

# MULTIPLE FORMS OF VALUATION IN THE HUMAN BRAIN

Thesis by

Klaus Wunderlich

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2010

(Defended November 9<sup>th</sup>, 2009)

© 2010

**Klaus Wunderlich**

**All Rights Reserved**

# Acknowledgements

I would like to thank my advisor and mentor John O’Doherty. I have learned a great deal from John about doing science and being a researcher and I am very grateful that I had the opportunity to be one of his apprentices. He has been a dedicated advisor, always full of actionable ideas, motivating, cheerful, and supportive. Even after he started his new lab in Dublin, John was always approachable and knew how to keep me encouraged when times were tough or when I was stuck. I would like to thank my co-advisors and collaborators Antonio Rangel and Peter Bossaerts. Antonio worked with us on my value studies and Peter helped with the Integration project and taught me about Bayesian decision theory. I also wish to acknowledge my other PhD committee members Ralph Adolphs, Richard Andersen, and Shinsuke Shimojo for their time and thoughtful advice.

Thanks to all current remaining members of the Olab, Jan Glaescher and Vikram Chib, and former members Ulrik Beierholm, Alan Hampton, Signe Bray, Vivian Valentin, Hackjin Kim, Eebie Tricomi and Tom Schonberg. They all have helped me over the years in one or the other ways, from minor technical problems, to bringing me up to speed on the latest methods. Research in John’s lab was not only very productive, but also collegial, and fun!

The Gordon and Betty Moore Foundation provided generous financial support throughout my graduate career, for which I will always be grateful. My Fellowship enabled me to explore different avenues of research and provided some extra degree of freedom during my graduate studies.

I have been lucky to work in a wide variety of research areas while at Caltech before settling in on the work presented in this thesis. All of my collaborators have influenced my growth as a researcher. In particular I would like to thank my rotation advisors Christof Koch and Erin Schuman. In Erin's lab I spent one year and had the rare chance to learn hands-on about recording from single cells in awake human patients.

Two professors from my undergraduate days in Germany had a particularly significant impact on my academic career. Michael Waldmann provided me with my first opportunities for research as an undergraduate in Psychology and encouraged me to follow my scientific curiosity from very early on in my studies. Stefan Treue introduced me to research in cognitive neuroscience and inspired my choice of graduate studies in the US.

I especially want to thank my parents Peter and Barbara. They have always supported me and I am grateful for all what they have done for me. I think it wasn't always easy for them with me being so far away in distance. And to my dear Jenny, for her love and all the shared experiences

I am very grateful that I had the chance to study at Caltech and I feel that the time here has influenced me for my life. Students at Caltech have the freedom to pursue cutting edge science, and they will find intellectual peers among their colleagues who are equally as curious and motivated as they are. Every effort is made to remove distractions from the lives of students and allow them to focus on research and learning. There are undoubtedly few institutions in the world, which can offer so much to students.

# Abstract

Our lives are defined by the decisions we make, often involving choices between different actions or goods. An important open problem in decision neuroscience is: what value signals are used in guiding the different types of choices, where are they stored in the brain, and how does the brain compare them to make a choice. We used fMRI in human subjects to address these questions in a variety of different choice settings: decisions between actions, economic choices, and more complex hierarchical decisions. We found evidence for a separate representation of two main forms of value signals in the human brain: precursors of choice, such as signals relating to the value of each available action or stimulus, and signals reflecting the consequence of the decision process by encoding the expected value of the option that is subsequently chosen. On the precursor side, we found action value signals in the supplementary motor cortex and stimulus value signals in the medial prefrontal cortex. Separate brain regions, most prominently the ventromedial prefrontal cortex, were involved in encoding the value of the chosen action or stimulus. Importantly, we found value chosen signals in stimulus decisions even when no actions were associated with choosing the stimuli, providing evidence for the hypothesis that the brain doesn't need the motor system to make such decisions but is capable of making economic choices completely within an abstract representation of goods. Furthermore, in action decisions, we found that activity in the dorsomedial frontal cortex resembles the output of a decision comparator, implicating this region in the computation of the decision itself. In a real world setting where multiple stimuli could potentially influence outcomes, an individual may consider a number of theories about which features are relevant for giving reward. We found that decision variables based on simultaneous integration of all evidences were better able to explain subjects' behavior and activity in the prefrontal cortex than those generated by an attention-gated approach, i.e., by first picking the theory that is most likely correct and then choosing accordingly. These results demonstrate that the human brain is capable of optimally integrating information, similar to an ideal Bayesian observer.

# Table of Contents

Acknowledgements .....	iii
Abstract.....	v
Table of Contents.....	vi
List of Figures.....	vii
List of Tables .....	viii
Nomenclature.....	ix
Chapter 1. Introduction.....	1
Decision making.....	1
Value signals in the brain .....	9
Chapter 2. Action-based decision making.....	18
Introduction .....	19
Results .....	23
Discussion.....	29
Methods.....	34
Supplementary Methods.....	43
Chapter 3. Economic choices .....	67
Introduction .....	68
Results .....	71
Discussion.....	73
Methods.....	75
Chapter 4. Optimal integration of multiple evidences.....	89
Report .....	90
Supplementary Materials.....	104
Methods.....	107
Chapter 5. Summary.....	126
Bibliography .....	132

# List of Figures

Figure 1.1 Model of decision making.....	16
Figure 1.2 Decision variables .....	17
Figure 2.1 Experimental Design and Behavior .....	38
Figure 2.2 Action values.....	39
Figure 2.3 Chosen values.....	40
Figure 2.4 Value comparison.....	41
Figure 3.1 Experimental Design and Behavior.....	84
Figure 3.2 Neural correlates of chosen value .....	85
Figure 3.3 Neural correlates of stimulus value .....	87
Figure 4.1 Task and Behavior.....	100
Figure 4.2 Activity reflecting the Integration model.....	102
Figure 4.3 Bayesian model comparison .....	103
Figure 5.1 Summary of value signals in the human brain .....	131
Figure S 2.1 Activations by action values with T-stats coloring.....	60
Figure S 2.2 Post-hoc effect sizes.....	61
Figure S 2.3 Local predominance bias .....	63
Figure S 2.4 Illustration of the competition difference model .....	64
Figure S 2.5 Illustration of the Drift Diffusion Model .....	65
Figure S 2.6 CDM steady state output.....	66
Figure S 4.1 Value chosen of the integration model .....	122
Figure S 4.2 Activity modulated by the attention-gated model.....	123
Figure S 4.3 Across-dimension certainty signal of the attention model.....	124
Figure S 4.4 Value – probability transformation.....	125

