-free’’ HMM using the leave-one-out method, thereby avoiding a non-independence bias in the voxel selection. We found that activity was better explained by precision signals estimated by the “model-based” HMM in both the aversive and appetitive sessions (aversive session: exceedance probability=0.99; appetitive session: exceedance probability=0.57).

Discussion

In this study, we used a Pavlovian conditioning task with a rudimentary higher-order structure in both appetitive and aversive domains to investigate whether neural activity in the human amygdala reflects learning that requires access to model-based representations. By comparing neural activity correlating with expected value signals generated by model-based versus model-free learning algorithms using a Bayesian model selection (BMS) procedure, we have been able to show that in at least some parts of the human amygdala, activity during Pavlovian conditioning is better accounted for by a model-based rather than a model-free algorithm.

One of the critical distinctions between the model-free and model-based learning algorithms in the present study is that while the expected value of a stimulus previously paired with the unpleasant outcome is still low following reversal of contingencies, because that was the value it had before reversal in a model-free system, the expected value of this stimulus will become high in a model-based system because it incorporates the knowledge that after a reversal stimulus values switch (i.e., there is full resolution of uncertainty when a reversal occurs). We have captured model-based representations in formal terms using an elementary Bayesian Hidden Markov computational model that incorporates the task structure (by encoding the inverse relationship between the cues and featuring a known probability that the contingencies will reverse).

Our behavioral analysis demonstrated that participants showed evidence of conditioned responses to the conditioned stimuli and thus successfully learnt the associations between the different cues and outcomes. In a trial-by-trial analysis in which we correlated reaction times against the model predictions, we found that the HMM model predicted changes in reaction times over time as a function of learning better than the model-free alternatives, and that indeed the model-free algorithms did not predict variation in reaction times significantly better than chance.

In the imaging data, we found trial-by-trial positive correlations of model-based expected values in an area consistent with the basolateral complex of the amygdala according to the Mai atlas in the appetitive session, and in areas in the likely vicinity of the centromedial complex in the aversive session (Mai, Paxinos, & Voss, 2008). It is interesting to note that activity in these same areas (i.e., basolateral versus centromedial complex) has been found to correlate with expected value signals generated by a simple RW model in a recent reward versus avoidance instrumental learning task (in an appetitive versus aversive context respectively) (Prevost et al., 2011). Using a BMS procedure, we found that amygdala activity correlating with expected value was better explained by model-based rather than model-free learning algorithms. Whereas the model-free system has received considerable attention in the past (J. P. O'Doherty, A. Hampton, & H. Kim, 2007), the more sophisticated and flexible model-based system has been more sparsely studied particularly in relation to its role in Pavlovian learning. Thus, our results point to the need for integrating model-based representations and their rich adaptability into our understanding of Pavlovian conditioning in general, and of the role of the amygdala in implementing this learning process in particular. Another important feature of the model-based algorithm featured in this study is that as well as keeping track of expected value, this model also keeps track of the degree of precision in the prediction of expected value over the course of learning. This precision starts off low at the beginning of a learning session with a new stimulus because the expected value computation is very uncertain at this juncture, but once outcomes are experienced in response to specific cues, the precision in the estimate quickly increases. However, this precision lessens again as the trial progresses because a reversal in the contingencies is increasingly expected to occur (hence the expected value becomes more and more uncertain). Signals correlating with precision were found to be located in the vicinity of the centromedial complex in both the appetitive and aversive sessions. Precision signals might play an important role in the directing of attentional resources toward stimuli in the environment. The presence of a precision signal in the centromedial amygdala in the present paradigm could be a key computational signal underpinning the putative role of this structure in directing attention and orienting toward affectively significant stimuli.

The presence of a precision-related signal in the amygdala during Pavlovian conditioning may relate to other findings in which the amygdala has been suggested to play a role in "associability", as implemented in a model-free algorithm such as the Pearce-Hall learning rule (Li et al., 2011; Roesch, Calu, Esber, & Schoenbaum, 2010). Associability as defined in such a model is essentially a model free computation of uncertainty, the inverse of precision: associability is maximal when the absolute value difference between expected and actual rewards is greatest. However, in our case, an associability signal is clearly distinct from the signal we observe in the amygdala in the centromedial complex (even leaving aside the fact the signal we found is negatively as opposed to positively correlated with uncertainty). First of all, because the signal in our HMM is model-based, it changes to reflect anticipated changes in task structure (such as a reversal), whereas Pearce-Hall associability does not change to reflect anticipated changes in task structure, rather, both change only reflexively once contingencies have reversed. The model-based nature of our signal was confirmed by comparing the precision signal generated by the model-based algorithm with that generated by a model-free version of our Hidden-Markov Model: activity in the amygdala was best accounted for by the precision signal generated by the model-based algorithm. Note that although this BMS comparison provided clear evidence in the aversive session, the evidence was much weaker in the appetitive session, in which case interpretation in favor of either model is difficult. However, it is interesting to note that 'aversive' precision signals in the amygdala were better accounted for by a model-based learning algorithm given the traditional view of the amygdala being associated with aversive processing, although this view has been considerably challenged in the past few years (Baxter & Murray, 2002; Murray, 2007).

Finally, we checked the correlation between the precision signal we found here and an associability signal generated by the Pearce-Hall learning rule, and we found the correlation between these signals to be essentially negligible (with r ranging from -0.06 to -0.14), as opposed to being strongly negatively or positively correlated, as would be anticipated were these signals to tap similar underlying processes.

The fact that in the present study we found model-based signals in the amygdala does indicate that this structure is capable of performing model-based inference even during Pavlovian conditioning. However, it is important to note that the findings of the present study do not rule out a role for this structure in model-free computations during Pavlovian conditioning. Indeed, while the model-free learning rules we used did not work very well in accounting for behavior on the task (as indexed by changes in reaction times), we did find some evidence (albeit weakly) of model-free value signals in the amygdala as generated by either a Rescorla-Wagner or Pearce-Hall learning rule. Indeed, while using our HMM model we did not find evidence for aversive-going expected value signals in the aversive session (i.e., by showing an increase in activity the more the unpleasant tasting liquid was expected), we did find such a signal correlating with expected value as computed by a Pearce-Hall learning rule. As a consequence, we cannot rule out a contribution for the amygdala in model-free computations. It is important to note however, that in many tasks in which neuronal activity was found in the amygdala to correlate with the predictions of model-free learning algorithms (Elliott, Newman, Longe, & William Deakin, 2004; A. N. Hampton, Adolphs, Tyszka, & O'Doherty, 2007; Prevost et al., 2011; Yacubian et al., 2006), such tasks were either not set up to discriminate the predictions of model-free versus model-based learning rules, or else the relevant model comparisons were not performed. Thus, it is entirely feasible that many of the computations found in the amygdala in previous studies correspond more closely to model-based as opposed to model-free learning signals. More generally, if indeed, both model-based and model-free signals are present in the amygdala during Pavlovian conditioning, then an important question for future research will be to address how and when these signals interact with each other.

To conclude, we have found in the present study evidence for the existence of model-based learning signals in the human amygdala during performance of a Pavlovian conditioning task with a simple task structure. These findings provide an important new perspective into the functions of the amygdala by suggesting that this structure may participate in model-based computations in which abstract knowledge of the structure of the world is taken into account when computing signals leading to the elicitation of Pavlovian conditioned responses. The findings also resonate with an emerging theme in the neurobiology of reinforcement learning whereby value signals are suggested to be computed via two mechanisms: a model-based and a model-free approach (P. Dayan & Daw, 2008; Doya et al., 2002). Whereas up to now, theoretical and experimental work on this distinction has tended to be focused on the domain of instrumental conditioning (N. D. Daw et al., 2005; Glascher et al., 2010; A. N. Hampton et al., 2006), the present study illustrates how similar principles may well apply even at the level of Pavlovian conditioning. Thus the distinction between model-based and model-free learning systems may apply at a much more general level across multiple types of associative learning in the brain. Furthermore, the present results provide evidence that model-based computations may be present not only in prefrontal cortex and striatum, but also in other brain structures such as the amygdala.

Materials and Methods

Subjects

Nineteen right-handed subjects (8 females) with a mean age of 22.21 ± 3.47 participated in the study. All subjects were free of neurological or psychiatric disorders and had normal or correct-to-normal vision. Written informed consent was obtained from all subjects, and the study was approved by the Trinity College School of Psychology research ethics committee.

Task Description

Subjects participated in a Pavlovian task where they had to learn associations between different cues (fractal images) and a pleasant (blackcurrant juice [Ribena, Glaxo-Smithkline, UK]), affectively neutral (artificial saliva made of 25mM KCl and 2.5 mM NaHCO3), or unpleasant (salty tea made of 2 black tea bags and 29g of salt per liter) flavor liquid. The task consisted of two sessions lasting approximately 22 minutes each. Each session was composed of 120 trials, leading to a total of 240 trials. In one of the sessions, subjects underwent an appetitive Pavlovian conditioning procedure whereby they were presented with cues leading to the subsequent delivery of either the pleasant flavor, or the affectively neutral one, while in the other aversive conditioning session subjects underwent an aversive conditioning procedure whereby they were presented with cues leading to the subsequent delivery of either the unpleasant flavor stimulus, or else the affectively neutral stimulus. The rationale for including the appetitive and aversive conditioning procedures in separate sessions as opposed to including both conditions intermixed within the same sessions was to avoid contrast effects observed in prior behavioral piloting whereby cues signaling the aversive outcome tended to overwhelm cues signaling the pleasant one such that both the pleasant and the neutral cue stimuli were viewed as relief stimuli (contrasted against the aversive outcome) (Seymour et al., 2005). Performing the appetitive and aversive conditioning procedures in separate sessions ensured robust behavioral conditioning in both the appetitive and aversive cases and largely avoided contrast effects between the appetitive and aversive conditions.

For both sessions, on each trial a cue was displayed randomly on either the left or right side of a fixation cross for 4 seconds. Following a well-established Pavlovian conditioning protocol (Gottfried, O'Doherty, & Dolan, 2002, 2003; J. O'Doherty et al., 2004), subjects were also instructed to indicate on which side of the screen the cue was presented by means of pressing the laterally corresponding button on a response box, yet they were also instructed that the subsequent outcomes were not contingent on their responses. This serves two purposes: it allows one to monitor the extent to which participants are paying attention to the cues on each trial, as well as offering a response time measure which can serve as an index of conditioning. The offset of the cue (after 4 secs) was followed by delivery of one of the liquid flavor stimuli with a probability of 0.6, or else no liquid stimulus was delivered. The next trial was triggered following a variable 2-11 secs inter-trial interval.

At the beginning of each session, subjects were presented with two novel fractal cues (not seen before in the course of the experiment): which we will denote as cue 1 and cue 2. In the appetitive session, cue 1 predicted the subsequent presentation of the pleasant liquid 60% of the time (or no liquid delivery 40% of the time), while in the aversive session cue 1 predicted the delivery of the aversive liquid 60% of the time (or no liquid delivery 40% of the time). cue 1 and cue 2 trials were presented in a randomly intermixed order. After 16 trials (8 trials of each type), a reversal of the cue-outcome associations was set to occur with a probability of 0.25 on each subsequent trial. The probabilistic triggering of the reversal after the 16th trial ensured that the onset of the reversal was not fully predictable by subjects. Once a reversal was triggered, cue 1 no longer predicted the appetitive or aversive outcome but instead was associated with delivery of the neutral outcome, while cue 2 now predicted the appetitive or aversive outcome. After another 16 trials (8 trials of each type) following the onset of the reversal, another event was triggered to occur with probability 0.25 on one of the subsequent trials: this time instead of a reversal, a completely novel pair of stimuli was introduced. One of these, cue 3, was now paired with the appetitive or aversive outcome, while cue 4 was now paired with the neutral outcome. These new cues were presented for a further 16 trials, and followed again after a probabilistic trigger of p=0.25 on each subsequent trial with a reversal of the associations. After the reversal, a new set of cues was introduced according to the same probabilistic rule and this was followed again by a reversal. Thus in total, 3 unique pairs of stimuli were used in each session and each of these pairs underwent a single reversal (Figure 11a,b). A completely different set of cues was used for each session, so that subjects experienced a total of 6 pairs of fractal stimuli throughout the whole experiment.

Within each session, the presentation order of the affective and neutral cue presentations was randomized throughout, with the one constraint being that the cue predicting the neutral tasting liquid delivery had to be delivered twice every four trials. This ensured that the appetitive and neutral cues, and aversive and neutral cues, were approximately evenly distributed in their presentation throughout the appetitive and aversive sessions, respectively. All fractal images were matched for luminance. The order of the sessions was counterbalanced across subjects so that half of the subjects started the experiment with the appetitive session and half of the subjects with the aversive session.

Subject Instructions

Before the conditioning session, subjects received the following task instructions:

“In each trial, an image will appear on the screen and may be followed by some liquid delivery. There are six different images per session. Each image will lead to either a pleasant, neutral, or unpleasant tasting liquid. You will have to learn these associations. However, during the experiment, this may change (or reverse), making image 1 associated with the liquid of image 2 and image 2 associated with the liquid of image 1. This reversal may actually happen more than once during the experiment and you have to fully pay attention and realize that it has happened. These cues may change during the experiment, so that you will have to learn these associations again with these new cues (which may also reverse).

At the beginning of each trial, the image will either appear on the left or right side of the screen. You will have to press the left button of the response pad if the image appears on the left side, or the right button if it appears on the right side. It is important that you press the button because we need to record your response times, although the trial will carry on if you don't press any button.

At the beginning and end of each session, we will ask you to rate different images and liquids. You will also have to rate these images in the middle of each session."

Apparatus

The pleasant, neutral, and unpleasant tasting liquids were delivered by means of three separate electronic syringe pumps positioned in the scanner control room. These pumps pushed 1 mL of liquid to the subject's mouth via ~10 m long polyethylene plastic tubes, the other end of which were held between the subject's lips like a straw, while they lay supine in the scanner.

Behavioral Measures

Affective evaluations of the fractal images and liquids

Participants were asked to provide subjective ratings indicating their perceived subjective hedonic evaluation for each of the 6 pairs of fractal images that were displayed. This was done during the experiment before each session, in the middle of each session (during the scanning), and at the end of each session by presenting a picture of the fractal alongside an instruction to rate the fractal for its pleasantness on a scale going from 1 (do not like at all) to 4 (strongly like). These ratings could therefore provide a behavioral measure of evaluative conditioning (Bray et al., 2008) at three different time points throughout the experiment. Furthermore, before and after the appetitive session, the pleasant and neutral liquids were rated for their subjective pleasantness using a scale ranging from -5 (very unpleasant) to +5 (very pleasant), and similarly the aversive and neutral liquids were rated before and after the aversive session.

Preference ranking test

Before the experiment started and after the experiment was over, participants were asked to make binary choices indicating their relative preferences for each of 16 different fractals (12 of which were included in the experiment; 6 each in the appetitive and aversive sessions respectively, while 4 of the fractals were not featured in either session). Each of the 16 fractals was paired with each other fractal. This test allowed us to estimate a preference ranking for each of the fractals, thereby potentially providing an additional and even more direct behavioral metric of evaluative conditioning beyond the pleasantness ratings.

Pupillary dilation

Pupil diameter was continuously measured during scanning using an MRI compatible integrated goggle and infrared eye tracking system (NordicNeuroLab AS, Bergen, Norway). Pupil reflex amplitude has been shown to be modulated by arousal level and can therefore be used as a physiological index of conditioning (Bitsios et al., 2004; Bray et al., 2008; Seymour et al., 2007). Pupil measurements could not be taken from 9 participants because space constraints within the head-coil alongside variations in head size meant that in some individuals the eye-tracker could not fit them comfortably.

Fluctuations in respiration and heart rate

Estimates of heart rate and respiration were recorded using a pulse oximeter positioned on the forefinger of subjects' left hand and a pressure sensor placed on the umbilical region. The time courses derived from these measures were used as a further physiological index of conditioning as well as being used separately to remove physiological noise from the fMRI data analysis (see fMRI data analysis).

Data Acquisition

Functional imaging was performed on a 3T Philips scanner equipped with an 8-channel SENSE (sensitivity encoding) head coil. Since the focus of our study was on the amygdala, we only acquired partial T2*-weighted images centered to include the amygdala while subjects were performing the task. These images also encompassed the ventral part of the prefrontal cortex, the ventral striatum, the insula, the hippocampus, the ventral part of the occipital lobe and the upper part of the cerebellum (amongst other regions). Nineteen contiguous sequential ascending slices of echo-planar T2*-weighted images were acquired in each volume, with a slice thickness of 2.2 mm and a 0.3 mm gap between slices (in-plane resolution: 1.58 x 1.63 mm; repetition time (TR): 2000 ms; echo time (TE): 30 ms; field of view: 196 x 196 x 47.2 mm; matrix: 128 x 128). A whole-brain high-resolution T1-weighted structural scan (voxel size: 0.9 x 0.9 x 0.9 mm) and three whole-brain T2*-weighted images were also acquired for each subject. To address the problem of spatial EPI distortions, which are particularly prominent in the medial temporal lobe (MTL) and especially in the amygdala, we also acquired gradient field maps. To provide a measure of swallowing motion, a motion-sensitive inductive coil was attached to the subjects' throat using a Velcro strap. The time course derived from this measure was used as a regressor of no interest in the fMRI data analysis. Finally, to account for the effects of physiological noise in the fMRI data, subjects' cardiac and respiratory signals were recorded with a pulse oximeter and a pressure sensor placed on the umbilical region and further removed from time-series images. We discarded the first 3 volumes before data processing and statistical analysis to compensate for the T1 saturation effects.

Preprocessing

All EPI volumes (partial scans acquired while subjects were performing the task and the three whole-brain functional scans acquired prior to the experiment) were corrected for differences in slice acquisition and spatially realigned. The mean whole-brain EPI was co-registered with the T1-weighted structural image, and subsequently, all the 'partial' volumes were co-registered with the registered mean whole-brain EPI image. 'Partial' volumes were then unwrapped using the gradient field maps. After the structural scan was normalized to a standard T1 template, the same transformation was applied to all the 'partial' volumes with a resampled voxel size of 0.9x0.9x0.9 mm. In order to maximize the spatial resolution of our data, no spatial smoothing kernel was applied to the data. These preprocessing steps were performed using the statistical parametric mapping software SPM5 (Wellcome Department of Imaging Neuroscience, London, UK).

Amygdalae Segmentation

Amygdalae Regions of Interest (ROIs) were manually segmented for each subject by a single observer using a pen tablet (Wacom Intuos3 Graphics Tablet) in FSL View (FSL 4.1.2). This program allows magnification and the simultaneous viewing of volumes in coronal, sagittal, and horizontal orientations. Amygdalae were manually outlined on each coronal image containing the amygdala using detailed tracing guidelines based on the Atlas of the Human Brain (Mai et al., 2008). Outlines were checked in horizontal and sagittal planes when they proved more valuable for the identification of structure boundaries. The anterior limit of the amygdala was defined using the horizontal and sagittal planes. The following guidelines were used: in its rostral part, the amygdala is bordered ventromedially by the entorhinal cortex, ventrally by the temporal horn of the lateral ventricle and subamygdaloid white matter, and laterally by white matter of the temporal lobe. Midrostrocaudally, the amygdala increases in size and is bordered ventromedially by a thin tract of white matter separating the amygdala and the entorhinal cortex, laterally by the white matter of the temporal lobe, and medially by the semiannular sulcus. Caudally, the amygdala is bordered dorsally by the substantia innominata and fibers of the anterior commissure, laterally by the putamen, ventrally by the temporal horn of the lateral ventricle and the alveus of the hippocampus, and medially by the optic tract.