# List of Tables

Table 2.1 Different types of value signals .....	56
Table 2.2 Activated regions.....	57
Table 2.3 Model performance .....	59
Table 3.1 Activated regions value chosen SC .....	88
Table 3.2 Activated regions value chosen AC .....	88
Table 3.3 Activated regions stimulus values .....	88
Table 4.1 BIC model fit.....	119
Table 4.2 Model cross-validation .....	120
Table 4.3 Activated regions.....	121



# Nomenclature

AC	action condition
ACC	anterior cingulate cortex
BOLD	blood oxygenation level dependant
EPI	echo-planar imaging
fMRI	functional Magnetic Resonance Imaging
LOO	leave one out
PFC	prefrontal cortex
SC	stimulus condition
OFC	orbitofrontal cortex
voxel	3D pixel

# Chapter 1. Introduction

## Decision making

Our lives are defined by the decisions we make. Humans evolved eons ago from other primates who lived in small groups and spent most of their waking hours foraging for a livelihood. When not searching for something to eat or drink, we were protecting ourselves from predators, selecting mates, and looking for safe places to live. Our success in accomplishing these tasks, crucial for survival, did neither arise due to particularly sharp senses nor to especially powerful physical endurance. Instead, we dominate this planet today because of our distinctive capacity for good decision-making. This skill has allowed us to leave the planet for brief periods of time, but has also permitted us to develop technologies and weapons that could render the planet uninhabitable if we make a few really bad decisions. As we made it to where we are today, it appears that human beings have an exceptional ability to choose appropriate means to achieve their ends.

Thinking about the question of how this amazing process works is thus one of the fascinating problems we need to solve in order to understand ourselves: what are the neural correlates underlying our valuation processes and which computations take place in our neural machinery when we make decisions. In this thesis I will try to shed light on some aspects of how the brain represents values that are used in the decision process and how it uses and compares these values in order to form a choice.

The decision-making capacity in the brain has been polished through natural selection to provide a versatile system that can quickly adapt to our environment and changing situations, with bad decisions punished in a dramatic manner. As the philosopher Willard Van Orman Quine [1] once commented: “Creatures inveterately wrong in their inductions have a pathetic but praiseworthy tendency to die before reproducing their kind”, in other words, animals that make bad predictions of the future tend to die before they can pass their genes on to the next generation. Thus it is clear that while the mechanism might not necessarily be best equipped to solve hypothetical math problems, it was shaped to be best adaptive to a variety of daily life situations ranging from foraging strategies to food selection.

Valuations of alternative options and decisions are even more ubiquitous in our modern day life. We constantly make choices either between different actions – think about playing a match of tennis – or between different goods, such as when shopping for groceries. A decision occurs every time when an organism is confronted by several discrete options, evaluates the merits of each, and finally selects one to pursue. While perceptual decisions, such as discrimination between different objects or detecting the motion direction of moving dot patterns, are exclusively based on objective sensory characteristic of the options, we were particularly interested in the mechanism of value-based decisions: those choices that are mandated by the subjective preference and experience of the individual. From the perspective of an outside observer, the decision-making mechanisms in animals or humans can be seen as a black box, mediating the sensory perception of available alternatives on the input side and a motor response on the output side (Figure 1.1).

### *Models of decision making*

Since the time of the behaviorists [2], scientists have aimed at explaining the processes underlying decision making by constructing theoretical models and then testing them on animals' or humans' behavior.

Decision making is commonly seen as a two-stage process. A learning process that works continuously throughout time and constantly evaluates and updates preferences for the available options. Each time we gain new experiences, the values associated with stimuli and actions are updated with the new information. The second part of the decision-making module works at every instance of a choice to select the option which is most desirable, i.e., the one that has the highest subjective value. The idea of two separate but connected decision modules first appeared in one of Barto's [3] early reinforcement learning papers, in which he showed that learning action-outcome relationships can be solved by a computational system with two neuron-like elements. This idea was then later developed further into the full Actor/Critic model of action selection. In this model, one of the units, the critic element, constructs the evaluation of different states using a temporal difference learning rule and the second, the associative search element, selects the correct action at each stage.

### *Evaluation*

The Rescorla-Wagner rule [4] is one of the most influential model in animal learning. It describes how predictive values of currently present conditioned stimuli are updated for the future by a discounted difference between the current prediction and the experienced outcome. Without doubt, one reason for its popularity is the extreme simplicity of the model despite its ability to explain most of the observed behavior during Pavlovian

conditioning, such as blocking, overshadowing, or conditioned inhibition. The most significant idea of the model is the postulation of a prediction error, i.e., learning occurs always when the outcome is different from the expectation. The greater this deviation is from the prediction, the bigger the amount of information that is learnt in any instance.

The Rescorla-Wagner learning rule describes the process of value acquisition in a trial-by-trial fashion. Over learning, the predictive stimulus acquires value  $V_i$ , where  $i$  is the  $i$ -th trial.

$$V_i = V_{i-1} + \alpha\delta$$

where  $\alpha$  is the learning rate, and  $\delta = R_i - V_{i-1}$  is the prediction error.

In this model, the prediction error signal is generated at the time of the expected outcome, and influences the value of the cue on the subsequent trial, discounted by the learning rate.

Though the Rescorla-Wagner model is suitable for explaining many basic forms of learning, alternative and more complex updating models were developed in later years. The most prominent out of these is probably temporal difference learning, which takes the expected value of outcomes not only at a single point but at multiple instances in the future into account [5].

While the original Rescorla-Wagner model was originally formulated for Pavlovian-type conditioning, it can also be used as an essential element to model goal-directed choice behavior. The model creates two decision variables for each stimulus: an expected value signal and a deviation of the outcome from the prediction. It is extremely useful to have a

representation of expected values for each individual option before an animal engages in any actions in order to obtain reward [6]. Consider the case of action decisions, where animals or humans learn to associate values with taking each action. The action value  $Q$  of action  $a$  is then

$$Q(a) = E(r|a)$$

where  $E$  is the expected reward given action  $a$ . These action values can be learned through a mechanism based on difference learning like the RW model described above. If action values are available at the time of choice for all possible actions under consideration, the obvious choice is to take the action that yields the highest value  $Q(a)$ . Following valuation we then need a second module that selects an action based on a comparison of these values.

### *Selection*

The best way to select an action is to find the subjectively optimal action out of all available alternatives at any given time (based on the individual preferences). This is a very difficult problem in the case that actions affect long-term outcomes, which are only reached through a cascade of different actions. In this case it is quite complex to recognize the causal relationship between action and specific outcome, a problem called credit assignment problem [5] in the literature.

One possible method of value comparison is the calculation of a value difference, a computation for which we will propose and test a potential neuronal implementation in Chapter 2.

The values are used as inputs to the decision process along with information about the current state of the individual, such as for instance, hunger. If the animal is satiated, the value of a normally very highly regarded food item is discounted and becomes less desirable [7].