Amygdalae Normalization

Because structures in the MTL exhibit significant inter-individual anatomic variability, the signal-to-noise ratio in group analyses is substantially limited in this area (Insausti et al., 1998). Atlas-based approaches used to register whole-brain EPI images across subjects (such as SPM) look for a global optimum alignment which is achieved under the limitations imposed by the available degrees of freedom, and which is at the expense of regional accuracy. Consequently, BOLD signals in the MTL may be underestimated or possibly missed (Miller, Beg, Ceritoglu, & Stark, 2005). Alignment of the MTL is substantially improved by a ROI-alignment (ROI-AL) approach, where segmentations of regions of interest (ROIs) are drawn on structural images and aligned directly, resulting in an increased statistical power (Yassa & Stark, 2009). The last iteration of this alignment tool is ROI-Demons, which has proven to be exceptionally accurate in the alignment of hippocampal subfields, for instance (http://darwin.bio.uci.edu/~cestark/roial/roial.html). Thirion's original demons algorithm has been implemented by Vercauteren and enforces smooth deformations by operating on a diffeomorphic space of displacement fields (Thirion, 1998; Vercauteren, Pennec, Perchant, & Ayache, 2007). Here, we used the implementation of ROI-Demons in the DemonsRegistration command-line tool (http://www.insight-journal.org/browse/publication/154). Our segmented amygdalae ROIs were registered with our amygdalae template based on 20 subjects from a previous study (Prevost et al., 2011) to serve as an initial model and to align all amygdalae using DemonsRegistration. The resulting registered amygdalae were then averaged in SPM5 (using ImCalc) to create a first model. Subsequently, the initial non-registered amygdalae were registered with this first model and the newly registered amygdalae were averaged to create a second model. We repeated the last two steps three more times in order to generate a more accurate model. We finally registered our initial amygdalae ROIs with the fifth model to generate the resulting displacement fields (or transformation calculations). These individual displacement fields were then applied to each subject's normalized EPI scans in order to specifically normalize their amygdalae to our template amygdalae. We applied the same transformation to each subject's structural scan before averaging all the aligned structural scans, to create an amygdalae-aligned average structural brain of our 19 subjects. Finally, amygdalar subdivisions were hand-drawn on our template amygdalae using the Atlas of the Human Brain (Mai et al., 2008). We delineated three sub-areas within the amygdala: the basolateral complex comprised of the basomedial, basolateral and lateral nuclei; the centromedial complex comprised of the central and medial nuclei; and the cortical complex (or cortical nucleus). In its most rostral part, the amygdala is exclusively composed of the basolateral complex. The cortical nucleus appears in the dorso-medial part of mid-rostral amygdala. The centromedial complex appears slightly more caudally than the cortical nucleus in the most dorsal part of the amygdala. The basolateral complex increases in size as one moves caudally from the anterior amygdala, has its maximal size midrostrocaudaully, and then decreases as one moves further back toward the caudal amygdala, whereas the cortical nucleus and centromedial complex slightly enlarge midrostrocaudally, but do not decrease in size as one moves further caudally within the amygdala. The cortical nucleus ends midcaudally, the basolateral complex ends in caudal amygdala, while the centromedial complex ends in the most caudal part of amygdala.

Computational Model analysis

To test whether amygdala activity was better explained by model-based or model-free learning algorithms, we correlated brain activity in this region with expected value signals estimated by a number of different computational models. In model-free learning algorithms, the agent is surprised when a reversal occurs and starts learning again after it happens, whereas in model-based learning algorithms, the agent expects the reversal and considers it as resolution of uncertainty and does not need to relearn. The two modes of learning are diametrically opposed in the current task, therefore allowing us to test whether amygdala is tracking model-based or model-free computations.

Model-based Learning Algorithms

HMM with dynamic expectation of change

For the model-based learning algorithm, we used a Hidden Markov Model (HMM). In this HMM, the inferred state of the environment is defined in terms of an association between cues and outcomes and is represented by the psychological variable S.There are three possible liquid outcomes in the experiment (pleasant and neutral in the appetitive session and unpleasant and neutral in the aversive session) and two cues on any given trial. The state values S_t are the possible combinations of cues and outcomes, for example $S_t = (cue\ 2, neutral\ liquid)$. Although the subjects were unaware that pleasant and unpleasant outcomes could not be delivered concurrently, this possible state value was omitted since it did not affect the results of the analyses. We also incorporated a binary-valued variable H in this HMM. The values of this hidden node determine whether ($H = 1$) or not ($H = 0$) the subject is expecting a reversal. A third random variable O represents the observed cue-outcome combination (see Figure 11d for a simple graphical representation of the model). The transition probabilities of the reversal variable H are

$$P(H_t|H_{t-1}) \ = \begin{pmatrix} 1-\alpha & \alpha \\ 0 & 1 \end{pmatrix}$$

Variable values are enumerated along the row and column axes. Each entry of the matrix represents the probability of moving from one value on trial $t-1$ (rows) to another on trial t (columns). At position (1,2), the α parameter is the probability of moving to the state of expecting a reversal ($H = 1$) from the $H = 0$ state. Once a subject begins expecting a reversal, they do not switch back. This is encoded in the asymmetry of the transition matrix. The time evolution of H represents a subject's growing expectation of a reversal in the cue-outcome association. After the presentation of a novel pair of cues, H is set to the zero state. The transitions for the state variable S are conditionally dependent on the reversal variable:

$$P(S_t|S_{t-1}, H_t) = \begin{pmatrix} 1-\beta & \beta \\ \beta & 1-\beta \end{pmatrix}$$

State reversals are inferred with a non-zero probability β when H is in the reversal expectation state ($H_t = 1$), otherwise $\beta = 0$ and $P\ (S_t|S_{t-1}, H_t = 0)$ is the identity matrix. Note that after the first trial following the presentation of novel cues, the subject has a nonzero probability of being in the reversal expectation state, thus they are always expecting a reversal to some degree and are prepared to react to an observation indicative of a contingency reversal. The posterior probability distribution $P\ (S_t)$ over the state values on trial t is determined by the prior state probability distribution $P\ (S_{t-1})$, the cue-outcome observation O_t, and the state transition probabilities:

$$\boldsymbol{Prior}(S_t) = \sum_{S_{t-1}\ states}\ \sum_{H_t\ states} P(S_t|S_{t-1}, H_t)\, P\,(H_t)\boldsymbol{Posterior}\,(S_{t-1})$$

$$\boldsymbol{Posterior}(S_t) = \frac{P(O_t|S_t)\boldsymbol{Prior}\,(S_t)}{\sum_{S_t states} P(O_t|S_t)\boldsymbol{Prior}\,(S_t)}$$

The prior over the state values at the beginning of a new set of cues is uniform. Beliefs are updated based on the likelihood of observing an outcome for a given cue and assuming a state such as “cue j is rewarding and this is likely to reverse soon”. For instance, if no reward is observed for cue j, then this state is given less credence because the likelihood that this occurs is low (0.4), and the expectation of reward for cue j is decreased. Significantly, expectations for the other cue are updated simultaneously, even if it is not implicated in the current trial. This is because a lower chance for the state “cue j is rewarding and this is likely to reverse soon” implies that the state “the other cue is rewarding and this is unlikely to reverse soon,” is more likely, and hence, the mathematical expectation of the reward upon presentation of the other cue increases.

The expected reward Q_j when presented with a given cue j is

$$Q_j(t) = E[R|cue\ j, trial\ t] = \sum_{R\ rewards}\ \sum_{S_t states} R\ P\ (R|S_t, cue\ c)P\ (S_t)$$

The reward R takes the values -1, 0, 1 for unpleasant, neutral, and pleasant rewards, respectively. Here, “E” denotes the mathematical expectation operator. This means that the forecast is correct on average for all possible outcomes given a specific history of past rewards for both cues.

Confidence in, or precision about, the identity of the current state can be measured by the extent to which there are differences in the posterior probabilities of the possible states given past experience and the cues presented. When these differences are high, one posterior probability is necessarily high, and hence, precision is high. Conversely, if all posterior probabilities are the same, precision is lowest. We measure precision on a given trial t using the inverse Shannon entropy of the posterior distribution of the state variable S:

$$\boldsymbol{Entropy}(S_t) = -\sum_{S_t states} P\ (S_t) \log P(S_t)$$

As more and more trials with no reward are experienced, the H node inputs a growing uncertainty about the identity of the current state into the HMM (since a reversal may have occurred in the absence of a rewarding outcome). Every time a new pair of cues is presented, precision is low but increases dramatically when the agent knows what particular state they are in (i.e., what the cue-liquid association is). Precision lowers again until the agent knows that a reversal has occurred, after which precision increases again. A random effects Bayesian approach was used for parameter fitting and model comparisons (note that we excluded one subject who failed to make motor responses from this analysis). Model parameters (such as α and β) were fixed *a priori* and the model fits were not sensitive to the specific values of these parameters. HMM estimation was performed via forward smoothing using the HMM toolbox for MATLAB (http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html).

Model-Free Learning Algorithms

Rescorla Wagner model.

In the Rescorla Wagner (RW) model, the new expected value at trial t + 1 for a given cue is based on the sum of the current expected value and the prediction error between the reward obtained and the expected value at time t, weighted by the learning rate (Rescorla, 1972):

$$Q_j(t+1) = Q_j(t) + \ \alpha \bullet (R(t) - Qj(t))$$

When j is a given cue, α is the learning rate with a range $0 \leq \alpha \leq 1$, and R (t) is the reward received on the current trial. If the valenced (pleasant or unpleasant) liquid was obtained on the current trial, R (t) = 1, else R (t) = 0. Hence there is one free parameter in this model, α. Note that using a random effects approach, we found that the optimal free parameters in the appetitive and aversive sessions averaged across subjects were 0.54 (SEM=0.09) and 0.18 (SEM=0.05) respectively.

Pearce Hall model.

This model differs from the Rescorla Wagner model (RW) in that it introduces an associability component and allows the effectiveness of the reinforcer to remain constant throughout conditioning. The associability values estimated by this model will decrease as the consequences of the conditioned stimulus become accurately predicted (J. M. Pearce & G. Hall, 1980). The expected values Q (t) of a given cue were updated according to:

$$Q_j(t+1) = Q_j(t) + S \bullet |R(t-1) - Qj(t-1)| \bullet R\ (t)$$

When j is a given cue, S is a free parameter governing the intensity of the CS, and R (t) is the reward received on the current trial. If the valenced (pleasant or unpleasant) liquid was obtained on the current trial, R (t) = 1, else R (t) = 0. In the Pearce Hall model (PH), the new expected value at trial t + 1 for a given cue is based on the sum of the current expected value and the product of the absolute value of the difference between the outcome obtained on the previous trial and the expected reward on the previous trial, and the outcome obtained on the current trial; this product is weighted by the free parameter. Hence there is one free parameter in this model, *S*. Note that using a random effects approach, we found that the optimal free parameters in the appetitive and aversive sessions averaged across subjects were 0.58 (SEM=0.09) and 0.40 (SEM=0.10), respectively.

In addition to the Rescorla-Wagner and Pearce-Hall models, we also tested a hybrid model introduced by Li et al., (2011), in which the Rescorla-Wagner rule is used to update value expectations, while the Pearce-Hall rule is used to set the learning rate. However, this hybrid model performed similarly to the Rescorla-Wagner and Pearce-Hall rules alone in terms of model-fits to the behavioral data and performed markedly worse than the HMM. Consequently, we do not consider this model further.

HMM Model with Static Expectations of Change

In order to further test whether amygdala activity is tracking precision signals from a model-based algorithm as opposed to more generically tracking precision signals computed in a model-free manner, we used a simpler version of the HMM described above, where precision signals resemble more closely what a "model-free" algorithm would estimate. In this version of the HMM, H is always set to the $H = 1$ state and thus the chance of a reversal happening is constant over time. As a result, in this HMM, precision starts low every time a new pair of cues is presented and increases substantially when the agent knows in which state they are, but because the chance of a reversal occurring does not increase over time, the precision remains high through the rest of the learning with that cue until a new pair of cues is introduced. In other words, there is no decrease in precision related to the anticipation of a change in the contingencies (which would come from having a model of when the contingencies are predicted to reverse), but instead a decrease in precision occurs only once a contingency change has occurred and been detected through trial and error experience (hence the algorithm is essentially model-free). Although the precision signals generated by our "model-based" and "model-free" HMM are very different, the expected reward signals from both signals are strongly correlated.

Baseline Model

Our baseline model simply assumes that rewards occur completely at random and no learning takes place. Hence, expected values for all trials are kept at a constant value of 0.5.

Model Comparison on Behavioral Data

To perform a formal model comparison on the behavioral conditioning data, we used the trial-by-trial reaction time data (measuring the length of time taken on each trial for participants to press a button to indicate which side of the screen the Pavlovian cue stimulus had been presented). Many previous studies have shown that changes in RTs to a Pavlovian cue are correlated with changes in associative encoding between cues and behaviorally significant outcomes (Bray & O'Doherty, 2007; Gottfried et al., 2003; Li et al., 2011; J.

O'Doherty et al., 2004). For each session separately, we log transformed and adjusted the RT data to account for a linear trend in RTs over time independently of trial type, as well as to remove the effects of changes in reaction time related to switching responses from one side of the screen to the other. This was done by regressing the log transformed RTs against a matrix containing a column of ones, a column accounting for the linear trend over time, and a column indicating whether participants switched their response from left to right or vice versa between the current and previous trial using the function regress in Matlab.

Using the same function, we then regressed these adjusted response times against the expected values generated by our ‘model-based’ HMM, our ‘model-free’ RW and PH models, and our baseline model. (For the baseline model, a small amount of noise was added to each expected value in order to compute the regression; without any noise the regression would not be calculable). This second regression analysis was run for each of these models, and cycled through all the possible learning rate parameters for the RW model, and CS intensity parameters for the Pearce-Hall model between 0 and 1, with increments of 0.001. This method returned Sum Squared Error (SSE) values for each of these parameter values, thereby allowing us to obtain the best fitting value for the free parameter for the appetitive and aversive sessions (i.e., the free parameter associated with the lowest SSE value). In order to compare the model goodness between these four different models, we converted the best SSE value of each session (appetitive and aversive) and each model into a Bayesian information criterion (BIC) value. The BIC adds a penalty proportional to the number of additional free parameters to the SSE value of each model, depending also on the number of degrees of freedom, which in this case is the total number of trials per session across all subjects (Schwarz, 1978). Using this procedure, we found that in both the appetitive and aversive sessions, the ‘model-based’ HMM outperformed the baseline model, and the baseline model outperformed both the ‘model-free’ RW and PH models (Table 3 and Table 4). Therefore, the ‘model-based’ HMM best fit our behavioral data, whereas the best fitting RW and PH models did not fit our behavioral data better than a random model. Hence, unlike

RW and PH, the ‘model-based’ HMM predicted RTs better than chance performance. Note that we did not regress the expected values generated by our simple HMM since they were highly correlated with that of our ‘model-based’ HMM.

fMRI Data Analysis

The event-related fMRI data were analyzed by constructing sets of δ (stick) functions at the time of cue presentation and at the time of outcome for the appetitive and aversive sessions. For our main GLM (illustrated in Figure 14 and Figure 15), additional regressors were constructed by using the expected values and the precision values generated by the model-based HMM as modulating parameters at the time of cue presentation. In order to compare model-based versus model-free learning algorithms in the amygdala, we ran three additional GLMs. For RW, the regressors were similar to our model-based HMM except that we did not have a regressor for precision which is not estimated by RW, and we added a modulating parameter for prediction error at the time of outcome. The regressors used in the GLM computed using PH model were the same as the ones used in our model-based HMM, except that the precision modulating parameter was replaced with an associability modulating parameter at the time of cue presentation. Finally, we ran a “model-free” HMM GLM using the same regressors as for our model-based HMM. All of these regressors were convolved with a canonical hemodynamic response function (HRF). The six scan-to-scan motion parameters derived from the affine part of the realignment procedure were included as regressors of no interest to account for residual motion effects. To account for motion of the subjects’ throat during swallowing, we added a regressor of no interest for swallowing motion. Finally, we also included 13 additional regressors to account for physiological fluctuations (4 related to heart rate, 9 related to respiration) which were estimated using the RETROICOR algorithm (Glover, Li, & Ress, 2000). 6 of the 38 (2 sessions x 19 subjects) log files could not be used to estimate these regressors due to a technical problem during data collection, and the missing physiological regressors were simply omitted for those sessions.

All of these regressors were entered into a general linear model and fitted to each subject individually using SPM5. The resulting parameter estimates for regressors of interest were then entered into second-level one sample t-tests to generate the random-effects level statistics used to obtain the results shown in Figure 14 and Figure 15. All reported fMRI statistics and p values arise from group random-effects analyses. We present our statistical maps at a threshold of $p < 0.005$, corrected for multiple comparisons at $p < 0.05$. To correct for multiple comparisons, we first used the 3dFWHMx function in AFNI to estimate the intrinsic smoothness of our data, within the area defined by a mask corresponding to our amygdala template. We then used the AlphaSim function in AFNI to estimate via Monte Carlo simulation an extent threshold for statistical significance that was corrected for multiple comparisons at $p < 0.05$ for a height threshold of $p < 0.005$ within the amygdala ROI.

Model Comparison on BOLD Data

In order to test whether the amygdala acts according to model-based or model-free learning algorithms, we used a Bayesian model selection procedure (BMS) to test which expected value signals estimated by model-based versus model-free learning algorithms better accounted for amygdala activity (Stephan, Penny, Daunizeau, Moran, & Friston, 2009). For both the appetitive and aversive sessions, we included in this model comparison individual betas averaged across voxels within a 4mm sphere centered on the peak voxels of the amygdalar activities correlating with either expected value signals for the HMM or the model-free algorithm using the leave-one out method, thereby avoiding a non-independence bias in the voxel selection (N. Kriegeskorte, W. K. Simmons, P. S. Bellgowan, & C. I. Baker, 2009). Using the spm_BMS function in SPM8, we compared expected value signals across all model-based (HMM) and model-free models separately for the appetitive and aversive sessions.

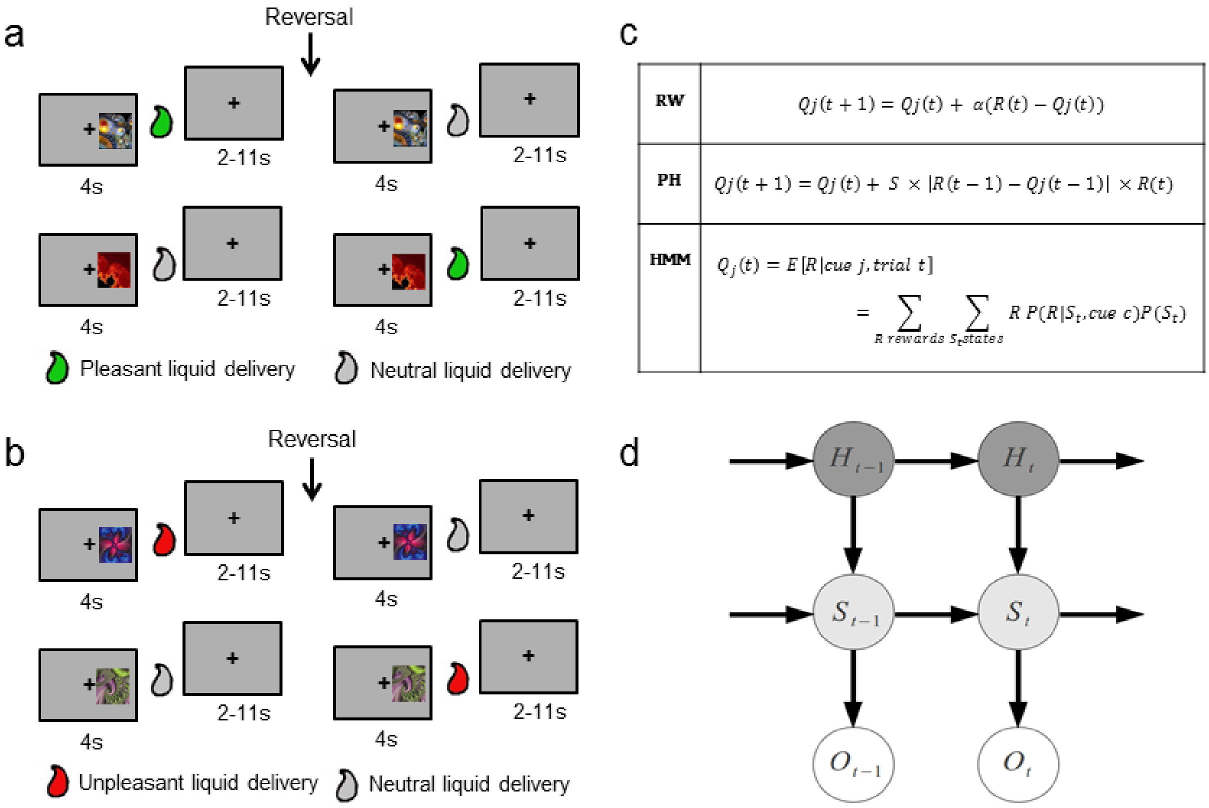
We used a similar approach to compare neural activity pertaining to precision signals estimated by our “model-based” and “model-free” HMMs. The difference between these two HMMs is that the “model-based” HMM does not allow for a reversal without moving from a “non-reversal state” to a “possible reversal state”. As a consequence, the precision values generated by these models are clearly distinguishable and thus easily comparable using a BMS (whereas the estimated expected rewards are strongly correlated). Again, we included in this model comparison voxels within a 4mm sphere centered on the peak voxels of the amygdalar activities correlating with precision signals for either the “model-based” HMM or the ‘’model-free’’ HMM using the leave-one out method. Here, we compared activity correlating with precision signals between the “model-based” and “model-free” HMM separately for the appetitive and aversive sessions (see Results section for the exceedance probabilities).

ROI Analyses

Functional regions of interest (ROIs) were defined using the MarsBaR toolbox (http://marsbar.sourceforge.net/). Beta estimates were extracted for each subject from the functional clusters of interest as they appeared on the statistical maps of a given contrast using the leave-one-out method to avoid a non-independence bias. They were then averaged across subjects to plot expected reward (Figure 14b) and precision (Figure 15b) according to three categories (category one corresponding to the lowest values and category three corresponding to the highest values).