### *Decision variables*

The trial-by-trial value signals that were predicted by the computational learning model can be used as a proxy for the values of the subject's true internal decision variables (Figure 1.2). If the model explains the value representation and states of the subjects' internal computations, we should be able to localize the neural correlates of the subject's decision variables using the traditional correlative techniques of systems neuroscience.

### *Action versus Stimulus based decision model*

It is a well-established belief among economists, psychologists, and neuroscientists that the brain solves decisions among different goods, such as when choosing between pizza and sandwich, much like it does action decisions by also first computing a value for each alternative and then selecting the one that has the highest value [5, 6, 8, 9]. However, neuroscientists have considered two possible alternative ways for how values might be compared to make a choice in these situations: by assigning values to all available goods (stimulus values) and then comparing these values directly in the space of stimuli, or instead by first transferring these values to the associated actions and then comparing those action values in action-space.

According to the goods-based model, economic choice is an independent cognitive module, with the actual neuronal processes underlying the decision taking place in a

space where goods are represented as such. This assumes a high level of mental abstraction. The available items are represented as objects in a space of goods, which is completely independent of representations about the sensory environment and the motor actions that would be required to obtain them. Values are then assigned to the individual goods. Of course, these values don't necessarily have to be fixed for each object but could also be described as a function of the subjects current mental and physiological state and thereby be dependant on the subjects desire for this object. Such abstract values do indeed exist in the primate orbitofrontal cortex [10], a prerequisite but not sufficient condition for the goods-based model. The key feature of the goods-based model is that economic values take place entirely in goods space. Then, once one good is chosen, the individual plans and executes a suitable motor action to implement the choice, without further valuation. Note that in this model the process of action selection is independent of stimulus selection and follows the evaluation between the desirability of the stimuli. It is merely an all-or-none process to select an appropriate action that helps to gain the good.

In contrast, in the action-based model, economic choices are embedded in premotor processes of action selection. Several learning models are variations of this idea, with the first references dating back to the associative behaviorist learning theories. In these early accounts [2], the behavior of the animal is described purely in terms of stimulus response associations and the choice problem is reduced to associative learning. During training, the animal learns the association between stimuli and rewarded motor response. In a more recent model driven by Paul Glimcher [11], values are learned through experience through reinforcement learning. Economic choices thus unfold as a process of action selection. Brain areas and neuronal populations responsible for action selection (e.g.,



LIP) represent a common pathway for different types of decision-making. The same neural hardware is used to process multiple kinds of valuations, and economic choice becomes a fundamental choice between actions. Action-based models have traditionally been more prominent among neuroscientists as they are built on theories of reinforcement learning, and provide a uniform model for universal problem solving which is flexible and adaptable.

However, planning and controlling movement is computationally very costly, and doing it every time even if no action is necessary at a stage would be a rather inefficient implementation. In contrast, a modular structure is more efficient because it breaks this process down into two separate components: choosing and moving. An important fact that favors the goods-based model is that lesions to parietal areas like LIP, areas associated with action-based decision making, do not typically influence economic choice per se. Parietal lesions rather produce visuo-spatial deficits such as hemi-neglect and Balint's syndrome [12]. However, economic choices are disrupted by OFC lesions, suggesting that motor areas are not strictly necessary for making economic choices, although they are required for implementing them.

Though we have proof for abstract representations of goods [10], it has yet not been shown that entire choice process takes place in this space. In principle, separating in time the choice between goods and the selection of action can test the two models. This is what we did in our second experiment.

*Measuring decision variables with fMRI*

In recent years, learning models have been applied to numerous neurobiological problems. One of the pioneers was Wolfram Schultz who used single-unit recording in the midbrain of awake monkeys and recorded from dopamine neurons while monkeys were trained in Pavlovian conditioning [13].

The trial-by-trial value signals that were predicted by the computational learning model can be used as a proxy for the values of the subject's true internal decision variables. If the model explains the value representation and states of the subject's internal computations, we should be able to localize the neural correlates of the subject's decision variables using the traditional correlative techniques of systems neuroscience.

## **Value signals in the brain**

The very subjective nature of value-based decisions means that different individuals have distinctive preferences for a group of available options. These preferences of the decision maker, the decision variables, are quantities internal to the subject's decision process; summarizing properties of the available behavioral options relevant to guiding choice. Decision variables can be thought of as linked to and guiding the process of option evaluation and action selection. There are a number of potential decision variables that might be used by animals and humans to form choices. It is now a well-established belief among economists, psychologists, and neuroscientists that the brain solves choice problems by first computing a value for each alternative and then selecting the one that

has the highest value [5, 6, 8, 9]. In addition to value signals there are prediction error signals used to update the value signal in future trials [14]. Other variables proposed in addition to more sophisticated decisions are risk and ambiguity [15].

In order to effectively compare different options and finally settle on a decision among them, it is necessary to have several different values represented along this process (Figure 1.2).

Action or stimulus values (together called option values) encode the value of each action or stimulus prior to choice and regardless of whether they are subsequently chosen or not. This means that individual option values of each action or stimulus need to be simultaneously represented in the brain to serve as input into the decision-making process [16-18]. These values are then compared in order to generate a choice. Finally, the value of the option that is selected, known as the chosen value, is tracked in order to be able to do reinforcement learning. In particular, by comparing the value of the outcome generated by the decision of the chosen value, the organism can compute a prediction error signal that can be used to update the action value of the chosen option. Note that while the option values are computed before the decision is made, the chosen value and outcome of the comparator process signals are computed afterwards.

Multiple value representation has already been found in the human and primate brain and we will review some evidence for value signals in various brain areas:

#### *Orbitofrontal and medial prefrontal cortex*

Patients with dementia or lesions in the OFC show impairment in their choices affecting a variety of areas. Fronto-temporal dementia can cause eating disorders, as these patients

may assign wrong values to appetitive stimuli [19]. In some laboratory tasks, OFC lesion patients are impaired in gambling tasks [20, 21], which have been commonly attributed to the impairment in judging risk. They also show inconsistencies in preference judgement, and they make inconsistent choices [22] and poor choices in the ultimatum game [23]. The case of Phineas Gage demonstrates that these impairments also extend to the social domain and while they are hardly noticeable in other cognitive standard tests such as IQ tests, they have a dramatic influence in the ability to lead a successful life [24].

While findings from these lesioned patients have demonstrated a contributing role of OFC to choice settings, more recent imaging and monkey electrophysiology studies have shown that neurons in OFC directly correlate with decision variables. Even without any choices, OFC activity encodes subjective preferences, as shown in an imaging experiment in which OFC activity was higher in response to direct pleasant unconditioned reinforcers such as taste as compared to neutral ones [25], and an electrophysiology study in which single cells encoded an amount of juice [26]. More interestingly, OFC seems to directly encode values in choice tasks. In an experiment by Padoa-Schioppa [10], monkeys choose between two different juices. Given in equal amounts, the thirsty monkeys had a strong preference for one juice but if the amount of the less-preferred juice was just enough increased then monkeys chose that juice. This paradigm allowed calculating the monkeys' indifference point and inferring their relative value of the juices to be used as proxy for the monkeys' subjective decision variable. Neurons that encoded the value of the chosen juice in a trial were frequently found in OFC. Interestingly, there were other neurons in OFC that encoded the offer value, a neuron that covaried in activity with the value of any one juice irrespective of whether

the monkey later chose it or not. Importantly, all neuronal responses were independent of visual or motor responses.

Interestingly, some of the neurons that encoded the offer value of the juice were menu invariant, meaning that their activity was not modulated by the other choices that were available at any given time [27].

### *Parietal cortex*

The parietal cortex has long been implicated in guiding attention [28] and linking the sensory with the motor system [29]. It would make a lot of sense to guide attention by valuation because this would allow us attend to those features in the world that are important to us. It would therefore be very plausible to find value representations in those systems of the brain. Indeed, the parietal cortex was one of the first areas where cells encoding value-dependant signals were found in awake monkeys.

Platt and Glimcher pioneered this research by recording from LIP neurons during a task in which they varied the amount of juice that they gave the monkeys. They found signals related to expected value during both tasks in which the monkeys were passive receivers of juice and not permitted to make any decisions [30] (similar to the study by Wallis in OFC), as well as in subsequent experiments in which monkeys were allowed to choose freely [31]. In subsequent tasks, the firing frequency of neurons was also similarly modulated by value in a foraging task [32] and a matching task [33].

However, the value signals that have been found in lateral intraparietal cortex (LIP) during saccadic action-based choice [33, 34] are also not pure action values since they are

strongly modulated by whether an action is subsequently taken. This suggests that instead of serving as inputs to the comparison process, they might reflect its output.

### *Midbrain, striatum*

Numerous evidences exist for learning-related signals in the striatum. These include value signals [35] for the chosen action as well as action-specific reward values for hand [17] and eye movements [36]. Striatum encodes chosen value also in non primates such as rats [37].

O’Doherty et al. [38] scanned human participants with functional magnetic resonance imaging while they engaged in instrumental conditioning. Their results suggest the encoding of signals related to predicting future reward in the ventral striatum, and information about the rewarding outcomes of actions in dorsal striatum, relating to the critic and actor in RL models.

Electrophysiological studies in non-human primates implicate the phasic firing of midbrain dopaminergic neurons in encoding reward-prediction errors [39]. FMRI studies of human learning have found evidence of reward-prediction error-related activity in known projection sites of dopaminergic cells, especially the ventral striatum, during learning with other forms of natural and abstract rewards such as juice or money [38, 40, 41] or faces [42]. The BOLD signature of the prediction error is more pronounced in learners compared to subjects who don’t learn on a task [43], providing evidence that the measured brain signal is in fact linked to behavior.

*Amygdala*

Schoenbaum et al. [44] examined neural activity in rat orbitofrontal cortex and basolateral amygdala during instrumental learning in an olfactory discrimination task. Neurons in both regions fired selectively during the anticipation of rewarding or aversive outcomes. This selective activity emerged early in training, before the rats had learned reliably to avoid the aversive outcome. The results support the concept that the basolateral amygdala and orbitofrontal cortex cooperate to encode information that may be used to guide goal-directed behavior.

Gottfried et al. [45] used reinforcer devaluation while measuring neural activity with functional magnetic resonance imaging in human subjects. They found that in the amygdala and orbitofrontal cortex, responses evoked by a predictive target stimulus were decreased after devaluation, whereas responses to the nondevalued stimulus were maintained. Thus, differential activity in the amygdala encodes the current value of reward representations accessible to predictive cues. Paton et al. [46] recorded the activity of individual amygdala neurons in monkeys while abstract images acquired either positive or negative value through conditioning. After monkeys had learned the initial associations, they reversed image value assignments and examined neural responses in relation to these reversals in order to estimate the relative contribution to neural activity of the sensory properties of images and their conditioned values. They show that changes in the values of images modulate neural activity, and that this modulation occurs rapidly enough to account for, and correlate with, monkeys' learning. Furthermore, distinct populations of neurons encoded the positive and negative values of visual stimuli.

A number of studies point to the direction that amygdala and OFC are tightly linked in the function of expected reward presentation. In a further study Schoenbaum [47] provides direct neurophysiological evidence of this cooperative function. They recorded from OFC in intact and ABL-lesioned rats learning odor discrimination problems. As rats learned these problems, they found that lesioned rats exhibited marked changes in the information represented in OFC during odor cue sampling. Lesioned rats had fewer cue-selective neurons in OFC after learning; the cue-selective population in lesioned rats did not include neurons that were also responsive in anticipation of the predicted outcome; and the cue-activated representations that remained in lesioned rats were less associative and more often bound to cue identity.



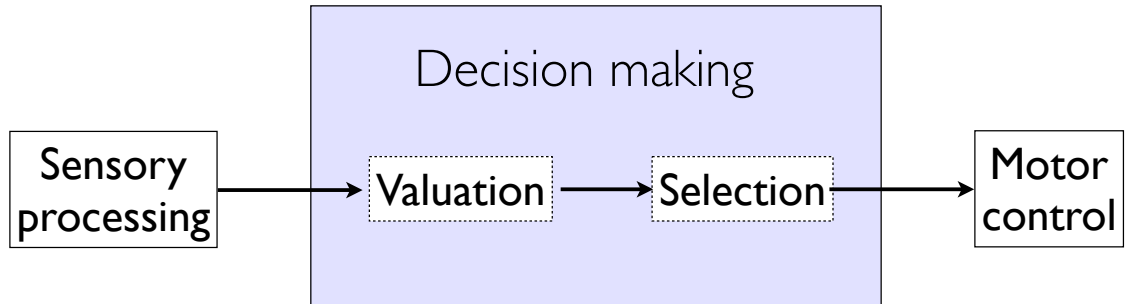


Figure 1.1 Model of decision making

Decision-making models describe the computational mechanism that links the sensory inputs with motor outputs. Commonly, the process of decision-making is divided into two stages, the first one being the valuation of the available alternative options and the second one the selection of the best action or stimulus.