Figures

Figure 11. Appetitive versus aversive Pavlovian learning task.



RW	$Qj(t+1) = Qj(t) + \alpha(R(t) - Qj(t))$
PH	$Qj(t+1) = Qj(t) + S \times \lvert R(t-1) - Qj(t-1)\rvert \times R(t)$
HMM	$Q_j(t) = E[R \mid cue\ j, trial\ t]$ $= \sum_{R\ rewards} \sum_{S_t states} R\ P(R \mid S_t, cue\ c) P(S_t)$

Figure 11. Sequence and timing of events in the appetitive (**a**) and aversive (**b**) sessions. On each trial, a cue was presented on one side of the screen for 4 seconds, followed by some liquid delivery 60% of the time. The trial ended with a 2-11s inter-trial interval. Each session started with the presentation of cue 1 and cue 2, leading 60% of the time to a pleasant or a neutral liquid delivery in the appetitive session or an unpleasant or a neutral liquid delivery in the aversive session. After a number of trials, a reversal occurred so that cue 1 now led to the liquid associated with cue 2, and cue 2 led to the liquid associated with cue 1. Subsequently, a new pair of cues was presented, which also reversed after a number of trials. In total, three new pair of cues were presented, and each of these pairs reversed once. **c**) Computational models used to estimate expected reward on each trial (Qj). The expected

rewards generated by the model-free learning algorithms (Rescorla-Wagner (RW) and Pearce-Hall (PH) were compared against a model-based learning algorithm (Hidden Markov Model or HMM) at both the behavioral and neural levels. **d**) Graphical model representation of the Bayesian HMM.

Figure 12. Behavioral results.

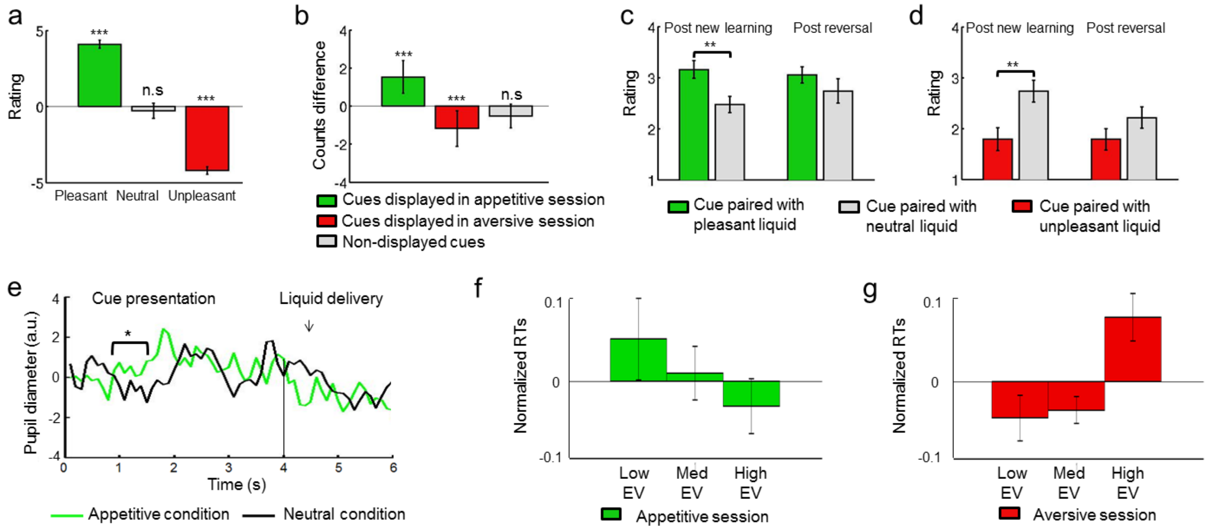


Figure 12. **a**) Ratings for the pleasant, neutral, and unpleasant liquids (-5 being very unpleasant and 5 very pleasant). *** indicates a significance of p<0.001 as computed by one sample t-tests comparing the mean of the different liquids against a mean of 0, n.s stands for not significant. **b**) Difference in the number of times a cue is preferred after - before the experiment. *** indicates a significance of p<0.001 as computed by one sample t-tests comparing the mean of the different liquids against a mean of 0. **c,d**) Ratings for the cue paired with the pleasant (**c**) or unpleasant (**d**) liquid and the cue paired with the neutral liquid after a few trials after a new pair of cue has been presented (Post new learning) and a few trials after a reversal has occurred (Post reversal). A rating of 1 indicates that participants strongly dislike the cue whereas a rating of 4 indicates that they strongly like it. ** indicates a significance of p<0.01 as computed by two sample t-tests comparing the means of the ratings for the cues paired with pleasant/unpleasant and neutral liquids. **e-g**) Conditioned responses. **e**) Time course for pupil diameter in response to cues paired with the pleasant liquid (green line) and the neutral liquid (black line) averaged across all trials in the appetitive session for the 10 subjects showing reliable amplitude in their pupil diameter. A one-tailed paired t-test for a time window 0.8-1.5s revealed a significant decrease in constriction when

participants were presented with cues paired with the pleasant liquid ($p<0.05$). **f,g**) Detrended and normalized response times averaged across low, medium, and high categories of expected values (EV) as determined by the model-based learning algorithm in the appetitive (**f**) and aversive (**g**) sessions.

Figure 13. Expected value and precision signal time series.

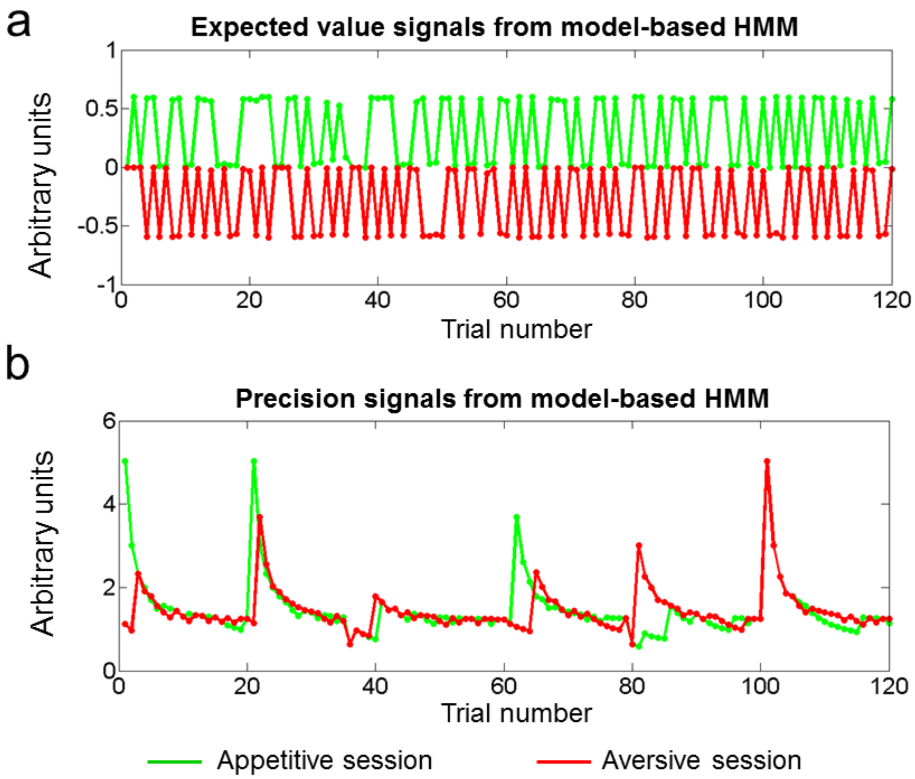


Figure 13. Plots showing expected value signals (a) and precision signals (b) from the model-based learning algorithm for the appetitive (green) and aversive (red) sessions for a typical participant.

Figure 14. Expected value signals from the model-based learning algorithm model in the amygdala.

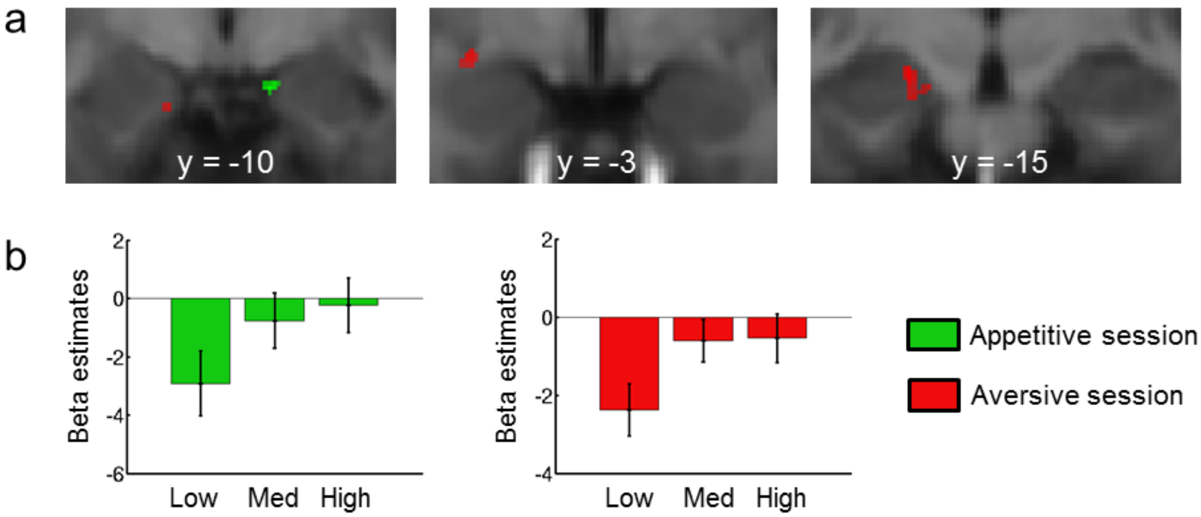


Figure 14. **a**) Blood oxygen level-dependent (BOLD) signals positively correlating with the magnitude of the expected value of the cue were found in the basolateral complex in the appetitive session (in green) and in the centromedial complex in the aversive session (in red). **b**) Plots showing the beta estimates for low, medium, and high categories of expected rewards in the appetitive (green) and aversive (red) sessions in the clusters activated using the leave-one-out method.

Figure 15. Precision signals from the model-based learning algorithm in the amygdala.

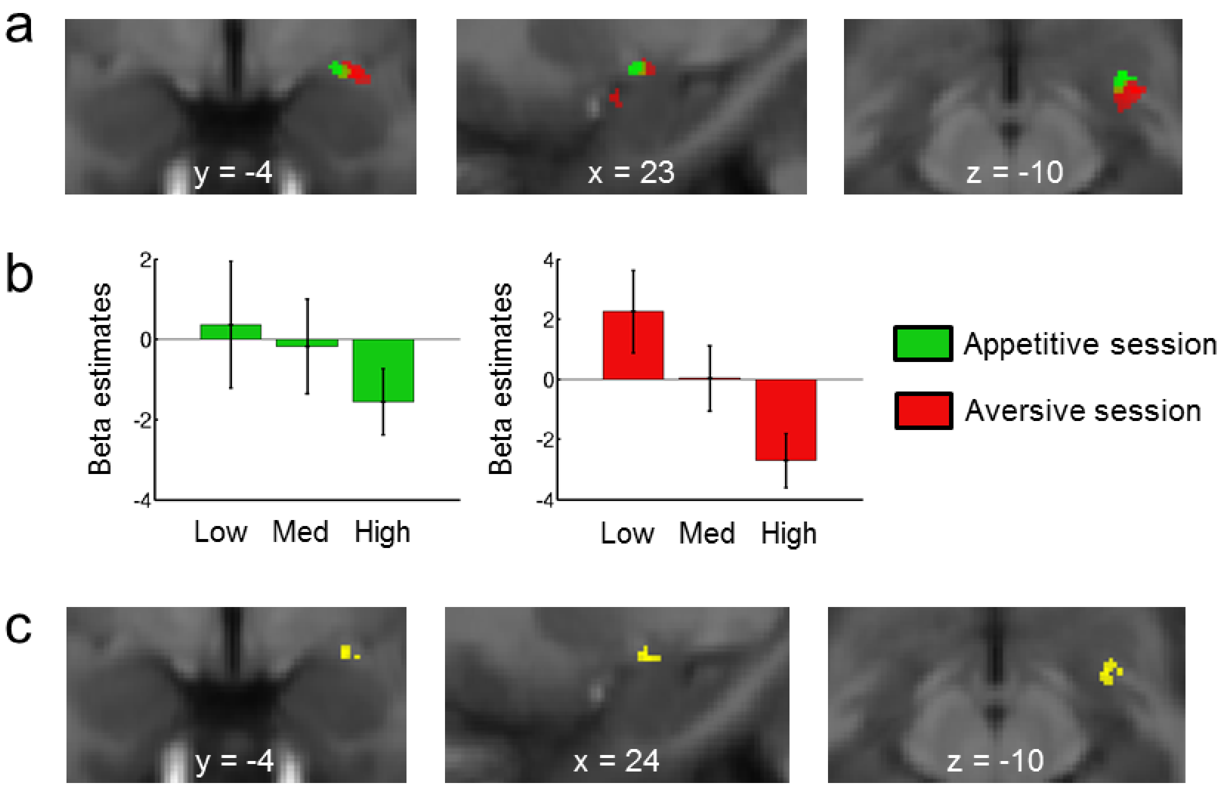


Figure 15. **a**) Blood oxygen level-dependent (BOLD) signals correlating with the precision of the cue were found in the centromedial complex of the amygdala in both the appetitive session (in green) and aversive session (in red). **b**) Plots showing the beta estimates for low, medium, and high categories of precision in the appetitive (green) and aversive (red) sessions in the clusters activated using the leave-one out method. **c**) Results from formal conjunction analysis of precision signals from the appetitive and aversive sessions in the centromedial complex.

Tables

Table 3. BIC scores.

Session	**HMM – RW**	**HMM - PH**
Aversive	Mean = -3.09	Mean = -5.08
	SEM = 1.05	SEM = 0.84
	p<0.01	p<0.0001
Appetitive	Mean = -6.22	Mean = -6.23
	SEM = 1.46	SEM = 1.53
	p<0.001	p<0.001

Table 3. Bayesian Information Criterion (BIC) values and standard errors to the mean (SEM) for the model-based learning algorithm (HMM) versus the model-free learning algorithms (RW and PH). A smaller BIC value indicates a better fit, therefore the model-based learning algorithm best fit the behavioral data. P-values as computed by paired t-tests on BIC values are also reported. HMM is significantly outperforming both model-free learning algorithms in both the appetitive and aversive sessions.

Table 4, Random effects test of models compared to baseline.

Session	HMM < Baseline	RW < Baseline	PH < Baseline
Aversive	<0.05	0.93	1
Appetitive	<0.001	1	1

Table 4. A random effects test of the models versus a baseline model was performed by simulating random expected value estimates (10,000 repetitions) and then computing a non-parametric p-value per subject as the fraction of repetitions in which the baseline BIC is lower than the model BIC. These p-values were then combined across subjects using Fisher's combined probability test. Only HMM outperforms the baseline model in both the appetitive and aversive sessions.

Chapter 5

BELIEF STATE-SPACE MANIPULATION IN HIERACHICAL INFERENCE

Hierarchical inference is a ubiquitous task for humans. From stock markets to social interactions, the natural world contains many hidden variables which may only be indirectly inferred based on conditionally related signals. Knowledge of the state of these ‘latent’ variables is required for optimal inference regarding the abstract decision structure of a given environment and therefore can be crucial to decision-making in a wide range of situations. Inferring the state of an abstract variable requires the generation and manipulation of an internal representation of beliefs over the values of the hidden variable. Here, we aimed to explore the learning strategies employed by human subjects in a hierarchical state-estimation task. The task contained two experimental conditions corresponding to whether or not “switches” could take place. We hypothesized that this key environmental feature would bias subject behavior between distinct learning strategies which differed depending on the subjects' manipulations of their belief state-space. Namely, in a “switch” condition, we expected participants to attempt to update beliefs over the entire belief state-space while in a “no-switch” condition, we hypothesized that they would exclude low-probability states from the inference process in order to minimize computational load.

Methods

Task

Participants engaged in an observational probabilistic hierarchical inference task akin to card-sorting tasks used to investigate executive functioning (Robbins, 1996). On each trial (Figure 16), subjects were presented with two stimuli composed of three features: color, motion, and shape. Each of these features had two exemplars. Color could be red or green, motion could be in the left or right direction, and shape could be a circle or a square. The two stimuli were randomly generated based on these exemplars in an anti-correlated fashion, i.e., one of the stimuli was red while the other was green, and similarly for the motion and shape dimensions. On any given trial, one of the exemplars was "active" but the subject was not told which exemplar this was. After the stimuli were presented for 1s, the stimulus containing the "active" exemplar was stochastically selected. The task of the subject was to infer which dimension contained the active exemplar. After a jittered ISI, subjects bet on which dimension contained the active exemplar by distributing $20 across the three dimensions. Subjects were paid the amount of money they put on the objectively correct dimension in a randomly chosen trial. Suppose that on such a trial, they bet $10 on color, $5 on motion, and $5 on shape. Then, if the correct dimension was actually motion, they would receive $5. It was explained to the subjects that they should bet on each dimension according to their beliefs and degree of uncertainty in their beliefs. It was explained to the subjects that they should pay close attention to the selection process and integrate information over several trials in order to form their beliefs. Each subject engaged in six blocks of 40 trials each. Half of the sessions contained unsignaled extra-dimensional switches in the active exemplar. That is, when a switch occurred, the new active exemplar came from a different dimension to the previous active exemplar. Switches could occur multiple times within the same block. Switch and no-switch blocks were randomly interleaved. At the start of each block, subjects were told whether they were in a switch or no-switch environment. Subjects were trained on simplified versions of the task immediately before the main experiment began.

Functional Magnetic Resonance Imaging

Twenty-two healthy right-handed Caltech students (mean age 24 years, SD 2.9, 14 male) volunteered to participate in this study. The data from two participants were excluded because their median performance was below that of a random decision-making algorithm. All participants gave informed consent and the study was approved by the Institutional Review Board of the California Institute of Technology. Functional imaging was performed with a 3 T Siemens Trio scanner. Forty-three contiguous interleaved transversal slices of echo-planar T2*-weighted images were acquired in each volume, with a slice thickness of 3mm and no gap (repetition time, 2340ms; echo time, 30ms; flip angle, 80°; field of view, 192mm^2; matrix, 64x64). Slice orientation was tilted 30° from a line connecting the anterior and posterior commissure. This slice tilt alleviates the signal drop in the OFC (Deichmann et al., 2003). We discarded the first three images before data processing and statistical analysis, to compensate for the T1 saturation effects. A whole-brain high-resolution T1-weighted structural scan (voxel size: 0.9 x 0.9 x 0.9mm^3) was also acquired for each subject.

Learning Model

Sensing Probability - Beta Distribution Parameter Updates

An alternative formulation would be to directly track the counts of each selection event on a per-dimension basis. This gives a very compact encoding which can be flexibly unpacked in order to generate probability representations. It is not dependent to any neuronal model of probability. We use a Dirichlet belief variable

$$B \sim Dirichlet(\alpha_1, \alpha_2, \alpha_3)$$

where the α_i parameters encode the strength of evidence within each dimension

$$\alpha_i := \alpha_i^+ := n_i^+ - n_i^-$$

Note that the index $+$ always refers to the feature which currently holds the balance of Probability, i.e.,

$$+ = argmax_{f=1,2} n^f$$

α_i can always be manipulated to reflect the number of selections of the alternative feature in that dimension by combining with n_c, the number of trials since the last changepoint:

$$n_i^- = \frac{n_c - \alpha_i^+}{2}$$

The mean and mode being, respectively,

$$\boldsymbol{p} := \frac{\boldsymbol{\alpha}}{\sum \boldsymbol{\alpha}}$$

$$\boldsymbol{x} := \frac{\boldsymbol{\alpha} - 1}{\sum \boldsymbol{\alpha} - 3}$$

The marginal distribution for a dimension i is

$$B_i \sim Beta(\alpha_i, \alpha_0 - \alpha_i)$$

where $\alpha_0 := \sum \boldsymbol{\alpha}$. This model suggests the following prediction errors:

Per-dimension prediction errors (with predictive coding):

$$\delta_i^{S=1} = \frac{\alpha_i + 1}{\alpha_0 + 1 - \alpha_i} - \frac{\alpha_i}{\alpha_0 - \alpha_i}$$

$$\delta_i^{S=0} = \frac{\alpha_i}{\alpha_0 + 1\ - \alpha_i} - \frac{\alpha_i}{\alpha_0\ - \alpha_i}$$

Per-dimension prediction errors (without predictive coding) could be modeled as follows:

$$\underline{B_i \sim Beta(n^+, n^-)}$$

$$\delta_i^{S=1} = \frac{n^+\ +\ 1}{n^+ + 1 + n^-} - \frac{n^+}{n^+ + n^-}$$

$$\delta_i^{S=0} = \frac{n^+}{n^+ + 1 + n^-} - \frac{n^+}{n^+ + n^-}$$

Across-dimension prediction errors are given by a divergence metric, although the expression for the KL-divergence between Dirichlets is relatively complicated, see (Gallistel, Krishan, Liu, Miller, & Latham, 2014) for the Beta version.

$$D_{KL}(B^{t+1}||B^t) = \sum_t log\left(\frac{B_i^{t+1}}{B_i^t}\right) B_i^{t+1}$$

where we define B^{t+1} as the Dirichlet variable computed up to trial t and $B^{t:t'}$ as the Dirichlet variable computed from trial t to t'. $B_{\boldsymbol{h}}^{t:t'}$ is the same Dirichlet variable but computed based on the hypothesis set $\boldsymbol{h}$. For BOLD signal modeling purposes, all prediction errors are presumed to be positive since information has no valence. We fit a prior for all α_i as a “softmax” parameter which is bounded from below by $\alpha_i\ = 2, n_i\ =\ 2$ for all variables.

Reasoning About Beliefs - Do I Need To Change My Beliefs?

This can be captured in a model by thresholding the divergence between a current belief and the posterior probability given the observed data:

$$n\, D_{KL}(B^{t'}||B^{t_u})\ >\ T_1$$

If T_1 is not exceeded, no further updating is performed on this trial, and the same beliefs are reported. This quantity is used in (Gallistel et al., 2014), however, it is a complex computation requiring perfect memory. It also raises the question, if $B^{t'}$ is being computed on every trial anyway, why not use that? A more computationally efficient and plausible method would be to threshold based on the history of "event prediction errors":

$$\sum\nolimits_{t_u:t} \log \frac{P_K(S)}{P_B(S)} = (t - t_u) \log 0.8 - \log P_B\left(S_{t_u:t}\right) > (t - t_u) T_1$$

where t' is the current trial, S is the observed selection event, $t - t_u$ is the number of trials since the last update, and K represents perfect knowledge of the generative process. This ratio compares how well your beliefs explain the environment dynamics versus the maximum possible prediction performance given perfect knowledge K. For example, the likelihood $L(S|K)$ of selection S on a given trial based on the generative model of the environment is $L(S|K) = 0.8$ in the switches condition and $L(S|K) = 0.7$ on the non-switches condition. This quantity is similar to the likelihood divided by the model evidence in Bayes' rule and thus probably results in an algorithm equivalent to that of (Gallistel et al., 2014). This could be computed on a trial-by-trial basis and averaged

$$\delta := \log\ L(S|K) - \log P_B(S)$$

$$\delta = \log L(S|K) - \sum_i \log L(S|B_i) P(B_i)$$

$$\delta = \log L(S|K) - \log \sum_E L(S|E) P(E|D) P(D)$$

$$\frac{\sum_t \delta_t}{n} > T_1$$

On a given trial, δ has a maximum value of $\log \frac{0.8}{0.2} = \log 4 = 1.386$ corresponding to an incorrect deterministic belief and a minimum value of $\log \frac{0.8}{0.8} = \log 1 = 0$ corresponding to a perfect prediction.

Reasoning About Beliefs - Hypothesis Elimination and Explaining Away

Hypothesis testing involves the elimination of hypotheses by thresholding beliefs. If the posterior probability of a particular hypothesis falls below a certain threshold T_2, then a conclusion is made that this hypothesis is incorrect and should not be considered any further. This is accomplished by directly threshold the posterior belief mode

$$x_i < T_2$$

then we eliminate dimension h_i as a hypothesis, set $\alpha_i = 0 \Rightarrow B_i = 0$, and no longer update. This will effect the updating over other dimensions via the "explaining away" phenomenon. For example:

$$P(h_1|S = 1, h_2 = 0) > P(h_1|S = 1)$$

Practically, we distribute the α_i variable amongst the other dimensions in proportion to their probabilities, e.g.,

$$(\alpha_1, \alpha_2, \alpha_3) \rightarrow \left(0, \alpha_2 + \left\lceil\frac{\alpha_1 \times \alpha_2}{\alpha_2 + \alpha_3}\right\rceil, \alpha_3 + \left\lfloor\frac{\alpha_1 + \alpha_3}{\alpha_2 + \alpha_3}\right\rfloor\right)$$

It is possible that subjects' would have some idea of the pattern of associations between features in previous trials and distribute α_i according to such memories, but this is probably quite noisy and we will disregard it. Practically, this means that we only "re-distribute" probability at the dimension level and not the exemplar level.

T_2 should depend on the number of hypotheses remaining, i.e., we would not use the same threshold to eliminate one of three hypotheses and also to eliminate one of two. Thus, T_2 is fitted as the product of a free parameter multiplied by the random chance level of probability which depends on the number of hypotheses remaining.

Reasoning about Beliefs - Re-evaluating Hypothesis Eliminations

People may have "second thoughts"—"was I really correct to eliminate hypothesis h_i?" If $x_i > T_2$ for the most recently eliminated hypothesis, we re-consider h_i as a hypothesis, update our hypothesis set $\boldsymbol{h}$, and re-compute B based on $S_{t_h:t}$ where t_h is the trial on which the last hypothesis elimination was performed. We only model the re-consideration of the last hypothesis elimination (rather than for example the last two) for two reasons: (i) bounded cognitive resources and (ii) re-consideration of two hypotheses is effectively a reversal. Note that rejecting a conclusion constitutes an “internal changepoint”. We assume that subjects cease to track information regarding dimensions other than the assumed correct dimension thus when a conclusion is rejected, the subject has no recent data to draw upon and therefore begins again with a flat prior.

Detecting Changepoints - Identification and Retrospection

Changepoint trials on which the active exemplar has been switched to another. We can detect such changes by comparing the likelihoods of competing explanations—either there was a changepoint on some trial or there wasn't. We compute a Bayes factor $K := \frac{P(S|M_1)}{P(S|M_0)}$ weighted by model priors $\frac{P(M_1)}{P(M_0)}$ in order to adjudicate. The model evidence for the no-changepoint model M_0 is the sum over hypothesis likelihood functions:

$$P(S_{t_c:t}|M_0) = \sum_i P(S_{t_c:t}|h_i)$$

For the changepoint model we need to sum over the possibility of the changepoint occurring on each trial:

$$P\big(S_{t_c:t}\big|M_1\big) = \sum_{t'}\left[\sum_i P(S_{t_c:t'}|h_i) + \sum_j P(S_{t':t}|h_i)\right] \times \frac{1}{t-t_c}$$

where $\frac{1}{t-t_c}$ represents the uniform probability that a trial t' contains the changepoint.

The Bayes factor K is not appropriate for model comparison in this case. In the prior model probabilities are not equal given that a changepoint relatively rare, thus we multiply the Bayes factor K by the prior odds:

$$\frac{P(M_1)}{P(M_0)} = \frac{\sum_{1:t-t_c} p_c \times (1-p_c)^{n-1}}{(1-p_c)^n} = \frac{(t-t_c)p_c}{1-p_c}$$

and threshold the following signal:

$$\frac{P(S|M_1)P(M_1)}{P(S|M_0)P(M_0)} > T_3$$

Note that these quantities can be computed simply by counting the number of hypothesis-consistent selection events occurring within the given timeframe (since the data likelihood is the same under all hypotheses). The minimum value of this quantity is 0 $(p_c \to 0)$ while the maximum value is ∞ $(p_c \to 1)$. Let us assume that $\max p_c = 0.33$ and $\min p_c = 0.005$, then a more realistic maximum value would be $\sim 2 \times 2 = 4$ and a more realistic minimum would be $\sim \frac{1}{2} \times 0.005 = 0.0025$.

If a changepoint is detected, we attempt to identify the specific trial on which this occurred via

$$t_c = \arg \min_{t_c < t' \leq t} \sum_{i,j} P\left(S_{t_c:t'} \middle| h_i\right) \times P\left(S_{t':t} \middle| h_j\right)$$

We also update our perception of the changepoint rate (also known as the hazard rate (Wilson, Nassar, & Gold, 2010)):

$$p_c \leftarrow \frac{n_c + \alpha_c - 1}{n + \alpha_c + \beta_c - 2}$$

where n_c is the number of perceived changepoints, n is the total number of trials, and α_c, β_c are hyperpriors to approximate the true changepoint rate. Note that this can handle the case where more than one hypothesis is still active. Just as with hypothesis elimination, reversals can be re-evaluated. In such a situation, we re-set the priors according to $S_{t_c:t}$.

Producing A Decision - Risk Attitudes and Dirichlet Modeling

Given their actual belief mode across dimensions encoded by $\boldsymbol{x}$, subjects may report a warped set of announced beliefs $\boldsymbol{\pi}$ according to their risk attitude. Introducing a risk-aversion parameter γ, we have:

$$\pi_i = \frac{x_i^{1/\gamma}}{\sum_j x_j^{1/\gamma}}$$

where $0 < \gamma < \infty$. This is isomorphic to re-scaling the belief parameters directly

$$\pi_i \coloneqq \alpha_i^{1/\gamma}$$

Model Fitting

Our main hypothesis was that subjects may use a hypothesis-testing strategy in order to simplify the state-space representation. As with statistical hypothesis testing (Liese & Miescke, 2008), this is accomplished by thresholding a summary statistic (in this case, a posterior belief) and rejecting hypotheses which did not meet this threshold. We assume that subjects subsequently eliminate this dimension from their belief state-space and no longer update this variable. We propose that this computational mechanism is a form of gating which might underpin executive attention (Lara & Wallis, 2014; Robbins, 1996). Parameter T_2 was a direct measure of this effect. The parametric value of T_2 measures the evidence $P(\boldsymbol{D}|h_i)$ required by each subject to reject a hypothesis h_i. Lower values of T_2 imply that more evidence is required in order to reject a hypothesis. In addition to fitting this parameter, we also fit[4] T_3, which controls subjects' sensitivity to switches, a risk attitude parameter, and a softmax parameter for a total of 4 parameters. We fit these parameters in two ways: (i) by standard maximum likelihood estimation (MLE) and (ii) by hierarchical Bayesian analysis (HBA). This was motivated by the fact that this task is particularly challenging and thus MLE may not be sufficient to accurately fit these parameters. Using HBA, we can "regularize" per-subject variability in task performance with group-level behavior, thus achieving a better fit. In addition, a direct comparison between these estimation methods has not been performed in the context of Bayesian inference and thus contributes to a growing suite of studies comparing these methods in the psychological literature (Farrell & Ludwig, 2008; Fox & Glas, 2001; Wiecki, Sofer, & Frank, 2013).

Objective Function

We used the cumulative Kullback-Leibler divergence between the model predictions and the actual announced beliefs of the subjects as a measure of model error. This was computed for

[4] In the switch condition only.

all sessions and thus represents a within-sample indicator of model prediction. This objective function was optimized during parameter estimation. Since more complex models tend to overfit, we need to regularize this objective function based on the number of parameters considered. We use the Bayesian Information Criterion (BIC) for the purposes of model comparison. A BIC formula based on KL-divergence is derived in Appendix B.

Maximum Likelihood Estimation

For MLE, we used the MYSTIC framework, a Python-based model-independent optimization package (McKerns, Strand, Sullivan, Fang, & Aivazis). We used the Nelder-Mead simplex solver, which wraps the *fmin* function in *scipy.optimize* (http://www.scipy.org/). This commonly used optimization algorithm ran until the relative change in the objective function dropped below 0.0001. Our parameters were constrained as described in Table 1.

Hierarchical Bayesian Analysis

Group and individual parameter were estimated simultaneously and related in a hierarchical model using Markov chain Monte Carlo sampling (MCMC) (Gelman, Carlin, Stern, & Rubin, 2003). Specifically, we used the Metropolis-Hastings method (MacKay, 2003). 100,000 were burned and 100,000 subsequently drawn to estimate the posterior over the model parameters. A comparison of three chains indicated that the sampling process had converged ($\hat{R} = 1$) (Brooks, Gelman, Jones, & Meng, 2011). The chains always converged in the switches condition but did not always converge in the no-switch condition despite the fact that subjects’ behavior is less noisy in the no-switch condition due to the relative simplicity of the environment. From a data analytics point of view, this might be intuitively explained by the fact that there is less independent behavioral data in the no-switch condition. Participants often converge on the correct dimension within the first 10 trials and then subsequently make very similar belief reports for the remainder of the experiment. We assumed that parameter values were normally distributed at the group-level and modeled hyperpriors for the group-level means. Specifically, we used a uniform prior (ranging over the corresponding parameter bounds) for the mean $\hat{\mu}$ and computed the group-level variance $\hat{\sigma}^2$ as

$$\hat{\sigma}^2 = \frac{1}{\sqrt{\hat{\mu}(m - \hat{\mu})/3}}$$

where m is the maximum possible value of that parameter. This ensured that the distribution was unimodal but flexible and also precluded the necessity of separately estimating the variability parameter (Ahn, Krawitz, Kim, Busemeyer, & Brown, 2011). MCMC was performed using the PyMC Python package (Patil, Huard, & Fonnesbeck, 2010) versions 2.3.

Methodological Comparison

In Figure 17 and Figure 18, samples from the hierarchical model of behavioral parameters are presented. In white are samples drawn from the prior before any sampling occurs, while in black are samples drawn from the stationary distribution after the chains had been determined to have converged. These figures suggest that two important characteristics of accurate parameter fitting are satisfied using HBA (i) the posterior probability density is concentrated smoothly around unique combinations of parameter values (as opposed to a noisy extended density function with a large non-trivial support) and (ii) the stationary posterior distributions are not dependent on the specification of a prior. In contrast to many decision-making paradigms (Wiecki et al., 2013), learning models result in dependencies between samples. This means that the data likelihood $P(D|\theta)$ is non-Markov. In order to compute $P(\boldsymbol{D}|\theta)$, the learning model must be re-run in its entirety for each new combination of parameter values θ. This computational bottleneck can be avoided by pre-computing all learning model predictions for a grid approximation of parameter values $\hat{\theta}$. The sampling algorithm can then draw on these pre-computed predictions instantly for the nearest-neighbor parameter values $\hat{\theta}$ in order to compute $P(\boldsymbol{D}|\theta) \approx P(\boldsymbol{D}|\hat{\theta})$. In Figure 19, I plot correlations between parameter fits based on sampling with and without “grid preparation”. Parameter values estimated using the two approaches were very highly correlated despite the dramatic reduction in processing time for sampling based on pre-prepared model predictions.

In order to compare the quality of HBA and MLE parameter fits, model predictions from both methods were Spearman correlated with the announced beliefs of the subjects for the switches and no-switches conditions separately (Figure 20). In paired t-tests across subjects and conditions, it was found that the HBA method led to model predictions that were more strongly correlated with the announced beliefs from the subjects (p=0.00022).

General Linear Modeling

SPM8 (http://www.fil.ion.ucl.ac.uk/spm/) was used for slice timing correction, volume realignment, and spatial normalization to the Montreal Neurological Institute (MNI) echoplanar imaging template. All volumes were then spatially smoothed using a three-dimensional Gaussian kernel (at a full-width-half-maximum of 8mm). Prior to GLM estimation, the data was also high-pass filtered by removing signal components oscillating at a frequency below 1/120Hz.

We estimated general linear models on a per-subject basis and then, in order to model random effects, performed second-level t-tests on sets of first-level contrasts (Penny, Holmes, & Friston, 2003). There were three timepoints of interest in each trial (Figure 1), namely (i) cue onset, (ii) stimulus selection, and (iii) simplex onset. Broadly speaking, these cues should elicit representations of prior information, updating based on the observed selection, and representation of posterior beliefs, respectively. Thus, we matched the relevant components of our learning model to their corresponding trial timepoints. Given that beliefs are represented in three dimensions, we used negative entropy as a univariate measure of belief for the purposes of linear modeling. We hypothesized that (i) posterior probabilities for each dimension would be represented in the same region and (ii) local mutual inhibition would imply that the neural signaling emanating from that region would scale with the strength of belief in any one dimension (Machens, Romo, & Brody, 2005, 2010; Strait, Blanchard, & Hayden, 2014). This is captured by negative entropy which is effectively a non-parametric measure of precision. During the update phase of the experiment, we used Jensen-Shannon divergence as an information-theoretic measure of the distance between the prior and the posterior. This can be thought of has a “Bayesian prediction error” triggered by the observation of a stimulus selection (see Appendix A).

Each no-switch session had seven regressors in total, four onsets for the cue, stimulus selection, simplex onset, and rating. Parametric modulators were included for the cue (negative entropy of the prior $-H(Prior)$), stimulus selection (Bayesian prediction error), and simplex onset (negative entropy posterior $-H(Posterior)$). In the switches condition an extra parametric modulator was added to the stimulus selection indicator regressor corresponding to T_3. Time series of head motion estimated during realignment were included as covariates of no interest. All results are displayed at p<0.001 uncorrected.

Discussion

Computational theories of prefrontal cortex broadly agree that one of the main functions of this area is to direct attention towards relevant goals and environment variables (E. K. Miller & J. D. Cohen, 2001). An important component of this process is estimating the current state of the environment (Gold & Shadlen, 2007). In many realistic scenarios (Stanley & Adolphs, 2013), the current state of the environment is hidden and must be estimated over many interactions with the environment from conditionally related signals. This process is made particularly difficult by the high-dimensionality of our stimulus-rich world (Robert C. Wilson & Yael Niv, 2011). The Wisconsin card-sorting task has often been used to measure a participant's ability to flexibly identify relevant environment variables and adapt to switches in the environment dynamics (Berg, 1948). Here we scan subjects engaged in a probabilistic hierarchical inference task inspired by the Wisconsin card-sorting task while they are scanned using functional magnetic resonance imaging. A computational analysis of their behavior revealed that they used a belief thresholding strategy (see Learning Model, Figure 21) in conditions where the environment is stable. In uncertain conditions, they did not eliminate hypotheses from their state-space and their behavior was well-approximated by a Bayesian algorithm.

We performed a model-based fMRI analysis using time series of internal variables generated from our learning model. Our results are consistent with the neural hypothesis that dlPFC holds and updates an internal model of the belief state-space. We find that activity in dlPFC scales with the certainty of priors beliefs at cue onset (Figure 22), and also reflects the Bayesian prediction error driven by stimulus selection (Glascher et al., 2010). Subsequently, dorsomedial prefrontal cortex activity scales with the certainty of posterior beliefs when subjects make their response (Figure 23). This is consistent with studies which show that dmPFC is more active in decision-making scenarios where knowledge gleaned from abstract state-space representations must be integrated into behavior (Alan N. Hampton et al., 2006). An important component of learning in the switches condition is identifying whether a changepoint has occurred. We find (Figure 24) that fronto-polar cortex tracks increasing evidence of a changepoint as expected due to previously acquired evidence that FPC is involved in computations regarding the favorability of alternative behavioral strategies (Boorman, Behrens, Woolrich, & Rushworth, 2009).

Figures

Figure 16. Task.

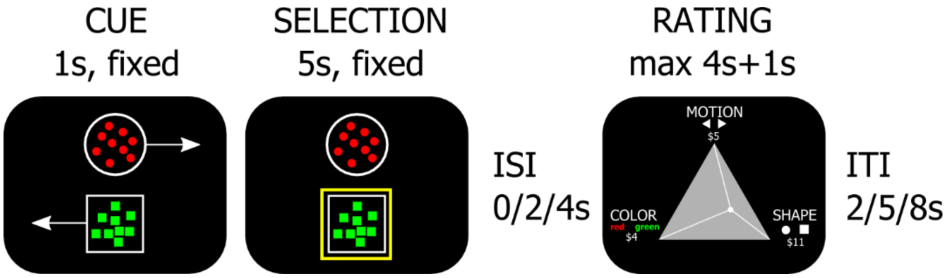


Figure 16. At cue onset, two stimuli are randomly generated in three binary dimensions. One second later, the environment stochastically selects one of the stimuli based on which exemplar is “active” (which is unknown to the subject). After an inter-stimulus interval of 0.2, or 4 seconds, subjects report their beliefs regarding which dimension is currently active on a continuous scale. Each trial is separated by an inter-trial interval of 2, 5, or 8 seconds.

Figure 17. MCMC example, group-level.