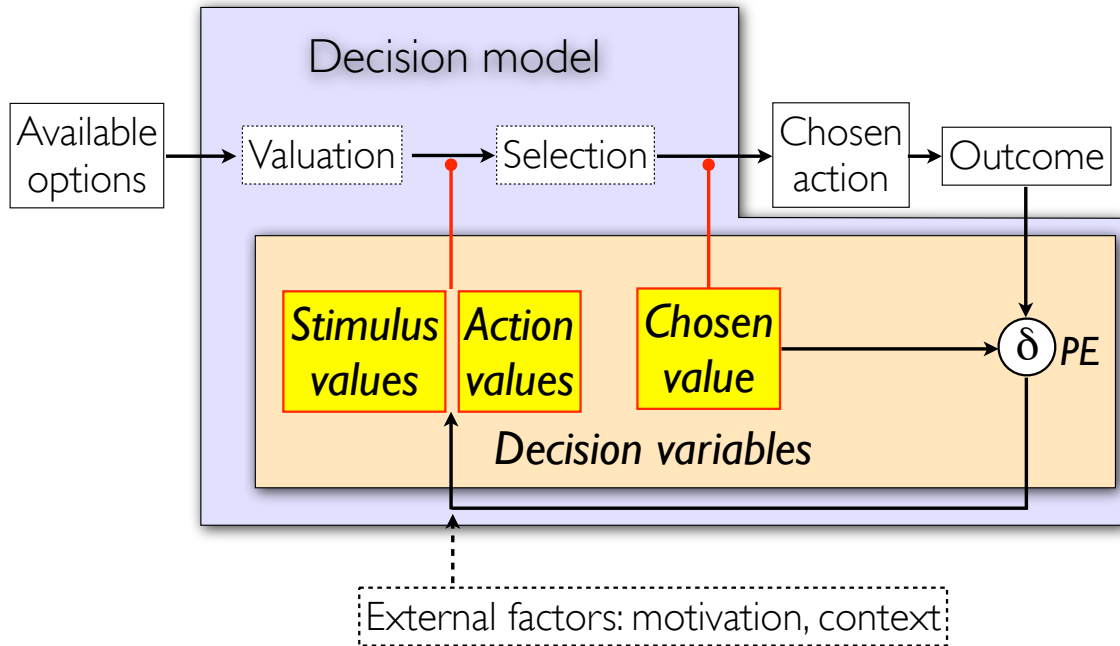


Figure 1.2 Decision variables

A simplified decision-making model based on reinforcement learning. Decision variables encode subjective states and values about individual preferences and are used in guiding the choice process. Values represent the desirability of the various available options. Together with motivational and contextual factors the stimulus or action values drive the decision towards the alternative with the highest subjective attractiveness. After the subject chooses an appropriate action, the outcome of the decision is compared to the expected outcome (chosen value) and any deviation is used to update future values through the predication error (PE). Since action and stimulus values are necessary to make a choice they are considered as inputs to the decision process. In contrast, value chosen signals are a consequence of the decision and thus regarded as output signal of the choice process.

## Chapter 2. Action-based decision making<sup>i</sup>

*Action-based decision making involves choices between different physical actions in order to obtain rewards. To make such decisions the brain needs to assign a value to each action and then compare them to make a choice. Using fMRI in human subjects we found evidence for action value signals in supplementary motor cortex. Separate brain regions, most prominently ventromedial prefrontal cortex, were involved in encoding the expected value of the action that was ultimately taken. These findings differentiate two main forms of value signals in the human brain: those relating to the value of each available action, likely reflecting signals that are a precursor of choice, and those corresponding to the expected value of the action that is subsequently chosen, and therefore reflecting the consequence of the decision process. Furthermore, we also found signals in the dorsomedial frontal cortex that resemble the output of a decision comparator, which implicates this region in the computation of the decision itself.*

---

<sup>i</sup> Adapted with permission from Klaus Wunderlich, Antonio Rangel, John P O’Doherty, “Neural computations underlying action based decision making in the human brain”, *Proceedings of the National Academy of Sciences, PNAS* **106**: 17199-17204 (2009).

## Introduction

Consider a goalkeeper trying to stop a soccer ball during a penalty kick. Within a brief amount of time he needs to choose between jumping to the left or right goal posts. Repeated play against the same opponents allows him to learn about their scoring tendencies, which can be used to compute the values of a left and a right jump prior to making a decision. It is a long established view in economics, psychology, and computational neuroscience that the brain makes choices among actions by first computing a value for each possible action, and then selecting one of them on the basis of those values [5, 8, 9]. This raises two fundamental questions in decision neuroscience: (1) Where in the brain are the values of different types of actions encoded? and (2) How and where does the brain compare those values to generate a choice?

An emerging theme in decision neuroscience is that organisms need to make a number of value related computations in order to make even simple choices [6]. Consider the case of action-based choice exemplified by the goalkeeper's problem. First, he needs to assign a value to each action under consideration. These signals, known as action values, encode the value of each action prior to choice and regardless of whether it is subsequently chosen or not, which allows them to serve as inputs into the decision making process [16-18]. Second, these action values are compared in order to generate a choice. Third, the value of the option that is selected, known as the chosen value, is tracked in order to be able to do reinforcement learning. In particular, by comparing the value of the outcome generated by the decision to the chosen value, the organism can compute a prediction error signal that can be used to update the action value of the chosen option. Note that

while the action values are computed before the decision is made, the chosen value and outcome of the comparator process signals are computed afterwards.

Although a rapidly growing number of studies have found neural responses that are correlated with some form of value signals, little is known about how the brain encodes action values or about how it compares them. This is central to understanding how the brain makes action-based choices. For example, a number of chosen value signals have been found in the orbital and medial prefrontal cortex [48, 49] and amygdala [46, 50]. Note that these signals are quite distinct from action values, and are not precursors to choice, because they reflect the value of the actions that were selected in the decision. For similar reasons, the value signals that have been found in lateral intraparietal cortex (LIP) during saccadic action-based choice [33, 34] are also not pure action values since they are strongly modulated by whether an action is subsequently taken. This suggests that instead of serving as inputs to the comparison process, they reflect its output. Several studies found orbitofrontal cortex to encode the value of different goals [10, 51, 52]. Though these signals are precursors of choice, they are not instances of action values since they are stimulus-based and independent of the action required to obtain them. To date only three monkey electrophysiology studies have found evidence for the presence of action-value signals for hand and eye movements in the striatum during simple decision making tasks [16-18]. This study extends their findings in three directions. First, as of yet no evidence has been presented for the existence of action value signals in the human brain. Second, by using fMRI we are able to look for action-value signals in the entire brain, whereas the previous electrophysiology studies have limited their attention to the striatum. As a result, no previous study has looked for action value signals in the cortex.

This is important because, as discussed below, there are a-priori reasons to believe that action value signals might be found in the motor and supplementary motor cortices. Finally, we investigate how such signals might get compared in order to actually compute the decision itself and where neuronal correlates of the output of this decision process are represented, an issue about which very little is known.