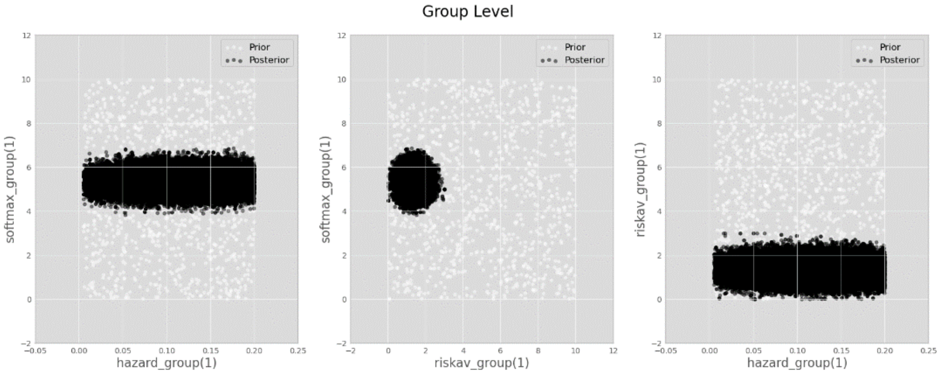


Figure 17. Samples from the prior are in white while samples drawn from the stationary distribution are in black. These samples are plotted for softmax, risk aversion, and hazard rate parameter distributions in the switches condition. One can see that the model is insensitive to the specific value of the hazard rate. In contrast, the softmax and risk aversion parameters combine focally in parameter space.

Figure 18. MCMC example output, individual-level.

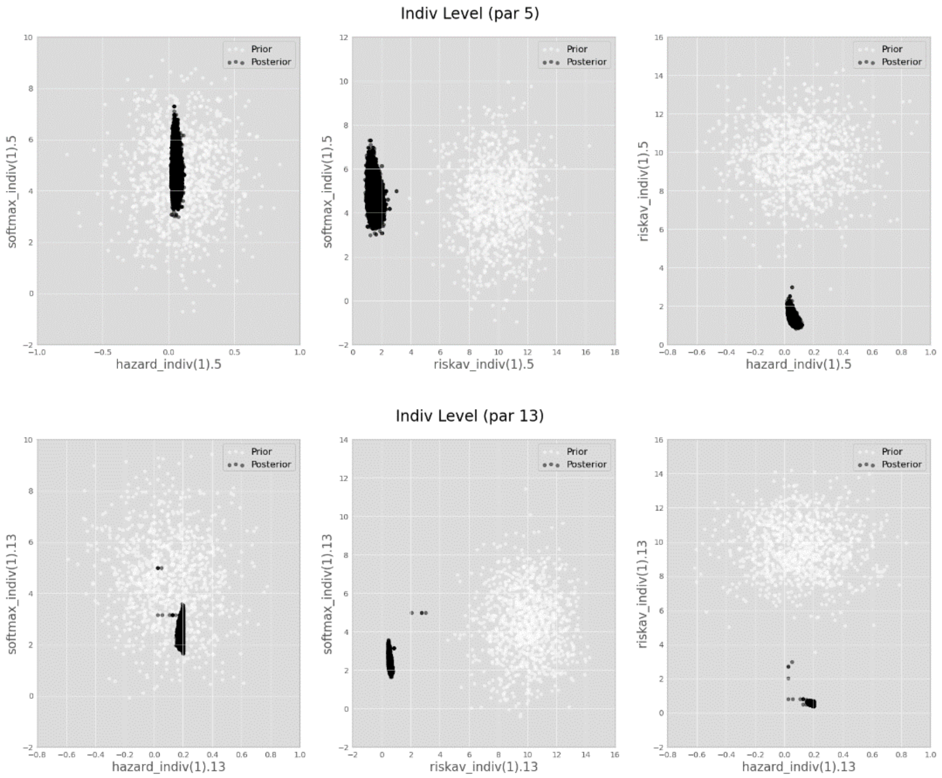


Figure 18. Plotted are the prior and stationary distributions for two subjects (5 and 13). Note the Gaussian priors in contrast to the uniform distributions used for the group mean hyperprior (Figure 17). The large divergences from the prior to the stationary posterior indicate that the model fitting procedure is not reliant on the specification of a strong prior (Bishop, 2006). One can see that, at the level of individual subjects, the probability density function is peaked on specific values of the hazard rate. This is in contrast to the group-level (Figure 17). This suggests a large variability in expected changepoint probability across subjects.

Figure 19, Grid preparation of model beliefs for HBA.

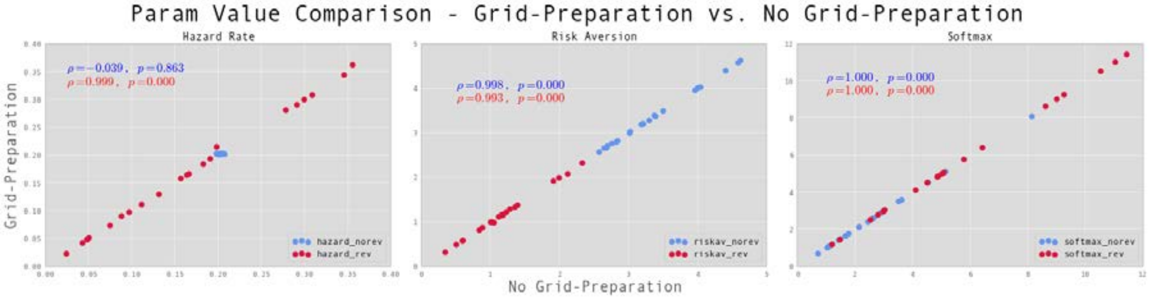


Figure 19. In order to speed up the HBA analysis, model beliefs were pre-computed as a function of all learning parameters. Then the sampling process need only to retrieve the pre-computed beliefs rather than compute them online. Here I compare the model fits with “no grid-preparation” (150,000 samples, ~3hrs) and without “grid-preparation” (10,000, ~36hrs). The results show that the correlation for all fitted parameters is > 0.99, thus vindicating this method. Note the lack of correlation for the hazard rate in the no-reversal condition as expected.

Figure 20. Correlation with announced beliefs computed using HBA and MLE techniques.

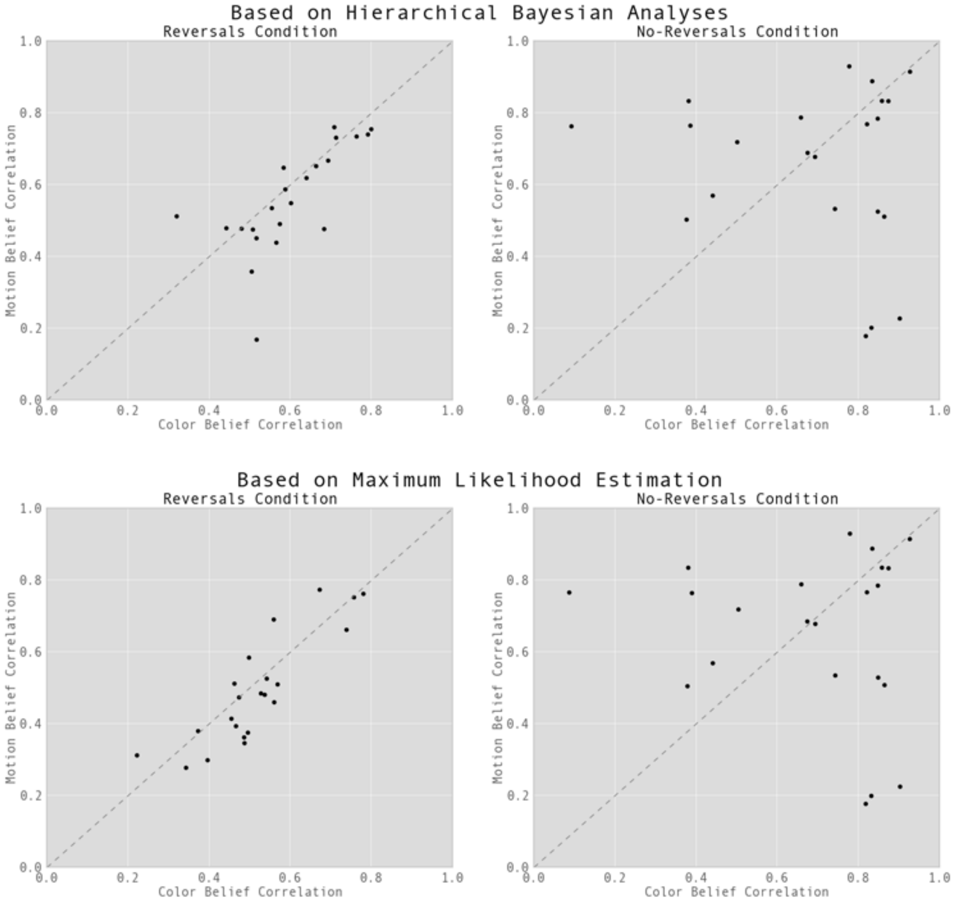


Figure 20. As a quantitative indication of the relative quality of the HBA and MLE fits, model predictions from both methods were Spearman correlated with the announced beliefs of the subjects. Correlations based on HBA are plotted in the top panels, while the bottom panels display the same data for MLE. This is done for the switches (or “reversals”) and no-switches (“no-reversals”) conditions separately. In paired t-tests across subjects and conditions, it was found that the HBA method led to model predictions that were more strongly correlated with the announced beliefs from the subjects (p=0.00022).

Figure 21. Model comparison.

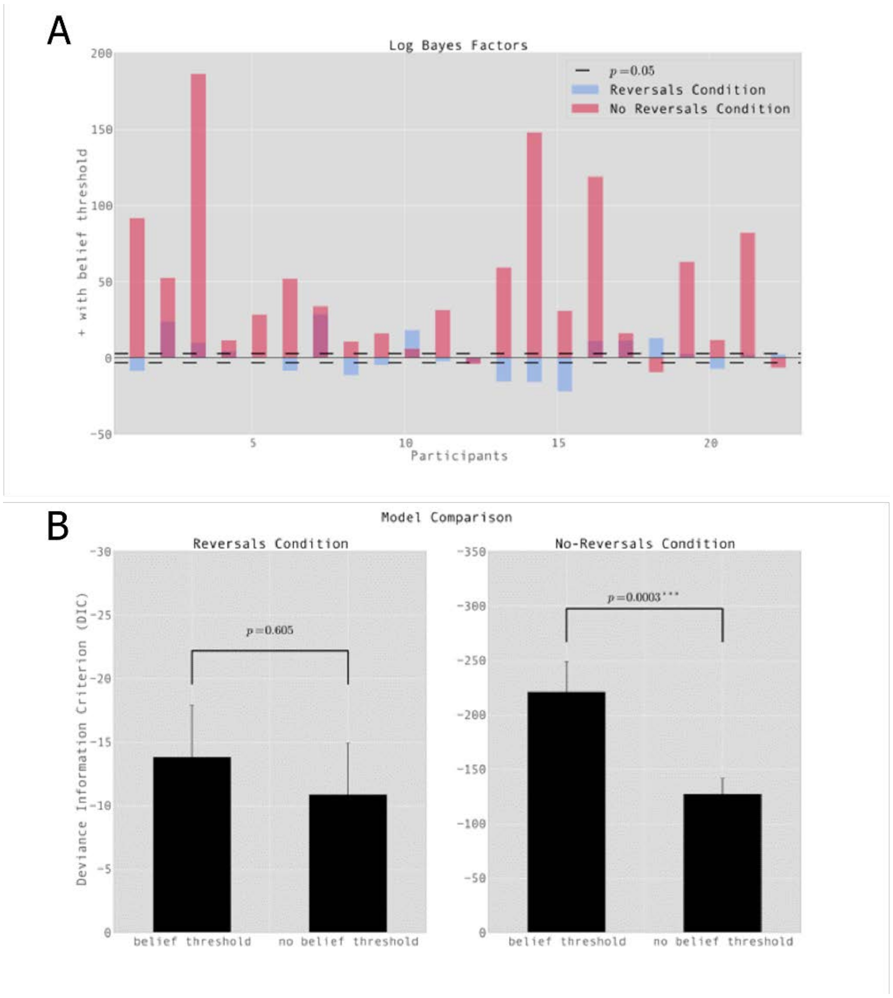


Figure 21. **a**) log Bayes factors for each subject plotted per condition. The larger the log Bayes factor, the better the fit of a belief thresholding model compared to a purely Bayesian learner. Log Bayes factors are estimated from 100,000 samples of the stationary MCMC chain. These results indicate strong evidence for a belief thresholding mechanism in the no-switches condition but are ambiguous in the switches condition. **b**) This is consistent with the use of deviance information criterion (DIC) as an indicator of model fit. DIC is a hierarchical generalization of BIC, which is easily computed via MCMC (Patil et al., 2010).

Figure 22. Neural activity at cue and simplex onsets.

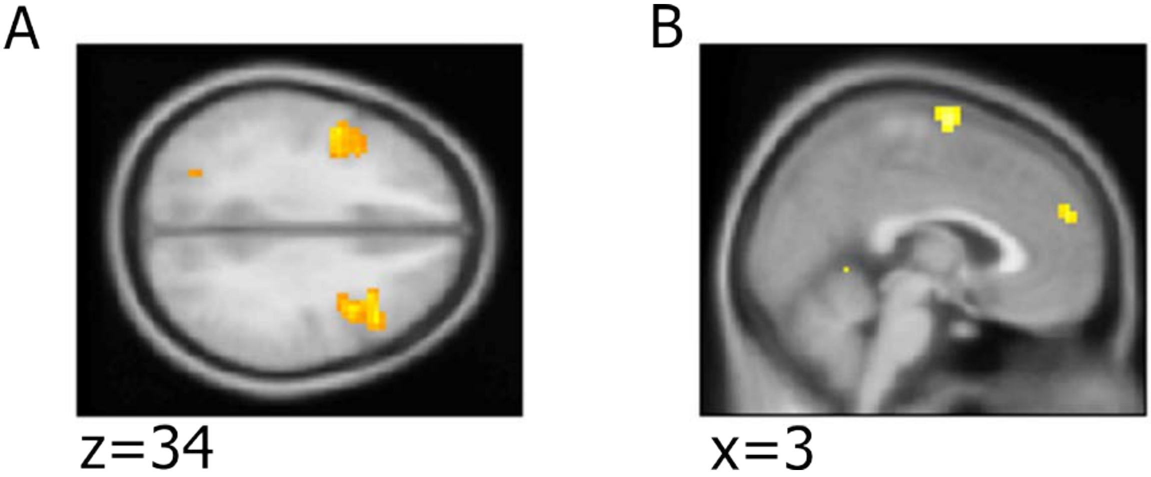


Figure 22. **a**) $-H(Prior)$ was represented in dlPFC bilaterally at cue onset. These clusters were positioned very similarly to those observed in (Glascher et al., 2010) for state prediction errors. **b**) At simplex onset, $-H(Posterior)$ correlated with activity in dorsomedial prefrontal cortex (dmPFC, t(19)=4.42, x=3, y=53, z=25) and premotor cortex.

Figure 23. Bayesian prediction errors in dlPFC and ventral striatum at stimulus selection onset.

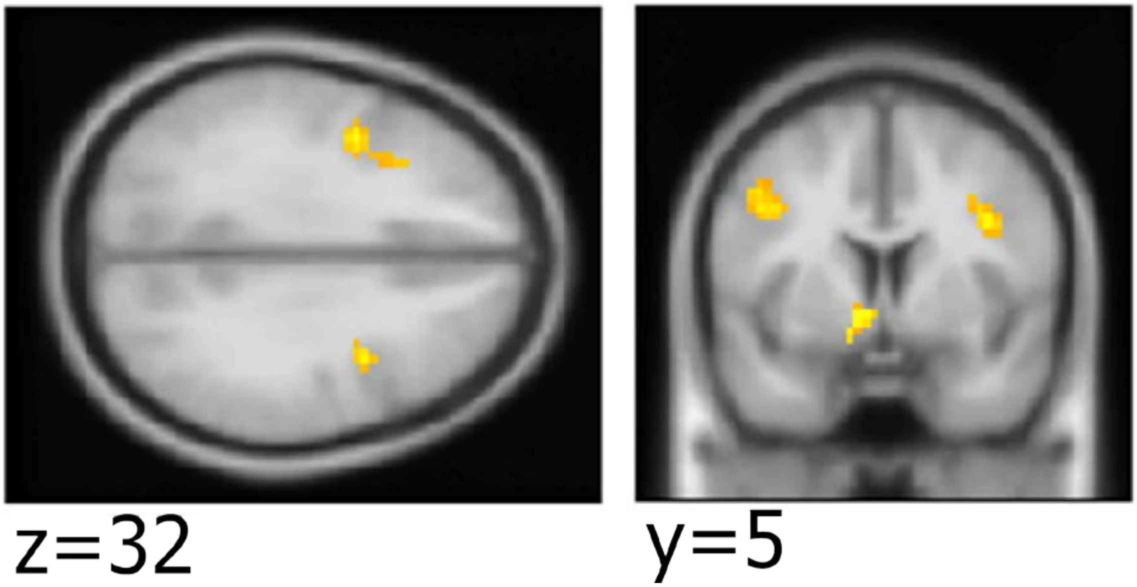


Figure 23. Bayesian prediction errors were observed in dlPFC and ventral striatum at the stimulus selection timepoint. Note that no $-H(Prior)$ correlation was observed in ventral striatum at the cue onset timepoint.

Figure 24. Neural correlates of changepoint model Bayes factors.

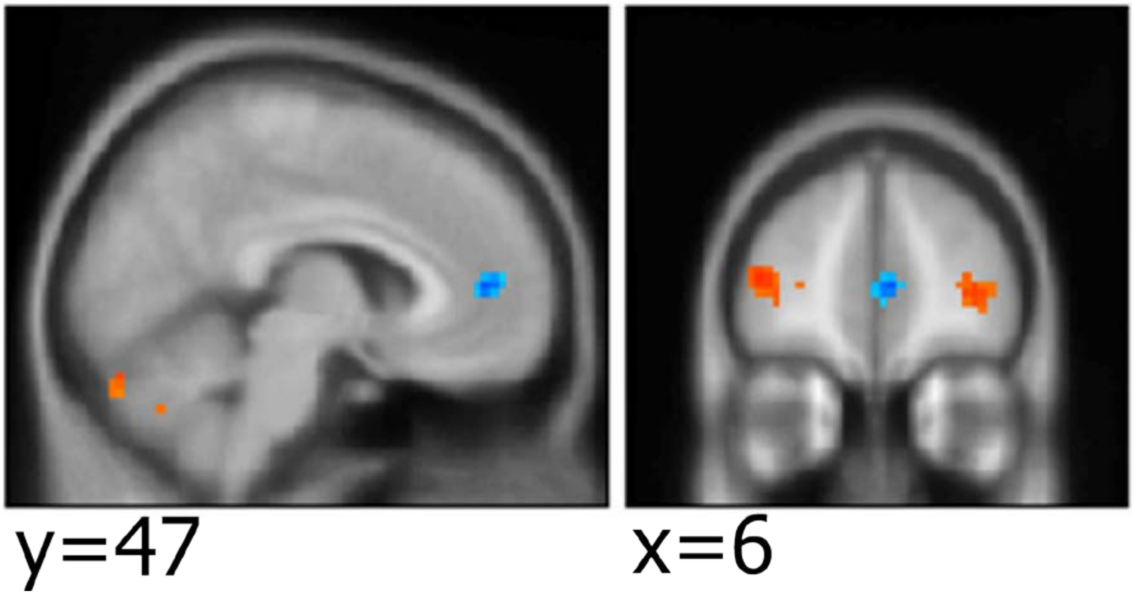


Figure 24. Increasing evidence of a changepoint (red) was represented by bilateral clusters in fronto-polar cortex. Activity in dorsal vmPFC scaled with evidence consistent with a changepoint not having occurred.

Tables

Table 5. Parameter bounds.

Parameter	*Low Bound*	*Upper Bound*
T_2 (testing threshold)	0	1
T_3 (switch sensitivity)	0	10
γ (risk)	0.005	10
τ (softmax)	2	20

Table 5. Parameter bounds for MLE and HBE analyses.