We studied these questions using fMRI in humans while subjects performed a variant of a two armed bandit task in order to obtain probabilistically delivered monetary rewards (Figure 2.1A). A critical feature of the task was that they had to select a motor response in one of two distinct response modalities: in every trial they could choose to make either a saccade to the right of a fixation cross, or to press a button with the right hand. This design allowed us to exploit the fact that different regions of the cortex are involved in the planning of eye and hand movements [53]. We hypothesized that value representations for the two actions would be separable within these cortical areas at the spatial resolution available to fMRI. The probability of being rewarded on each of the two actions drifted randomly over time and was independent of the probability of being rewarded on the other (Figure 2.1B). This characteristic ensured that value estimates for eye and hand movements were uncorrelated, which gave us maximum sensitivity with which to dissociate the neural representations of the two action values.

In order to look for neural correlates of action values we had to estimate the value of taking each action in every trial. We calculated the action values using a computational reinforcement-learning (RL) model in which the value of each action,  $V_{\text{eye}}$  and  $V_{\text{hand}}$ , was updated in proportion to a prediction error on each trial (see Table 2.1 for a summary of how the different types of value signals relate to the components of the experiment). The

model also assumed that action selection in every trial followed a soft-max probability rule based on the difference of the estimated action values [48]. To test for the presence of action value signals in the brain we took the model predicted trial-by-trial estimates of the two action values and entered these into a regression analysis against the fMRI data. In addition to a whole brain screening for the presence of action value signals, we specifically looked for them in areas known to be involved in the planning of motor actions, including supplementary motor cortex [54-57] and lateral parietal cortex [58, 59]. Given that both of these areas have previously been shown to contain value related signals for movements in nonhuman primates, and that they are closely interconnected with the area of motor cortex involved in carrying out motor actions [60-62], we considered these areas prime candidates for containing action-value representations that could then be used to guide action-based choices. It is important to emphasize, however, that the tasks used in previous studies did not make it possible to determine if the value signals identified were chosen values or action values.

We also looked for areas that are involved in comparing the action values to make a choice. Two areas of a-priori interest were the anterior cingulate cortex (ACC) and the dorsal striatum. ACC has been previously implicated in action-based choice, both in the context of a human imaging study reporting activity in this area during a task involving choices between different actions compared to a situation involving responses guided by instruction [63], and in a monkey lesion study where ACC lesions produced an impairment in action-outcome based choice but not in mediating changes in responses following errors [64]. Dorsal striatum has been implicated in both goal-directed and habitual instrumental responding for reward in rodents [65, 66]. Moreover, human fMRI

studies reveal increased activity in both of these regions when subjects make choices in order to obtain reward compared to an otherwise analogous situation in which the rewards are obtained without the need to make a choice [38, 67-69].

The most simple type of comparison process would be to compute a difference between the two action values. We tested for such a difference, but as we had no a priori hypothesis about the directionality of the computation, we tested for both the difference between the value of the action chosen and the value of action not chosen ( $V_{\text{chosen}} - V_{\text{unchosen}}$ ), and one involving the opposite difference ( $V_{\text{unchosen}} - V_{\text{chosen}}$ ). As we found evidence for such an action value comparison signal in the brain, we then proposed a simple computational model to provide a conceptual explanation as to how such a signal could reflect the output of a computationally plausible decision mechanism.

## Results

### *RL model fits to behavioral choice data*

A comparison of the choice probabilities predicted by the RL model and the soft-max procedure to subjects' actual behavior suggests that the model matches subjects behavior well. Figure 2.1C compares both variables for a typical subject. Figure 2.1D compares the predicted choice probability (binned) against the actual choice probabilities for the group. A similar linear regression analysis at the individual level generated an average  $R^2$  across subjects of 0.83 and regression coefficients that were significant at  $p < 0.001$  in each subject.



*Action values*

We found neural activity correlating with the action values for making a hand movement in left supplementary motor area (SMA; Figure 2.2A, Table S2). A region of interest (ROI) analysis showed that activity in this area satisfied the properties of a hand action value: it was sensitive to the value of hand movements, and it showed no response selectivity to the value of eye movements (Figure 2.2B). Activity in lateral parietal cortex, anterior cingulate cortex and right dorsal putamen also correlated with hand action values. In contrast, activity in a region of left supplementary motor cortex anterior to the SMA (pre-supplementary eye fields, preSEF, Figure 2.2A, Table S2) correlated with action values for eye movements. A similar ROI analysis showed that the area satisfied the properties of an eye action value: it was sensitive to the value of eye movements, but showed no sensitivity to the value of hand movements (Figure 2.2B). We tested this by performing a 2-way ANOVA with the factors of area (SEF vs. SMA) and action value (eye vs. hand). There was no significant main effect for either area or action value but the interaction was significant at  $p=0.03$  ( $F=5.6$ ,  $df=1$ ). Another important feature of an action value signal is that, since it is a precursor of choice, it should not depend on which action is actually chosen. We tested for this property by computing the following two voxel-wise interaction contrasts ( $V_{e|eye} - V_{e|hand} \neq 0$  and  $V_{h|hand} - V_{h|eye} \neq 0$ ). We found no significant interaction between action value and chosen action in either SMA or preSEF at  $p<0.05$  uncorrected. A post-hoc plot of the average percent signal change within each cluster plotted as a function of high and low action values are shown in Figure S 2.2.

One potential explanation for these correlations is that activity in the SEF and SMA reflect motor preparation. To help exclude the possibility we carried out two additional analyses. First, we estimated a model that used reaction times (RT) as a proxy index of the degree of motor preparation on a given trial and found hand and eye RTs did not show the same pattern of differential correlations in SMA and SEF as exhibited by our action-value regressors. Second, we estimated a version of our main general linear model in which the RTs were included as a covariate of no interest alongside our action-value signals, and found the action-value results in SMA and SEF still survived at  $p < 0.005$  uncorrected. Both results suggest that simple motor preparation is unlikely to account for the correlations with action values identified above.

Another alternative potential explanation for the correlations between activity in SMA/pre-SEF and action-values is that signal fluctuations in these areas depend on the degree to which subjects currently choose those motor actions. For example when the value of a hand movement is high, the subject may tend to choose hand actions more often, and therefore activity in SMA may be increased as a result of enhanced overall motor excitability. We tested for this possible confound by regressing BOLD signals against the degree to which subjects' favored one or another action in the recent past. We found activation most prominently within the occipital lobe, primary motor cortex, cerebellum, and dorsal medial frontal gyrus. However, we did not find any significant correlation within our previously identified action value areas, ruling out this possible explanation for the action-value signals in SMA, SEF and elsewhere (see Table 2.2 and Figure S 2.3).

### *Chosen values*

We then looked for correlates of the value of the action that is chosen on a particular trial, irrespective of its modality. Consistent with previous findings [48, 49], we found chosen value modulated activity in a number of brain areas, most prominently the ventromedial prefrontal cortex extending onto the medial orbital surface (Figure 2.3A, Table S2). The parietal cortex, including bilateral IPS and right LIP, were also activated by this contrast.

We also tested for areas correlating with the chosen value only on occasions when the action chosen was a hand movement, and for areas correlating with chosen value only on trials in which the eye movement was chosen. Intriguingly, we found evidence for a topographical arrangement of action specific chosen value signals in vmPFC along the anterior-posterior axis, whereby a mid-vmPFC region correlated with hand values only when hand movements were selected, and a region of more posterior vmPFC correlated with the value of eye movements only on trials when eye movements were selected. These two action specific representations were both located caudal to the value chosen signal reported above (Figure 2.3B).

### *Action value comparison: decision computation*

The most straightforward decision process to compare the action values is to compute the value difference and choose the one with the higher value. We looked for areas in which BOLD activity was correlated with the value difference between the two action values. As any difference in values can be computed by subtracting the lower from the higher value and also by subtracting the higher value from the lower value, and we had no a priori hypothesis for the directionality of the computation, we tested for correlates of both. We did not find any areas where activity was correlated with  $V_{\text{chosen}} - V_{\text{unchosen}}$  at

our omnibus statistical threshold of  $p < 0.001$  uncorrected. However, we found a strong correlation with  $V_{\text{unchosen}} - V_{\text{chosen}}$  in anterior cingulate cortex, extending dorsally into Brodmann area 9 (dmFC; Figure 2.4A, Table 2.2).

In order to provide a conceptual explanation as to how the brain might implement the value difference computation, we constructed a computational model called the Competition Difference Model (CDM). This model is a simple neural network that carries out value comparisons by stochastic mutual inhibition between two populations of neurons: one encoding the value of a hand movement, and one encoding the value of an eye movement. The model takes into account the stochasticity that leads to non-optimal choices in a proportion of the trials, consistent with actual behavior choices. It produces an output that closely resembles but is not identical to, the value comparison regressor used above (See Figure S 2.6 for details). In order to validate the model behaviorally, we compared the performance of the model on subjects' actual choice behavior and found that the model predicted subjects' actual choices as well as the soft-max procedure used with reinforcement learning (Table 2.3). We then used the output of this model as a parametric regressor in our fMRI analysis instead of the value difference. This model was found to correlate robustly with activity in the same anterior cingulate cortex region identified as correlating with the value difference (Figure 2.4B). Thus, the model proposed here provides a possible description of the output of a decision comparator, and captures activity related to such a comparison process in anterior cingulate cortex.

An important question is the relationship between our suggested role for dmFC/ACC in

the decision comparison process and prior findings implicating this area in error monitoring [70]. An error-monitoring signal would be strongest on trials where subjects chose the lower valued action and in which there is a large difference between the values of the two available actions (as on those trials it should be most clear to the subjects that they have erroneously chosen the “wrong” action). However, we still find significant correlations between the dmFC signal and our decision model when we restrict the analysis to trials in which the value of the chosen action largely exceeded the value of the unchosen action (Figure 2.4C). Another possibility is that subjects were deliberately choosing the lower value action in some trials to explore and anticipated in those deliberate “error” trials a negative outcome. We therefore also looked for regions that had stronger activity during the choice period on trials that were subsequently not rewarded compared to subsequently rewarded trials. Activity in the frontal poles showed such a pattern but not activity in dmFC/rostral ACC. Together this suggests that the decision signal is unlikely to be accounted solely as a side effect of error monitoring.

Another pertinent issue is the extent to which the activity in dmFC/ACC is related to conflict monitoring, another cognitive function that has been attributed to this area [71]. In order to compare these explanations we constructed a measure of decision conflict by testing for areas showing a maximal response when the action values are the same, and a minimal response when they are maximally different. We found that activity in rostral dmFC is significantly better explained by the decision signal than by this simple decision conflict signal (Figure 2.4D), and this is true even if the measure of decision conflict includes subject specific biases to either eye or hand movements. To further address this

point, we also tested for correlations of reaction time with decision difficulty, but did not observe any such influences ( $r = 0.01$  across all subjects).

## Discussion

Action based decision-making involves different kinds of value signals that play specific roles in the various stages of the decision process (Figure 2.4E). Action values are by definition precursors of choice that are used to guide the decision process. Here we provide evidence that action-values for different physical actions are present in the supplementary motor area. These value signals are not modulated by choice, i.e., they are present for a given action on trials when that action is chosen and on trials when that action is not chosen. We found neural correlates for action-values in supplementary motor cortex, an area traditionally associated with motor planning [72]. This finding supports the hypothesis that during decisions involving motor actions, action-value signals are encoded in brain regions directly connected with, and involved in, the generation of motor output [73]. This finding is broadly consistent with a number of previous studies that have investigated the role of the motor system in decision making. For example, two studies [61, 62] have found monkey medial premotor cortex involved in the entire discrimination process between haptic stimuli, and the findings of another study [60] suggest that formation of the decision and formation of the behavioral response share a common level of neural organization.

In contrast to the supplementary motor cortex, activity in both the ventromedial prefrontal cortex and intraparietal sulcus pertained predominantly to the value of the action that is chosen. Such signals are a consequence of the decision process, emerging

after the subject has decided which action will ultimately be taken. We suggest that the functional contribution of such signals to the decision process is likely not in guiding choice directly, but rather in learning the action values. Reinforcement learning theory stipulates that updating of action-values occurs via a prediction error, which computes the difference between actual and expected outcomes [5, 39, 74]. A major function of the chosen value signals in these areas could be to facilitate the generation of a prediction error signal that can then be used to update future action values. It is notable that the two signals required to compute a prediction error, namely the actual outcome and the expected outcome (of the chosen action) are both represented in ventromedial prefrontal cortex [48, 49, 75, 76]. Therefore this region is ideally placed to facilitate computation of the prediction error signal that could then be transferred on to dopaminergic neurons in the midbrain for subsequent broadcast [77, 78]. Another intriguing feature of our results is that we observed a number of different types of chosen value signals within vmPFC. While one region of vmPFC was responsive to the value of the chosen action irrespective of whether that action was a hand or an eye, distinct regions more posterior within vmPFC appear to be sensitive to action specific chosen values. These findings provide evidence that values of different types of movement might be represented separately within ventromedial prefrontal cortex, adding further support to the suggestion that this region plays a role in encoding the value of chosen actions, as well as possibly contributing to encoding stimulus-values [79]. The apparent topographical arrangement of action modality specific value signals within vmPFC may relate to distinct cortico-striatal loops concerned with processing hand and eye movements [80].