Chapter 6

CONCLUSIONS

The studies described in this thesis address several questions in the neurobiological and psychological literatures regarding the representation of variables while the brain engages in several standard learning and decision-making paradigms. Since a groundbreaking study in 2001 (Haxby et al., 2001), the use of pattern analysis tools in functional neuroimaging has become more prevalent to the point where it forms a robust minority of studies (Norman, Polyn, Detre, & Haxby, 2006). As an investigative tool, pattern identification techniques (Bishop, 2006) provides a complimentary approach to the univariate analyses typically used in functional neuroimaging experiments (Coutanche, 2013). This allows us to evaluate some underlying assumptions of previous neuroimaging studies as to the meaning of representation-via-correlation (John P. O'Doherty, Alan Hampton, & Hackjin Kim, 2007). For example, if the activity in a region *correlates* with the value of chosen actions (Klaus Wunderlich, Rangel, & O'Doherty, 2009), is that region *representing* the chosen action? Can we detect evidence for a distinct neural firing pattern that is unique to a particular action? Similar arguments can be made for other decision variables such as environmental states and values (Vikram S. Chib et al., 2009). MVPA can provide evidence in support or against such conclusions as shown in Chapter 3. Based on a time-delayed binary choice paradigm, we decoded representations of decision variables at different points in time throughout the decision process. We *a priori* defined several regions of interest based on related univariate functional imaging studies (McNamee, Rangel, & O'Doherty, 2013; Saori C. Tanaka, Bernard W. Balleine, & John P. O'Doherty, 2008; Elizabeth Tricomi et al., 2009)

and tested whether the corresponding representations were present in these regions while a decision is made (Table 2). Our results were broadly consistent with our current assumptions regarding the functional role played by each region within the model-free vs. model-based framework (Dolan & Dayan, 2013). In particular, the representation of outcome identity in prefrontal cortex (both dorsolateral and ventromedial portions) is supported by a wealth of lesion data (Fellows & Farah, 2003) and the presence of outcome representations in anterior caudate is congruent with a fMRI study showing that anterior caudate activity scales with the learned correlation between an action and an outcome (Saori C. Tanaka et al., 2008). We found that actions but not outcomes were represented in dorsolateral striatum. Previous functional imaging (Elizabeth Tricomi et al., 2009) found this region to be particularly implicated in habitual action control (see Chapter 1 for a review). For example, after multiple sessions of free responding, subjects (who were deemed to be insensitive to outcome devaluation and thus acting habitually) were found to have an overall increased level of activity in posterior putamen (Elizabeth Tricomi et al., 2009). Neurophysiological studies are less clear on the DLS vs. DMS distinction between action and outcome encoding. Both action and outcome signaling were present in both these regions in rodent models (Thomas A. Stalnaker, Gwendolyn G. Calhoon, Masaaki Ogawa, Matthew R. Roesch, & Geoffrey Schoenbaum, 2010). How should be integrate the clear differences found in our MVPA study? First, a failure to reject to the null hypothesis does not imply that a particular representation does not exist in a region. Combined electrophysiological/fMRI studies have shown that some signaling detected in the visual stream using single-unit recordings is inaccessible to fMRI (Dubois, de Berker, & Tsao, 2015). Taking the aforementioned example of action but not outcome encoding in DLS, we can reasonably conclude that either (a) action encoding is stronger than outcome encoding in DLS, or (b) that actions and outcomes are represented differently (e.g., neuronal response and/or feature space characteristics may vary), and that the action encoding mechanism is more amenable to fMRI. In (Dubois et al., 2015), it was concluded the spatial clustering of neurons with similar response profiles was necessary for fMRI-MVPA decoding while sparseness did impact classification performance. A second possible source of divergence between the results of fMRI-MVPA and single-unit experiments is the nature of a typical electrophysiological analysis. Usually, neurons are categorized by their response profiles and counts of neurons whose activity significantly correlates with task variables are tabulated. This essentially univariate analysis may lead to erroneous negative results whenever representations are encoded in a distributed fashion. For example, if a stimulus is encoded in a vector space spanned by a set of attribute dimensions and neurons only encode a one of these dimensions than analyzing these neurons in isolation may not reveal the hypothesized distributed representations in this region.

In Chapter 3, we show that stimuli and actions are combined into unique distributed representations. To our knowledge, these are the first published results showing that the brain combines internal and external variables to form new variables. Going forward, it is possible that many other combinations of internal and external variables may similarly be utilized by the brain. In the social domain, the combination of an agent representation along with an action representation would be one example. These results also emphasize a deeper philosophical point regarding psychology and neuroscience research — in a typical experiment, we do not know what representations are being used by the brain's algorithms. Given the fundamental interplay between representation and algorithm (Marr & Poggio, 1976), it seems that this theme would serve as a rich seam of research. In particular, it highlights how crucial neural data can be since it seems that no psychological experiment could provide evidence for such representational strategies. Determining the functional role of such representations and how they emerge will be critical to understanding how action control systems in the brain operate and interact (Dolan & Dayan, 2013; O'Doherty, Lee, & McNamee, 2015). The theory of predictive state representations may provide a computational framework in which to understand these representations (Littman, 1996). Briefly, states are represented by the preceding history of state-action combinations that led to that state. For example, if an agent transitions through S1-A1-S2-A2-S3 then state S3 is represented as S1-A1-S2-A2. This scheme addresses the question of how a completely novice agent, with no prior experience of the environment, consolidates its experiences into a structured representation of their environment. Subjects might be scanned in more complex environments while similar decoding analyses are performed in order to confirm whether subjects build representations of novel environments in a similar manner to that predicted by this theory.

MVPA can also determine *how* a variable is being represented. An important question is this regard is “what is the feature space in which a variable is transmitted?” There are multiple approaches to this issue within the MVPA framework. One prominent technique is representational dissimilarity analysis (Kriegeskorte, Mur, & Bandettini, 2008) where a distance measure is applied to neural data vectors usually followed by a cluster analysis. Furthermore, the resulting representational dissimilarity matrix can be embedded in a Euclidean space for visualization. An experimenter would infer that clustered data samples are represented in the same manner and, most importantly, that the distance between data samples represents the distance between samples in feature space. One drawback of this approach is that results can depend strongly on the distance measure used (Kriegeskorte et al., 2006). Thus, this approach is particularly well suited to large sets of diverse stimuli in order to ensure robustness. In Chapter 2, we addressed this question of neural representation in the case of value signaling in the human brain using an alternative approach (McNamee et al., 2013). Instead of “model-free” representational dissimilarity analyses, we ran multiple decoding analyses using different data labels. From these results we logically deduced representational characteristics of the neural data. For example, we provide the first evidence that value is represented in a manner which is dependent on the class of stimuli being evaluated. In addition, the loci of stimulus-dependent and stimulus-independent value representations was suggestive of a stimulus-to-value pathway from central orbitofrontal cortex to medial orbitofrontal cortex to more dorsal regions of ventromedial prefrontal cortex. In addition, these results pointed to the presence of multiple functional gradients in the ventromedial prefrontal cortex. Namely, a posterior-to-anterior gradient with respect to stimulus abstractness along the medial orbital surface and a ventral-to-dorsal gradient along which the complexity of value encoding decreased. That is, value signaling tended to be more distributed along the orbital surface. Several primate experimental datasets are consistent with various components of this broad neurobiological theory (Cai & Padoa-Schioppa, 2012; Kringelbach & Rolls, 2004; Wallis & Kennerley, 2011) but further experimentation is required for confirmation. In order to corroborate the gradient hypotheses, one could repeat a similar experiment as the one performed here but with a more controlled set of stimuli which evenly spanned the spectrum of "abstractness". Probably the most difficult challenge of such a study would be to define and precisely measure the psychological notion of stimulus abstractness on a continuous scale. Published electrophysiological data is consistent the aforementioned ventral-dorsal gradient however the functional implications of this network-level are not understood (S. Kennerley et al., 2011). Theoretic analysis of data from multi-electrode arrays would be an ideal approach however implantation in such deep cortical regions remains difficult. In the absence of an empirical dataset, theoretic analyses could still proceed in order to make predictions regarding the functional role of these distinct value signaling strategies. Testable behavioral predictions could be generated for patients with brain lesions (Gläscher et al., 2012). Ideally, two populations of such patients would be used. One population with lesions along the orbital surface and the other more dorsally. A rough conjecture would be that the former population would be inhibited when evaluating novel goals online while the latter would make noisier choices for goals with learned, or easily computable, values (e.g., monetary sums). Finally, the premise of this study ignores "the other half" of decision valuation — the costs of actions themselves. Based on recent neuroimaging (Croxson, Walton, O'Reilly, Behrens, & Rushworth, 2009), electrophysiological (Hillman & Bilkey, 2012), and especially lesion-based evidence (Camille, Tsuchida, & Fellows, 2011; Rudebeck et al., 2008), one hypothesis would be that a parallel action-to-value transformation occurs in dorsomedial portions of the brain and that both stimulus-to-value and action-to-value signals are integrated in a "final common value pathway" in subgenual cingulate in preparation for action initiation. Although the feature space for actions could be very high-dimensional (corresponding to all possible degrees of freedom in the human body), probably a relatively small subspace is relevant for valuation (e.g., energy cost) thus value gradients may be more difficult to detect.

The fundamental distinction between the results of Chapters 2 and 3 is that the focus of the analysis on variable and algorithm respectively. Chapter 2 answers the question of how a variable (in this case, goal value) is represented while Chapter 3 effectively asks "how is an algorithm represented?" In other words, what are the input and output variables of the algorithm. This latter question can be addressed using univariate fMRI analyses since our inference is based on the presence or absence of particular variables. Chapter 4 provides an example of this. From the representational point of view, we ask "how is Pavlovian learning represented in the amygdala?" Due to the presence of neural activity correlating with precision signals that can only be obtained from learning based on a higher-order model of the environment structure, we conclude that Pavlovian conditioning, in the amygdala, is model-based on nature. This result could potentially be applied to a wide range of research paradigms which focus on the amygdala, for example, fear conditioning and the identification of emotion from facial expressions (Adolphs, 2010). In the case of facial expressions, one could ask how our knowledge regarding the mental state of another human being impacts our classification of their emotional state. This paradigm would be amenable to a model-based analysis and, based on our results here, we would hypothesize that amygdala activity would be sensitive to the belief structure a participant holds for another person (Koster-Hale & Saxe, 2013).

With respect to the task in Chapter 4, one could ask how participants might come to understand the task structure without instruction. How could one determine what the relevant variables are in the environment for decision-making? In Chapter 5, we examine how humans construct belief structures online. Participants engaged in a probabilistic Wisconsin Card-Sorting Task (Berg, 1948) and reported their posterior beliefs on a continuous scale

(Figure 16). Behavioral analyses (Figure 21) showed that humans dynamically adjust the process by which decision state-spaces are constructed. In particular, it was found subjects were less likely to reject hypotheses regarding the environment dynamics when unsignaled changepoints could occur (Robert C. Wilson & Yael Niv, 2011). This behavior is consistent with an optimal observer model. Intuitively, it means that humans try to keep track of as many variables as possible in order to be able to adapt rapidly to an environment changepoint. If there are no changepoints (as was the case in half the sessions), there is no reason to do encode and update all possible variables, and instead, one can focus on those which are relevant in the current state of the environment. More generally, one can view these results as integrating reinforcement learning with cognitive science (Chater, Tenenbaum, & Yuille, 2006; Téglás et al., 2011) — we studied how subjects construct cognitive representations via reinforcement. Neurally, activity in dorsolateral prefrontal cortex (dlPFC) was correlated the precision of the prior beliefs on each trial (as measured by negative entropy) and the divergence between the prior and posterior as evidence is accumulated (a “bayesian prediction error”). This is consistent with the functional characterization of dlPFC as a “working memory” module (P. S. Goldman-Rakic, 1995). Our results provide a computationally precise, model-based point of view which complements the prior literature (John P. O'Doherty et al., 2007). We can view our results within the classical framework of attention whereby dlPFC evaluates the relevance of state-space components (i.e., the color, motion, and shape variables) to the current decision problem. Recent electrophysiological data in monkeys in an analogous region is congruent with this interpretation (S. Tremblay, Pieper, Sachs, & Martinez-Trujillo, 2014). Despite the sophisticated nature of the computational mechanisms observed, this thresholding model could lead to sub-optimal inferences regarding the structure of the world and thus decisions. A hypothesis regarding the dynamics of the environment might be erroneously eliminated due to an unlikely run of events. The development of these cognitive “blind spots” could potentially be present in a wide range of decision-making paradigms and is ripe for further investigation. For example, one could model the “jumping-to-conclusions” bias (Averbeck, Evans, Chouhan, Bristow, & Shergill, 2011) using the computational model developed in Chapter 5 (P. Read Montague et al., 2012). This bias is often observed in clinical populations with disruptions in dopaminergic pathways in the brain such as seems to be the case with schizophrenia (Rolls, Loh, Deco, & Winterer, 2008). Working memory impairments have already been shown to underpin reinforcement learning deficits in schizophrenic patients (Collins, Brown, Gold, Waltz, & Frank, 2014) which suggest, in concert with the results described here, that disruptions in the mesocortical pathway (which projects to dlPFC) might impede schizophrenic patients ability to construct and maintain accurate models of the decision environment.

APPENDIX A

Processing of rewards and percepts are typically modeled using different tools in computational neuroscience based on the scientific development of these disciplines. Learning is typically modeled using Bayesian formalisms in the perceptual domain (Gold & Shadlen, 2007) while associative learning or, more generally, reinforcement learning (RL) is used in reward-based decision-making (Peter Dayan & Berridge, 2014; John P. O'Doherty, 2011) with some notable exceptions (A. Courville, Daw, & Touretzky, 2004; A. C. Courville et al., 2006; Nathaniel D. Daw & Courville, 2007). Based on the simple functional analogy between information-theoretic log-probabilities and economic value as exemplified in the computational convergence of decision-making models across these domains (Krajbich, Armel, & Rangel, 2010), we show Bayesian updating is equivalent to reinforcement learning in “log space”. It turns out that the learning rate parameter in RL is effectively a temperature or softmax parameter in Bayesian updating. Let us consider the simplest possible application of Bayesian inference—the estimation of a single parameter value h given a stream of data D. We assume that we have an *a priori* defined model which specifies the data likelihood values $P(D|h)$. Based on a single sample, Bayes’ theorem gives the following update:

$$\begin{aligned} P(h) &\leftarrow \frac{P(D|h)P(h)}{P(D)} \\ &\leftarrow \frac{P(D|h)P(h)}{\sum_h P(D|h)P(h)} \end{aligned}$$

In log space,

$$\log P(h) \leftarrow \log P(D|h) + \log P(h) - \log \sum_h P(D|h)P(h)$$

This can be re-arranged into a linear update rule based on summing a prediction error

$$\log P(h) \leftarrow \log P(h) + \left[\log P(D|h) - \log \sum_h P(D|h)P(h)\right]$$

What is missing in this expression is a learning rate. After adding a learning rate α, we will re-interpret this equation in “exp space”.

$$\log P(h) \leftarrow \log P(h) + \alpha \left[\log P(D|h) - \log \sum_h P(D|h)P(h) \right]$$

$$\log P(h) \leftarrow \log P(h) + \log P(D|h)^\alpha - \log \left(\sum_h P(D|h)P(h) \right)^\alpha$$

$$\log P(h) \leftarrow \log P(h) + \log \frac{P(D|h)^\alpha}{P(D)^\alpha}$$

$$P(h) \leftarrow P(h) \times \left(\frac{P(D|h)}{P(D)} \right)^\alpha$$

Here, $\left(\frac{P(D|h)}{P(D)}\right)^\alpha$ can be interpreted as a ratio prediction error. It measures the ratio between the evidence of data D given perfect knowledge of the variable under investigation h against the current prediction. The prediction error is scaled by α which weights its contribution to the new estimate of $P(h)$ as a learning rate would.

APPENDIX B

Here we derive the Bayesian Information Criterion (Schwarz, 1978) for KL-divergence. The basic definition of KL-divergence D is

$$D(P,Q) := \sum_i P_i \log \frac{P_i}{Q_i}$$

but we will use it in the following equivalent form[5]:

$$D(P,Q) := -H(P) - \langle \log Q \rangle_P$$

The BIC formula is

$$-2 \log P(x|M) \approx BIC := -2 \log \max_\theta L(x|\theta, M) + k(\log n - \log 2\pi)$$

For $n \gg 0$, this is effectively

$$BIC \approx -2 \log \max_\theta L(x|\theta, M) + k \log n$$

In terms of log-likelihoods based on our model-predicted P_M and announced belief distributions P_A,

$$D(P_A, P_M) = -H(P_A) - \langle \log \max_\theta L(x|\theta, M) \rangle_{P_A}$$
$$= -\frac{1}{3n} \sum_{i,t} \log \max_\theta L(x_{it}|\theta, M) + C$$

assuming a flat prior on the data where n is the number of trials and the factor of 3 refers to the three belief dimensions. Thus minimizing KL-divergence is equivalent to maximizing log-likelihood and we have

$$BIC \approx -2 \log \max_\theta L(x|\theta, M) + k \log 3n$$

[5] See http://www.hongliangjie.com/2012/07/12/maximum-likelihood-as-minimize-kl-divergence/.

$$= -2[-3n\, D(P_A, P_M)] + \; k \log 3n$$
$$= 6n\, D(P_A, P_M) + \; k \log 3n$$

Note that KL-divergence is additive over samples

$$D(P_{1:n}, Q_{1:n}) = D(P_{1:j}, Q_{1:j}) \; + \; D(P_{j:n}, Q_{j:n})$$

Note that $2D(P, Q)$ is the G-statistic, to which χ^2 is now understood to be an approximation[6]. There are only two degrees of freedom in our data and predictions, namely two out of the three beliefs. In our derivation, this implies that the prior over the “third'” belief is 1 since it is uniquely defined by $\sum_{i=1\ldots3} P_A(i) = 1$. Thus we simply drop this belief vector from our BIC score to get,

$$BIC = \; 4n\, D(P_A, P_M) + \; k \log 2n$$

where P_A and P_M are computed over the color and motion dimensions only.

[6] See e.g. http://strimmerlab.github.io/statisticalthinking/pdf/c4.pdf, page 4.

BIBLIOGRAPHY

Adolphs, R. (2010). Conceptual challenges and directions for social neuroscience. *Neuron, 65*(6), 752-767. doi: 10.1016/j.neuron.2010.03.006

Ahmad, S., & Yu, A. J. (2013). Active sensing as Bayes-optimal sequential decision-making. *Proceedings of the Twenty-ninth Annual Conference on Uncertainty in Artificial Intelligence (UAI).*

Ahn, W.-Y., Krawitz, A., Kim, W., Busemeyer, J. R., & Brown, J. W. (2011). A model-based fMRI analysis with hierarchical Bayesian parameter estimation. *Journal of Neuroscience, Psychology, and Economics, 4*(2), 95-110. doi: 10.1037/a0020684

Averbeck, B. B., Evans, S., Chouhan, V., Bristow, E., & Shergill, S. S. (2011). Probabilistic learning and inference in schizophrenia. *Schizophrenia Research, 127*(1-3), 115-122. doi: 10.1016/j.schres.2010.08.009

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nat Rev Neurosci, 10*(9), 659-669. doi: 10.1038/nrn2667

Badre, D., Doll, B. B., Long, N. M., & Frank, M. J. (2012). Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration. *Neuron, 73*(3), 595-607. doi: 10.1016/j.neuron.2011.12.025

Balleine, B., & O'Doherty, J. P. (2010). Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology, 35*(1), 48-69. doi: 10.1038/npp.2009.131

Balleine, B. W., Daw, N. D., & O'Doherty, J. P. (2008). Multiple forms of value learning and the function of dopamine. In P. W. Glimcher, C. Camerer, E. Fehr, & R. A. Poldrack (Eds.), *Neuroeconomics: decision making and the brain* (pp. 367-385). New York: Elsevier.

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology, 37*(4-5), 407-419. doi: 10.1016/S0028-3908(98)00033-1

Balleine, B. W., & Ostlund, S. B. (2007). Still at the Choice-Point. *Annals of the New York Academy of Sciences, 1104*(1), 147-171.

Barto, A. G., & Mahadevan, S. (2003). Recent advances in hierarchical reinforcement learning. *Discrete Event Dynamic Systems, 13*(4), 341-379.

Baxter, M. G., & Murray, E. A. (2002). The amygdala and reward. *Nat Rev Neurosci, 3*(7), 563-573. doi: 10.1038/nrn875

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50*(1-3), 7-15. doi: 10.1016/0010-0277(94)90018-3

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral science, 9*(3), 226-232.

Beckmann, M., Johansen-Berg, H., & Rushworth, M. (2009). Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *The Journal of* *neuroscience : the official journal of the Society for Neuroscience, 29*(4), 1175-1190. doi: 10.1523/JNEUROSCI.3328-08.2009

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nat Neurosci, 10*(9), 1214-1221.

Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of general psychology, 39*, 15-22. doi: 10.1080/00221309.1948.9918159

Bishop, C. (2006). *Pattern Recognition and Machine Learning*: Springer.

Bitsios, P., Szabadi, E., & Bradshaw, C. M. (2004). The fear-inhibited light reflex: importance of the anticipation of an aversive event. *Int J Psychophysiol, 52*(1), 87-95. doi: 10.1016/j.ijpsycho.2003.12.006

Boorman, E., Behrens, T., Woolrich, M., & Rushworth, M. (2009). How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron, 62*(5), 733-743. doi: 10.1016/j.neuron.2009.05.014

Botvinick, M. M., Niv, Y., & Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. *Cognition, 113*(3), 262-280. doi: 10.1016/j.cognition.2008.08.011

Bradley, M. M. (2009). Natural selective attention: orienting and emotion. *Psychophysiology, 46*(1), 1-11. doi: 10.1111/j.1469-8986.2008.00702.x

Bray, S., & O'Doherty, J. (2007). Neural coding of reward-prediction error signals during classical conditioning with attractive faces. *J Neurophysiol, 97*(4), 3036-3045. doi: 10.1152/jn.01211.2006

Bray, S., Rangel, A., Shimojo, S., Balleine, B., & O'Doherty, J. P. (2008). The neural mechanisms underlying the influence of pavlovian cues on human decision making. *J Neurosci, 28*(22), 5861-5866. doi: 10.1523/JNEUROSCI.0897-08.2008

Brooks, S., Gelman, A., Jones, G. L., & Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*: Chapman and Hall/CRC.

Buchel, C., & Dolan, R. J. (2000). Classical fear conditioning in functional neuroimaging. *Curr Opin Neurobiol, 10*(2), 219-223.

Cai, X., & Padoa-Schioppa, C. (2012). Neuronal Encoding of Subjective Value in Dorsal and Ventral Anterior Cingulate Cortex. *Journal Of Neuroscience, 32*(11), 3791-3808. doi: 10.1523/JNEUROSCI.3864-11.2012

Camille, N., Tsuchida, A., & Fellows, L. K. (2011). Double dissociation of stimulus-value and action-value learning in humans with orbitofrontal or anterior cingulate cortex damage. *Journal Of Neuroscience, 31*(42), 15048-15052. doi: 10.1523/JNEUROSCI.3164-11.2011

Carmichael, S., & Price, J. (1996). Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. *Journal of comparative neurology, 371*(2), 179-207.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: conceptual foundations. *Trends in Cognitive Sciences, 10*(7), 287-291. doi: 10.1016/j.tics.2006.05.007

Chen, Y., Namburi, P., Elliott, L. T., Heinzle, J., Soon, C. S., Chee, M. W. L., & Haynes, J.-D. (2011). Cortical surface-based searchlight decoding. *NeuroImage, 56*(2), 582-592. doi: 10.1016/j.neuroimage.2010.07.035

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. doi: 10.1523/JNEUROSCI.2575-09.2009

Chumbley, J., Worsley, K., Flandin, G., & Friston, K. (2010). Topological FDR for neuroimaging. *NeuroImage, 49*(4), 3057-3064. doi: 10.1016/j.neuroimage.2009.10.090

Clithero, J., Smith, D. V., Carter, R. M., & Huettel, S. (2011). Within- and cross-participant classifiers reveal different neural coding of information. *NeuroImage, 56*(2), 699-708. doi: 10.1016/j.neuroimage.2010.03.057

Collins, A. G. E., Brown, J. K., Gold, J. M., Waltz, J. A., & Frank, M. J. (2014). Working Memory Contributions to Reinforcement Learning Impairments in Schizophrenia. *Journal Of Neuroscience, 34*(41), 13747-13756. doi: 10.1523/JNEUROSCI.0989-14.2014

Corrado, G., & Doya, K. (2007). Understanding neural coding through the model-based analysis of decision making. *J Neurosci, 27*(31), 8178-8180. doi: 10.1523/JNEUROSCI.1590-07.2007

Courville, A., Daw, N. D., & Touretzky, D. S. (2004). Similarity and discrimination in classical conditioning: A latent variable account. *NIPS, 17*.

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends Cogn Sci, 10*(7), 294-300. doi: 10.1016/j.tics.2006.05.004

Coutanche, M. N. (2013). Distinguishing multi-voxel patterns and mean activation: why, how, and what does it tell us? *Cognitive, Affective & Behavioral Neuroscience, 13*(3), 667-673. doi: 10.3758/s13415-013-0186-2

Croxson, P. L., Johansen-Berg, H., Behrens, T., Robson, M. D., Pinsk, M. A., Gross, C. G., . . . Rushworth, M. (2005). Quantitative investigation of connections of the prefrontal cortex in the human and macaque using probabilistic diffusion tractography. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 25*(39), 8854-8866. doi: 10.1523/jneurosci.1311-05.2005

Croxson, P. L., Walton, M. E., O'Reilly, J. X., Behrens, T. E. J., & Rushworth, M. F. S. (2009). Effort-based cost-benefit valuation and the human brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 29*(14), 4531-4541. doi: 10.1523/JNEUROSCI.4515-08.2009

Damasio, H., Grabowski, T., Frank, R., Galaburda, A. M., & Damasio, A. R. (1994). The return of Phineas Gage: clues about the brain from the skull of a famous patient. *Science (New York, N.Y.), 264*(5162), 1102-1105.

Daw, N. D., & Courville, A. C. (2007). The pigeon as particle filter. *Advances in Neural Information Processing Systems*.

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. doi: 10.1016/j.neuron.2011.02.027

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. doi: 10.1038/nn1560

Dayan, P., & Berridge, K. C. (2014). Model-based and model-free Pavlovian reward learning: Revaluation, revision, and revelation. *Cognitive, Affective & Behavioral Neuroscience, 14*(2), 473-492. doi: 10.3758/s13415-014-0277-8

Dayan, P., & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cogn Affect Behav Neurosci, 8*(4), 429-453. doi: 10.3758/CABN.8.4.429

Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nat Neurosci, 3 Suppl*, 1218-1223. doi: 10.1038/81504

de Wit, S., & Dickinson, A. (2009). Associative theories of goal-directed behaviour: a case for animal–human translational models. *Psychological Research PRPF, 73*(4), 463-476.

de Wit, S., Watson, P., Harsay, H. A., Cohen, M. X., van de Vijver, I., & Ridderinkhof, K. R. (2012). Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. *J Neurosci, 32*(35), 12066-12075. doi: 10.1523/JNEUROSCI.1088-12.2012

Dearden, R., Friedman, N., & Andre, D. (1999). *Model based Bayesian Exploration.*

Deichmann, R., Gottfried, J. A., Hutton, C., & Turner, R. (2003). Optimized EPI for fMRI studies of the orbitofrontal cortex. *NeuroImage, 19*(2 Pt 1), 430-441.

Delgado, M. R., Olsson, A., & Phelps, E. A. (2006). Extending animal models of fear conditioning to humans. *Biol Psychol, 73*(1), 39-48. doi: 10.1016/j.biopsycho.2006.01.006

Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 308*(1135), 67-78.

Dolan, R. J., & Dayan, P. (2013). Goals and Habits in the Brain. *Neuron, 80*(2), 312-325. doi: 10.1016/j.neuron.2013.09.007

Donoso, M., Collins, A. G., & Koechlin, E. (2014). Human cognition. Foundations of human reasoning in the prefrontal cortex. *Science, 344*(6191), 1481-1486. doi: 10.1126/science.1252254

Doya, K., Samejima, K., Katagiri, K., & Kawato, M. (2002). Multiple model-based reinforcement learning. *Neural Comput, 14*(6), 1347-1369. doi: 10.1162/089976602753712972

Duarte, A., Henson, R. N., Knight, R. T., Emery, T., & Graham, K. S. (2010). Orbito-frontal cortex is necessary for temporal context memory. *Journal of Cognitive Neuroscience, 22*(8), 1819-1831. doi: 10.1162/jocn.2009.21316

Dubois, J., de Berker, A. O., & Tsao, D. Y. (2015). Single-Unit Recordings in the Macaque Face Patch System Reveal Limitations of fMRI MVPA. *Journal Of Neuroscience, 35*(6), 2791-2802. doi: 10.1523/JNEUROSCI.4037-14.2015

Elliott, R., Dolan, R. J., & Frith, C. D. (2000). Dissociable functions in the medial and lateral orbitofrontal cortex: evidence from human neuroimaging studies. *Cerebral Cortex, 10*(3), 308-317.

Elliott, R., Newman, J. L., Longe, O. A., & William Deakin, J. F. (2004). Instrumental responding for rewards is associated with enhanced neuronal response in subcortical reward systems. *NeuroImage, 21*(3), 984-990. doi: 10.1016/j.neuroimage.2003.10.010

Engel, Y. (2005). *Algorithms and Representations for Reinforcement Learning.* Retrieved from http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.124.6809

Fanselow, M. S., & LeDoux, J. E. (1999). Why we think plasticity underlying Pavlovian fear conditioning occurs in the basolateral amygdala. *Neuron, 23*(2), 229-232.

Farrell, S., & Ludwig, C. J. H. (2008). Bayesian and maximum likelihood estimation of hierarchical response time models. *Psychonomic bulletin & review, 15*(6), 1209-1217. doi: 10.3758/PBR.15.6.1209

Fellows, L. K., & Farah, M. J. (2003). Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm. *Brain, 126*(Pt 8), 1830-1837. doi: 10.1093/brain/awg180

Fellows, L. K., & Farah, M. J. (2005). Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans. *Cerebral Cortex, 15*(1), 58-63. doi: 10.1093/cercor/bhh108

FitzGerald, T. H. B., Seymour, B., & Dolan, R. J. (2009). The role of human orbitofrontal cortex in value comparison for incommensurable objects. *Journal Of Neuroscience, 29*(26), 8388-8395. doi: 10.1523/JNEUROSCI.0717-09.2009

Formisano, E., & Kriegeskorte, N. (2012). Seeing patterns through the hemodynamic veil – the future of pattern-information fMRI. *NeuroImage.* doi: 10.1016/j.neuroimage.2012.02.078

Fox, J.-P., & Glas, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using gibbs sampling. *Psychometrika, 66*(2), 271-288. doi: 10.1007/BF02294839

Franklin, D. W., & Wolpert, D. M. (2011). Computational mechanisms of sensorimotor control. *Neuron, 72*(3), 425-442. doi: 10.1016/j.neuron.2011.10.006

Friston, K., Schwartenbeck, P., Fitzgerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2013). The anatomy of choice: active inference and agency. *Front Hum Neurosci, 7*, 598. doi: 10.3389/fnhum.2013.00598

Fuster, J. n. M. (2008). *The Prefrontal Cortex* (4th ed.): Elsevier Ltd.

Gallistel, C. R., Krishan, M., Liu, Y., Miller, R., & Latham, P. E. (2014). The perception of probability. *Psychological Review, 121*(1), 96-123. doi: 10.1037/a0035232

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2003). *Bayesian Data Analysis* (Second ed.): Chapman and Hall/CRC.

Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychol Rev, 117*(1), 197-209. doi: 10.1037/a0017808

Gershman, S. J., & Niv, Y. (2010). Learning latent structure: carving nature at its joints. *Curr Opin Neurobiol, 20*(2), 251-256. doi: 10.1016/j.conb.2010.02.008

Gläscher, J., Adolphs, R., Damasio, H., Bechara, A., Rudrauf, D., Calamia, M., Paul, L., Tranel, D. (2012). Lesion mapping of cognitive control and value-based decision making in the prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 109*(36), 14681-14686. doi: 10.1073/pnas.1206608109

Glascher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron, 66*(4), 585-595. doi: 10.1016/j.neuron.2010.04.016

Glover, G. H., Li, T. Q., & Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magn Reson Med, 44*(1), 162-167.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience, 30*, 535-574. doi: 10.1146/annurev.neuro.29.051605.113038

Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron, 14*(3), 477-485. doi: 10.1016/0896-6273(95)90304-6

Goldman-Rakic, P. S. (1996). Regional and cellular fractionation of working memory. *Proceedings of the National Academy of Sciences, 93*(24), 13473-13480.

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2002). Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *J Neurosci, 22*(24), 10829-10837.

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science, 301*(5636), 1104-1107. doi: 10.1126/science.1087919

Gremel, C. M., & Costa, R. M. (2013). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nature communications, 4*.

Hampton, A. N., Adolphs, R., Tyszka, M. J., & O'Doherty, J. P. (2007). Contributions of the amygdala to reward expectancy and choice signals in human prefrontal cortex. *Neuron, 55*(4), 545-555. doi: 10.1016/j.neuron.2007.07.022

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *Journal Of Neuroscience, 26*(32), 8360-8367. doi: 10.1523/JNEUROSCI.1010-06.2006

Hampton, A. N., & O'Doherty, J. P. (2007). Decoding the neural substrates of reward-related decision making with functional MRI. *Proceedings of the National Academy of Sciences of the United States of America, 104*(4), 1377-1382. doi: 10.1073/pnas.0606297104

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V., & Pollmann, S. (2009). PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics, 7*(1), 37-53. doi: 10.1007/s12021-008-9041-y

Hare, T., Camerer, C., & Rangel, A. (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. *Science, 324*(5927), 646-648. doi: 10.1126/science.1168450

Hare, T., Malmaud, J., & Rangel, A. (2011). Focusing attention on the health aspects of foods changes value signals in vmPFC and improves dietary choice. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 31*(30), 11077-11087. doi: 10.1523/JNEUROSCI.6383-10.2011

Hare, T., O'Doherty, J., Camerer, C., Schultz, W., & Rangel, A. (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *Journal Of Neuroscience, 28*(22), 5623-5630.

Hare, T. A., Camerer, C. F., Knoepfle, D. T., & Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *Journal Of Neuroscience, 30*(2), 583-590. doi: 10.1523/JNEUROSCI.4089-09.2010

Hare, T. A., Malmaud, J., & Rangel, A. (2011). Focusing attention on the health aspects of foods changes value signals in vmPFC and improves dietary choice. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 31*(30), 11077-11087. doi: 10.1523/JNEUROSCI.6383-10.2011

Harlow, J. M. (1848). Passage of an iron rod throught the head. *Boston Med. Surg. J., 39*, 389-393.

Harris, A., Adolphs, R., Camerer, C., & Rangel, A. (2011). Dynamic Construction of Stimulus Values in the Ventromedial Prefrontal Cortex. *PLoS ONE, 6*(6), e21074-e21074. doi: 10.1371/journal.pone.0021074

Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning* (2nd ed.): Springer.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430. doi: 10.1126/science.1063736

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature reviews. Neuroscience, 7*(7), 523-534. doi: 10.1038/nrn1931

Hikosaka, O., Sakamoto, M., & Usui, S. (1989). Functional properties of monkey caudate neurons. III. Activities related to expectation of target and reward. *J Neurophysiol, 61*(4), 814-832.

Hillman, K. L., & Bilkey, D. K. (2012). Neural encoding of competitive effort in the anterior cingulate cortex. *Nature Neuroscience, 15*(9), 1290-1297. doi: 10.1038/nn.3187

Insausti, R., Juottonen, K., Soininen, H., Insausti, A. M., Partanen, K., Vainio, P., . . . Pitkanen, A. (1998). MR volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. *AJNR Am J Neuroradiol, 19*(4), 659-671.

Jimura, K., & Poldrack, R. A. (2012). Analyses of regional-average activation and multivoxel pattern information tell complementary stories. *Neuropsychologia, 50*(4), 544-552. doi: 10.1016/j.neuropsychologia.2011.11.007

Johansen, J. P., Cain, C. K., Ostroff, L. E., & LeDoux, J. E. (2011). Molecular mechanisms of fear learning and memory. *Cell, 147*(3), 509-524. doi: 10.1016/j.cell.2011.10.009

Jones, J. L., Esber, G. R., McDannald, M. A., Gruber, A. J., Hernandez, A., Mirenzi, A., & Schoenbaum, G. (2012). Orbitofrontal Cortex Supports Behavior and Learning Using Inferred But Not Cached Values. *Science, 338*(6109), 953-956. doi: 10.1126/science.1227489

Kable, J., & Glimcher, P. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience, 10*(12), 1625-1633. doi: 10.1038/nn2007

Kahnt, T., Chang, L. J., Park, S. Q., Heinzle, J., & Haynes, J. D. (2012). Connectivity-Based Parcellation of the Human Orbitofrontal Cortex. *Journal Of Neuroscience, 32*(18), 6240-6250. doi: 10.1523/JNEUROSCI.0257-12.2012

Kahnt, T., Heinzle, J., Park, S. Q., & Haynes, J.-D. (2010). The neural code of reward anticipation in human orbitofrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 107*(13), 6010-6015. doi: 10.1073/pnas.0912838107

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience, 8*(5), 679-685. doi: 10.1038/nn1444

Kennerley, S., Behrens, T., & Wallis, J. (2011). Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nature Neuroscience*. doi: 10.1038/nn.2961

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. doi: 10.1162/jocn.2009.21100

Killcross, S., & Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb Cortex, 13*(4), 400-408.

Kim, H., Lee, D., & Jung, M. W. (2013). Signals for previous goal choice persist in the dorsomedial, but not dorsolateral striatum of rats. *J Neurosci, 33*(1), 52-63. doi: 10.1523/JNEUROSCI.2422-12.2013

Kim, H., Shimojo, S., & O'Doherty, J. P. (2011). Overlapping responses for the expectation of juice and money rewards in human ventromedial prefrontal cortex. *Cerebral cortex (New York, N.Y. : 1991), 21*(4), 769-776. doi: 10.1093/cercor/bhq145

Kim, H., Sul, J. H., Huh, N., Lee, D., & Jung, M. W. (2009). Role of striatum in updating values of chosen actions. *The Journal of neuroscience, 29*(47), 14701-14712.

Kimchi, E. Y., Torregrossa, M. M., Taylor, J. R., & Laubach, M. (2009). Neuronal correlates of instrumental learning in the dorsal striatum. *Journal of Neurophysiology, 102*(1), 475-489.

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science, 302*(5648), 1181-1185. doi: 10.1126/science.1088545

Konidaris, G., Scheidwasser, I., & Barto, A. G. (2012). Transfer in reinforcement learning via shared features. *The Journal of Machine Learning Research, 13*(1), 1333-1371.

Koster-Hale, J., & Saxe, R. (2013). Theory of mind: a neural prediction problem. *Neuron, 79*(5), 836-848. doi: 10.1016/j.neuron.2013.08.020

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience, 13*(10), 1292-1298. doi: 10.1038/nn.2635

Krajbich, I., Camerer, C., Ledyard, J., & Rangel, A. (2009). Using neural measures of economic value to solve the public goods free-rider problem. *Science, 326*(5952), 596-599. doi: 10.1126/science.1177302

Kriegeskorte, N. (2009). Relating Population-Code Representations between Man, Monkey, and Computational Models. *Frontiers in Neuroscience, 3*(3), 363-373. doi: 10.3389/neuro.01.035.2009

Kriegeskorte, N., Cusack, R., & Bandettini, P. (2010). How does an fMRI voxel sample the neuronal activity pattern: compact-kernel or complex spatiotemporal filter? *NeuroImage, 49*(3), 1965-1976. doi: 10.1016/j.neuroimage.2009.09.059

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America, 103*(10), 3863-3868. doi: 10.1073/pnas.0600244103

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 2*, 4-4. doi: 10.3389/neuro.06.004.2008

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience, 12*(5), 535-540. doi: 10.1038/nn.2303

Kringelbach, M. L., & Rolls, E. T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Progress in Neurobiology, 72*(5), 341-372. doi: 10.1016/j.pneurobio.2004.03.006

Kumaran, D., & Maguire, E. A. (2009). Novelty signals: a window into hippocampal information processing. *Trends Cogn Sci, 13*(2), 47-54. doi: 10.1016/j.tics.2008.11.004

Kwok, C., & Fox, D. (2004). Reinforcement learning for sensing strategies. *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*: IEEE.

Lara, A. H., & Wallis, J. D. (2014). Executive control processes underlying multi-item working memory. *Nature Neuroscience, advance on*. doi: 10.1038/nn.3702

Lau, B., & Glimcher, P. W. (2007). Action and outcome encoding in the primate caudate nucleus. *J Neurosci, 27*(52), 14502-14514. doi: 27/52/14502 [pii]

10.1523/JNEUROSCI.3060-07.2007

Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural Computations Underlying Arbitration between Model-Based and Model-free Learning. *Neuron, 81*(3), 687-699. doi: 10.1016/j.neuron.2013.11.028

Levy, D. J., & Glimcher, P. (2011). Comparing Apples and Oranges: Using Reward-Specific and Reward-General Subjective Value Representation in the Brain. *Journal Of Neuroscience, 31*(41), 14693-14707. doi: 10.1523/JNEUROSCI.2218-11.2011

Levy, I., Lazzaro, S. C., Rutledge, R. B., & Glimcher, P. (2011). Choice from Non-Choice: Predicting Consumer Preferences from Blood Oxygenation Level-Dependent Signals Obtained during Passive Viewing. *Journal Of Neuroscience, 31*(1), 118-125. doi: 10.1523/JNEUROSCI.3214-10.2011

Li, J., Schiller, D., Schoenbaum, G., Phelps, E. A., & Daw, N. D. (2011). Differential roles of human striatum and amygdala in associative learning. *Nat Neurosci, 14*(10), 1250-1252. doi: 10.1038/nn.2904

Libby, W. L., Jr., Lacey, B. C., & Lacey, J. I. (1973). Pupillary and cardiac activity during visual attention. *Psychophysiology, 10*(3), 270-294.

Liese, F., & Miescke, K. (2008). *Statistical Decision Theory*: Springer.

Lin, A., Adolphs, R., & Rangel, A. (2011). Social and monetary reward learning engage overlapping neural substrates. *Social cognitive and affective neuroscience, 7*(3), 274-281. doi: 10.1093/scan/nsr006

Littman, M. (1996). *Algorithms for Sequential Decision Making.*

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature, 412*(6843), 150-157.

Louie, K., & Glimcher, P. W. (2012). Efficient coding and the neural representation of value. *Annals of the New York Academy of Sciences, 1251*(1), 13-32. doi: 10.1111/j.1749-6632.2012.06496.x

Machens, C. K., Romo, R., & Brody, C. D. (2005). Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science, 307*(5712), 1121-1124. doi: 10.1126/science.1104171

Machens, C. K., Romo, R., & Brody, C. D. (2010). Functional, but not anatomical, separation of "what" and "when" in prefrontal cortex. *Journal Of Neuroscience, 30*(1), 350-360. doi: 10.1523/JNEUROSCI.3276-09.2010

MacKay, D. (2003). *Information Theory, Inference, and Learning Algorithms*: Cambridge University Press.

Mackey, S., & Petrides, M. (2010). Quantitative demonstration of comparable architectonic areas within the ventromedial and lateral orbital frontal cortex in the human and the macaque monkey brains. *The European journal of neuroscience, 32*(11), 1940-1950. doi: 10.1111/j.1460-9568.2010.07465.x

Mai, J. k., Paxinos, G., & Voss, T. (2008). *Atlas of the Human Brain* (3rd edition ed.): Elsevier.

Maldjian, J. A., Laurienti, P. J., Kraft, R. A., & Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage, 19*(3), 1233-1239.

Marr, D. C., & Poggio, T. (1976). From understanding computation to understanding neural circuitry. *Neurosciences Research Progress Bulletin, 15*(3), 470-488.

McClure, S. M., Berns, G., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron, 38*(2), 339-346.

McKerns, M. M., Strand, L., Sullivan, T., Fang, A., & Aivazis, M. A. G. (2011). *Building a Framework for Predictive Science.*

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

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*, 167-202. doi: 10.1146/annurev.neuro.24.1.167

Miller, M. I., Beg, M. F., Ceritoglu, C., & Stark, C. (2005). Increasing the power of functional maps of the medial temporal lobe by using large deformation diffeomorphic metric mapping. *Proc Natl Acad Sci U S A, 102*(27), 9685-9690. doi: 10.1073/pnas.0503892102

Milner, B. (1963). Effects of different brain lesions on card sorting: The role of the frontal lobes. *Archives of Neurology, 9*(1), 90-00.

Misaki, M., Kim, Y., Bandettini, P., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *NeuroImage, 53*(1), 103-118.

Mitchell, T. M. (1997). *Machine Learning*. New York, New York, USA: McGraw-Hill.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing Atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.

Montague, P. R., & Berns, G. (2002). Neural Economics and the Biological Substrates of Valuation. *Neuron, 36*(2), 265-284.

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Sciences, 16*(1), 72-80. doi: 10.1016/j.tics.2011.11.018

Mukherjee, S., Golland, P., & Panchenko, D. (2003). Permutation Tests for Classification. *Journal of Machine Learning Research, 1*, 1-48.

Mur, M., Bandettini, P., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI--an introductory guide. *Social cognitive and affective neuroscience, 4*(1), 101-109. doi: 10.1093/scan/nsn044

Murray, E. A. (2007). The amygdala, reward and emotion. *Trends Cogn Sci, 11*(11), 489-497. doi: 10.1016/j.tics.2007.08.013

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. (2011). Encoding and decoding in fMRI. *NeuroImage, 56*(2), 400-410. doi: 10.1016/j.neuroimage.2010.07.073

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J.-B. (2005). Valid conjunction inference with the minimum statistic. *NeuroImage, 25*(3), 653-660. doi: 10.1016/j.neuroimage.2004.12.005

Nicotra, A., Critchley, H. D., Mathias, C. J., & Dolan, R. J. (2006). Emotional and autonomic consequences of spinal cord injury explored using functional brain imaging. *Brain, 129*(Pt 3), 718-728. doi: 10.1093/brain/awh699

Nishimoto, S., Vu, A., & Gallant, J. (2010). Decoding human visual cortical activity evoked by continuous time-varying movies. *Journal of Vision, 9*(8), 667-667. doi: 10.1167/9.8.667

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences, 10*(9), 424-430. doi: 10.1016/j.tics.2006.07.005

O'Doherty, J., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron, 38*(2), 329-337.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science, 304*(5669), 452-454. doi: 10.1126/science.1094285

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. doi: 10.1016/j.conb.2004.10.016

O'Doherty, J. P. (2007). Lights, camembert, action! The role of human orbitofrontal cortex in encoding stimuli, rewards, and choices. *Annals of the New York Academy of Sciences, 1121*, 254-272.

O'Doherty, J. P. (2011). Contributions of the ventromedial prefrontal cortex to goal-directed action selection. *Annals of the New York Academy of Sciences, 1239*(1), 118-129. doi: 10.1111/j.1749-6632.2011.06290.x

O'Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Ann N Y Acad Sci, 1104*, 35-53. doi: 10.1196/annals.1390.022

O'Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences, 1104*, 35-53. doi: 10.1196/annals.1390.022

O'Doherty, J. P., Kringelbach, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience, 4*(1), 95-102. doi: 10.1038/82959

O'Reilly, J. X., Jbabdi, S., & Behrens, T. E. (2012). How can a Bayesian approach inform neuroscience? *Eur J Neurosci, 35*(7), 1169-1179. doi: 10.1111/j.1460-9568.2012.08010.x

O’Doherty, J. P., Lee, S. W., & McNamee, D. (2014). The structure of reinforcement-learning mechanisms in the human brain. *Current Opinion in Behavioral Sciences*. doi: 10.1016/j.cobeha.2014.10.004

O’Doherty, J. P., Lee, S. W., & McNamee, D. (2015). The structure of reinforcement-learning mechanisms in the human brain. *Current Opinion in Behavioral Sciences, 1*, 94-100. doi: 10.1016/j.cobeha.2014.10.004

Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., & Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America, 89*(13), 5951-5955.

Ongur, 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. doi: 10.1093/cercor/10.3.206

Op de Beeck, H. P. (2010). Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? *NeuroImage, 49*(3), 1943-1948. doi: 10.1016/j.neuroimage.2009.02.047

Padoa-Schioppa, C. (2011). Neurobiology of economic choice: a good-based model. *Annual Review of Neuroscience, 34*, 333-359. doi: 10.1146/annurev-neuro-061010-113648

Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature, 441*(7090), 223-226. doi: 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. *Nature Neuroscience, 11*(1), 95-102. doi: 10.1038/nn2020

Patil, A., Huard, D., & Fonnesbeck, C. J. (2010). PyMC: Bayesian Stochastic Modelling in Python. *Journal of statistical software, 35*(4), 1-81.

Pavlov, I. (1927). *Conditioned Reflexes*. London: Oxford University Press.

Payzan-LeNestour, E., & Bossaerts, P. (2011). Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. *PLoS Comput Biol, 7*(1), e1001048. doi: 10.1371/journal.pcbi.1001048

Payzan-Lenestour, E., & Bossaerts, P. (2012). Do not Bet on the Unknown Versus Try to Find Out More: Estimation Uncertainty and "Unexpected Uncertainty" Both Modulate Exploration. *Front Neurosci, 6*, 150. doi: 10.3389/fnins.2012.00150

Payzan-LeNestour, E., Dunne, S., Bossaerts, P., & O'Doherty, J. P. (2013). The neural representation of unexpected uncertainty during value-based decision making. *Neuron, 79*(1), 191-201. doi: 10.1016/j.neuron.2013.04.037

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87*(6), 532-552.

Penny, W. D., Holmes, A. P., & Friston, K. J. (2003). Random effects analysis. In R. S. J. Frackowiak, K. J. Friston, C. Frith, R. Dolan, C. J. Price, S. Zeki, J. Ashburner, & W. D. Penny (Eds.), (2nd ed.): Academic Press.

Pereira, F., & Botvinick, M. (2011). Information mapping with pattern classifiers: a comparative study. *NeuroImage, 56*(2), 476-496. doi: 10.1016/j.neuroimage.2010.05.026

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage, 45*(1 Suppl), S199-209. doi: 10.1016/j.neuroimage.2008.11.007

Picard, N., Matsuzaka, Y., & Strick, P. L. (2013). Extended practice of a motor skill is associated with reduced metabolic activity in M1. *Nature Neuroscience, 16*(9), 1340-1347. doi: 10.1038/nn.3477

Plassmann, H., O'Doherty, J. P., & Rangel, A. (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *Journal Of Neuroscience, 27*(37), 9984-9988. doi: 10.1523/JNEUROSCI.2131-07.2007

Poggio, T., & Ullman, S. (2013). Vision: are models of object recognition catching up with the brain? *Ann N Y Acad Sci, 1305*, 72-82. doi: 10.1111/nyas.12148

Preuschoff, K., & Bossaerts, P. (2007). Adding Prediction Risk To The Theory Of Reward Learning. *Ann N Y Acad Sci.*

Prevost, C., McCabe, J. A., Jessup, R. K., Bossaerts, P., & O'Doherty, J. P. (2011). Differentiable contributions of human amygdalar subregions in the computations underlying reward and avoidance learning. *Eur J Neurosci, 34*(1), 134-145. doi: 10.1111/j.1460-9568.2011.07686.x

Prévost, C., McNamee, D., Jessup, R. K., Bossaerts, P., & O'Doherty, J. P. (2013). Evidence for Model-based Computations in the Human Amygdala during Pavlovian Conditioning. *PLoS Computational Biology, 9*(2), e1002918-e1002918. doi: 10.1371/journal.pcbi.1002918

Rangel, A. (2013). Regulation of dietary choice by the decision-making circuitry. *Nature Neuroscience, 16*(12), 1717-1724. doi: 10.1038/nn.3561

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature reviews. Neuroscience, 9*(7), 545-556. doi: 10.1038/nrn2357

Rangel, A., & Clithero, J. A. (2014). The Computation of Stimulus Values in Simple Choice. In P. W. Glimcher, E. Fehr, C. Camerer, & R. A. Poldrack (Eds.), (pp. 125-147): Academic Press.

Rangel, A., & Hare, T. (2010). Neural computations associated with goal-directed choice. *Current Opinion in Neurobiology, 20*(2), 262-270. doi: 10.1016/j.conb.2010.03.001

Rescorla, R. A. (1972). A theory of Pavlovian conditioning: Variations in effectiveness of reinforcement and nonreinforcement. . In N.-Y. A. Century-Crofts (Ed.), *Classical conditioning II: Current research and Theory*: Black A.H, P.W.H.

Ribas-Fernandes, J. J., Solway, A., Diuk, C., McGuire, J. T., Barto, A. G., Niv, Y., & Botvinick, M. M. (2011). A neural signature of hierarchical reinforcement learning. *Neuron, 71*(2), 370-379. doi: 10.1016/j.neuron.2011.05.042

Ridderinkhof, K. R. (2014). Neurocognitive mechanisms of perception-action coordination: A review and theoretical integration. *Neurosci Biobehav Rev.* doi: 10.1016/j.neubiorev.2014.05.008

Robbins, T. W. (1996). Dissociating executive functions of the prefrontal cortex. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 351*(1346), 1463-1470; discussion 1470-1461. doi: 10.1098/rstb.1996.0131

Roesch, M. R., Calu, D. J., Esber, G. R., & Schoenbaum, G. (2010). Neural correlates of variations in event processing during learning in basolateral amygdala. *J Neurosci, 30*(7), 2464-2471. doi: 10.1523/JNEUROSCI.5781-09.2010

Rolls, E. T., Loh, M., Deco, G., & Winterer, G. (2008). Computational models of schizophrenia and dopamine modulation in the prefrontal cortex. *Nature reviews. Neuroscience, 9*(9), 696-709. doi: 10.1038/nrn2462

Rudebeck, P. H., Behrens, T. E., Kennerley, S. W., Baxter, M. G., Buckley, M. J., Walton, M. E., & Rushworth, M. F. S. (2008). Frontal cortex subregions play distinct roles in choices between actions and stimuli. *Journal Of Neuroscience, 28*(51), 13775-13785. doi: 10.1523/JNEUROSCI.3541-08.2008

Rushworth, M. F. S., & Behrens, T. E. J. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience, 11*(4), 389-397. doi: 10.1038/nn2066

Rushworth, M. F. S., Mars, R. B., & Summerfield, C. (2009). General mechanisms for making decisions? *Current Opinion in Neurobiology, 19*(1), 75-83. doi: 10.1016/j.conb.2009.02.005

Russell, S., & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach* (3rd ed.): Prentice Hall.

Samejima, K., Ueda, Y., Doya, K., & Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science, 310*(5752), 1337-1340.

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nat Neurosci, 16*(4), 486-492. doi: 10.1038/nn.3331

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology, 57*, 87-115. doi: 10.1146/annurev.psych.56.091103.070229

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593-1599.

Schwartenbeck, P., Fitzgerald, T., Dolan, R. J., & Friston, K. (2013). Exploration, novelty, surprise, and free energy minimization. *Front Psychol, 4*, 710. doi: 10.3389/fpsyg.2013.00710

Schwartenbeck, P., FitzGerald, T. H., Mathys, C., Dolan, R., & Friston, K. (2014). The Dopaminergic Midbrain Encodes the Expected Certainty about Desired Outcomes. *Cereb Cortex*. doi: 10.1093/cercor/bhu159

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics, 6*(2), 461-464.

Sescousse, G., Redouté, J., & Dreher, J.-C. (2010). The architecture of reward value coding in the human orbitofrontal cortex. *Journal Of Neuroscience, 30*(39), 13095-13104. doi: 10.1523/JNEUROSCI.3501-10.2010

Seymour, B., Daw, N., Dayan, P., Singer, T., & Dolan, R. (2007). Differential encoding of losses and gains in the human striatum. *J Neurosci, 27*(18), 4826-4831. doi: 10.1523/JNEUROSCI.0400-07.2007

Seymour, B., O'Doherty, J. P., Koltzenburg, M., Wiech, K., Frackowiak, R., Friston, K., & Dolan, R. (2005). Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nat Neurosci, 8*(9), 1234-1240. doi: 10.1038/nn1527

Shadlen, M. N., & Kiani, R. (2013). Decision making as a window on cognition. *Neuron, 80*(3), 791-806. doi: 10.1016/j.neuron.2013.10.047

Simon, D. A., & Daw, N. D. (2011). Neural correlates of forward planning in a spatial decision task in humans. *J Neurosci, 31*(14), 5526-5539. doi: 10.1523/JNEUROSCI.4647-10.2011

Spence, K. W. (1950). Cognitive versus stimulus-response theories of learning. *Psychological Review, 57*(3), 159-172.

Stalnaker, T. A., Calhoon, G. G., Ogawa, M., Roesch, M. R., & Schoenbaum, G. (2010). Neural correlates of stimulus-response and response-outcome associations in dorsolateral versus dorsomedial striatum. *Frontiers in integrative neuroscience, 4*, 12-12. doi: 10.3389/fnint.2010.00012

Stanley, D. A., & Adolphs, R. (2013). Toward a Neural Basis for Social Behavior. *Neuron, 80*(3), 816-826. doi: 10.1016/j.neuron.2013.10.038

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *Neuroimage, 46*(4), 1004-1017. doi: 10.1016/j.neuroimage.2009.03.025

Strait, C. E., Blanchard, T. C., & Hayden, B. Y. (2014). Reward Value Comparison via Mutual Inhibition in Ventromedial Prefrontal Cortex. *Neuron, 82*(6), 1357-1366. doi: 10.1016/j.neuron.2014.04.032

Sutton, R., & Barto, A. (1998). *Reinforcement Learning: An Introduction*: MIT Press.

Sutton, R. S., Precup, D., & Singh, S. (1999). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence, 112*(1), 181-211.

Symmonds, M., Bossaerts, P., & Dolan, R. J. (2010). A behavioral and neural evaluation of prospective decision-making under risk. *Journal Of Neuroscience, 30*(43), 14380-14389. doi: 10.1523/JNEUROSCI.1459-10.2010

Szczepanski, S. M., & Knight, R. T. (2014). Insights into Human Behavior from Lesions to the Prefrontal Cortex. *Neuron*. doi: 10.1016/j.neuron.2014.08.011

Tanaka, S. C., Balleine, B. W., & O'Doherty, J. P. (2008). Calculating consequences: brain systems that encode the causal effects of actions. *Journal Of Neuroscience, 28*(26), 6750-6755. doi: 10.1523/JNEUROSCI.1808-08.2008

Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science, 332*(6033), 1054-1059. doi: 10.1126/science.1196404

Tesauro, G. (1995). Temporal difference learning and TD-Gammon. *Communications of the ACM, 38*(3), 58-68. doi: 10.1145/203330.203343

Thirion, J. P. (1998). Image matching as a diffusion process: an analogy with Maxwell's demons. *Med Image Anal, 2*(3), 243-260.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review, 55*(4), 189-208.

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science, 315*(5811), 515-518. doi: 10.1126/science.1134239

Tremblay, L., & Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. *Nature, 398*(6729), 704-708. doi: 10.1038/19525

Tremblay, S., Pieper, F., Sachs, A., & Martinez-Trujillo, J. (2014). Attentional Filtering of Visual Information by Neuronal Ensembles in the Primate Lateral Prefrontal Cortex. *Neuron*. doi: 10.1016/j.neuron.2014.11.021

Tricomi, E., Balleine, B., & O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *European Journal of Neuroscience, 29*(11), 2225-2232. doi: 10.1111/j.1460-9568.2009.06796.x

Tusche, A., Bode, S., & Haynes, J.-D. (2010). Neural responses to unattended products predict later consumer choices. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 30*(23), 8024-8031. doi: 10.1523/jneurosci.0064-10.2010

Valentin, V. V., Dickinson, A., & O'Doherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human brain. *J Neurosci, 27*(15), 4019-4026.

Vercauteren, T., Pennec, X., Perchant, A., & Ayache, N. (2007). Non-parametric diffeomorphic image registration with the demons algorithm. *Med Image Comput Comput Assist Interv, 10*(Pt 2), 319-326.

von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*: Princeton University Press.

Wallis, J. D. (2007). Orbitofrontal cortex and its contribution to decision-making. *Annual Review of Neuroscience, 30*, 31-56. doi: 10.1146/annurev.neuro.30.051606.094334

Wallis, J. D. (2011). Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nature Neuroscience, 15*, 13-19. doi: 10.1038/nn.2956

Wallis, J. D., & Kennerley, S. W. (2011). Contrasting reward signals in the orbitofrontal cortex and anterior cingulate cortex. *Annals of the New York Academy of Sciences, 1239*, 33-42. doi: 10.1111/j.1749-6632.2011.06277.x

Walton, M. E., Behrens, T. E. J., Buckley, M. J., Rudebeck, P. H., & Rushworth, M. F. S. (2010). Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron, 65*(6), 927-939. doi: 10.1016/j.neuron.2010.02.027

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in neuroinformatics, 7*, 14-14. doi: 10.3389/fninf.2013.00014

Wilson, R. C., Nassar, M. R., & Gold, J. I. (2010). Bayesian online learning of the hazard rate in change-point problems. *Neural Computation, 22*(9), 2452-2476. doi: 10.1162/NECO_a_00007

Wilson, R. C., & Niv, Y. (2011). Inferring relevance in a changing world. *Frontiers in Human Neuroscience, 5*, 189-189. doi: 10.3389/fnhum.2011.00189

Wolpert, D. M., Diedrichsen, J., & Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nat Rev Neurosci, 12*(12), 739-751. doi: 10.1038/nrn3112

Wunderlich, K., Beierholm, U., Bossaerts, P., & O'Doherty, J. P. (2011). The human prefrontal cortex mediates integration of potential causes behind observed outcomes. *Journal of Neurophysiology*. doi: 10.1152/jn.01051.2010

Wunderlich, K., Dayan, P., & Dolan, R. J. (2012). Mapping value based planning and extensively trained choice in the human brain. *Nat Neurosci, 15*(5), 786-791. doi: 10.1038/nn.3068

Wunderlich, K., Rangel, A., & O'Doherty, J. P. (2009). Neural computations underlying action-based decision making in the human brain. *Proceedings of the National Academy of Sciences of the United States of America, 106*(40), 17199-17204. doi: 10.1073/pnas.0901077106

Yacubian, J., Glascher, J., Schroeder, K., Sommer, T., Braus, D. F., & Buchel, C. (2006). Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. *J Neurosci, 26*(37), 9530-9537. doi: 10.1523/JNEUROSCI.2915-06.2006

Yang, T., & Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature, 447*(7148), 1075-1080. doi: 10.1038/nature05852

Yassa, M. A., & Stark, C. E. (2009). A quantitative evaluation of cross-participant registration techniques for MRI studies of the medial temporal lobe. *NeuroImage, 44*(2), 319-327. doi: 10.1016/j.neuroimage.2008.09.016

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur J Neurosci, 19*(1), 181-189.

Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. *Eur J Neurosci, 22*(2), 513-523.

Yu, A. J., & Dayan, P. (2003). Expected and unexpected uncertainty. *Advances in Neural Information Processing Systems 15*.