Our results also suggest that the dmFC/ACC plays a role in the decision process. Interestingly, this area has been previously implicated in action-based choice in the context of a human neuroimaging study reporting activity in this area during a task involving choices between different actions compared to a situation involving responses guided by instruction [63], and single neuron recordings have shown that cells in this area were activated only by particular action-reward combinations [81]. Another study suggests that this region plays a part in processing the reward information for motor selection [82]. Consistent with our findings, Seo and Lee [58] found neural signals resembling the difference between action values in this region. In addition, ACC lesions have been shown to produce an impairment in action-outcome based choice, but not in mediating changes in responses following errors [64, 83]. Our results provide evidence that these deficits might be the results of impairments in the mechanisms in ACC/dmFC responsible for comparing action values. Heekeren et al. [84, 85] (see also [86]) have used fMRI to look for regions that might be involved in computing perceptual decisions. They found evidence that activity in left dorso-lateral PFC encodes a value signal that is proportional to the absolute value difference between the two signals, while our value difference related signal is represented in dorso-medial PFC. The specific form of the comparison signal we found in ACC was well captured by a simple network model which we called the Competition Difference Model. This model relies on a mutual inhibitory competition between distinct populations of neurons representing eye and hand movements in order to generate a decision. Although it bears a conceptual relationship to many other models used to generate decisions such as the drift diffusion model (DDM) [87-89], the predictions of the CDM and the DDM model are in fact very different (see



supplementary materials for more details). Indeed, while the CDM model provides a good account for the comparison signal we observed in ACC, the DDM model fails to capture such an output. Because our study was not designed to address the presence of DDM-related signals we cannot rule out the contribution of such computations to the decision process. However, it is worth noting that while there is now considerable evidence concerning the applicability of the DDM model to the neural mechanisms underlying decision making in the perceptual domain [90, 91], to our knowledge very little evidence exists regarding the applicability of such a model to value-based decision making. Thus, it is possible that these two different types of decisions rely on distinct computational processes.

Interestingly, the signal reflecting the output of the action value comparator represented the difference between the action not chosen and the action chosen, instead of the more intuitive difference given by the action chosen minus the action not chosen. A speculative interpretation for this finding is that the outcome of the comparator process is used to inhibit the opposite action, instead of exciting the motor plan that it represents. Interestingly, our activation pattern looks very similar to one found in a study of volitional motor inhibition [92]. Such a mechanism based on pre-innervation and inhibition could provide a better execution speed after the values become available compared to a mechanism where the motor response is planned only after the decision has been made. Though we cannot distinguish between excitatory and inhibitory processes based on the measured BOLD [93], our hypothesis resonates with previous findings that pre-innervation and inhibition play an important role in motor execution and volitional action initiation [94, 95].

Since activity in ACC/dmFC has been associated with error monitoring and conflict detection in previous studies, we carried out several controls to help exclude the possibility that the activity we observed in this area can be explained by these alternative computations. We emphasize that our results don't rule out a contribution of ACC to either conflict or error monitoring, but rather suggest that these explanations are unlikely to account fully for the results we observe here. Instead we provide a mechanistic account for how action comparison signals in ACC/dmFC could form an integrated part of the decision process. Note that because of limitations in the spatial and temporal resolution of our fMRI signal it is not possible to determine whether the signal we observe reflects solely the output of a decision comparator or whether the dmFC/ACC is involved in the comparison process itself. Therefore the possibility exists that the actual computation of the decision is carried out elsewhere and the output then transferred to dmFC/ACC.

Choices between different physical actions, such as those studied here, represent a large subset of the decisions made by humans and other animals. The present study has identified neural mechanisms involved in these types of choices and provides insight into the general neural mechanism that might be involved in action-based decision-making. An important question for future studies is whether similar mechanisms are at play when goal directed decisions are made between more abstract choices not tied to specific physical actions.

## Methods

### *Subjects*

23 healthy subjects (10 female; 18–29 years old; right-handed, assessed by self-report with an adapted version of the Edinburgh handedness inventory [96]) with no history of neurological or psychiatric illness participated in the study. The study was approved by the Institutional Review Board of the California Institute of Technology.

### *Experimental design and task*

The task is a variant of a two-armed bandit problem in which subjects chose between two actions: a button press with the right index finger, and a saccade from a central fixation cross to a target located at 10 degrees of visual angle in the right hemifield. In every trial each action yielded either a prize of ten cents, or nothing. We did not reveal the exact reward per trial to subjects before the experiment but instead instructed them only that they will get a small amount of money for each rewarded trial. At the end of the experiment subjects were paid their accumulated earnings in addition to a flat amount of 25\$.

The probability ( $Q_{i,t}$ ) of action  $i$  being rewarded in trial  $t$  evolved over time as a decaying Gaussian random walk process, with  $Q_{i,t+1} = \max(0, \min(1, \lambda Q_{i,t} + (1 - \lambda)\theta + v))$ ; where the decay parameter  $\lambda$  was 0.79836, the decay center  $\theta$  was .50, and the diffusion noise  $v$  was zero-mean Gaussian with standard deviation  $\sigma_d = .208$ . Five different probability trajectories were generated using this method and were assigned across subjects randomly. The task consisted of two sessions of 150 trials separated by a short break.

There were three trial types. In free-choice trials (150 trials) the subject had to choose one of the two actions and both were rewarded according to their current reward schedule. Free-choice trials were pseudo-randomly interspersed with forced-choice trials (50 eye trials and 50 hand trials) and null-choice trials (50 trials). Subjects were instructed that in forced trials only the displayed action would be rewarded with its current probability, while the other action would lead to a zero prize with certainty. Subjects did not get a prize in null-trials, but were still required to make a choice.

The task was presented via back projection on a translucent screen, viewable through a headcoil mounted mirror. Subjects chose the hand action by pressing a button on a button box with their right index finger. Eye positions were monitored at 120 Hz with a long-range infrared eye-tracking device (ASL Model L6 with control unit ASL 6000, Applied Science Laboratories, Bedford, MA). An eye action during the choice period was registered when the median horizontal eye coordinate during the past 200 msec exceeded 8 degrees of visual angle to the right from fixation. Subjects were instructed to maintain fixation during the entire experiment when not deliberately making a saccade.

#### *Reinforcement Learning (RL) model*

A RL model was used to estimate the value that the brain assigned to the two actions on the basis of trial-by-trial experience. In this study we used a version of RL called Q-learning, where action values are updated using a simple Rescorla-Wagner rule (see supplementary methods for details).

#### *Computational model of the choice process (decision model)*

We were also interested in identifying brain regions involved in comparing the action

values in order to make decisions. The most basic value comparison process that one could consider involves calculating the difference between the action values in order to identify and select the largest one. A problem with such a model is that it does not account for the choice stochasticity that is observed in the data, and thus it cannot explain behavior in those trials where subjects chose the action with the lower action value. We therefore constructed an extremely simple neural network type model that characterizes the properties of aggregate activity that identify putative decision making regions. We then use these trial-by-trial predictions as parametric regressors in our fMRI analysis to identify areas where the value comparison computation might be carried out (see supplementary methods for details).

#### *fMRI data acquisition and analysis*

Data were acquired with a 3T scanner (Trio, Siemens, Erlangen, Germany) using an eight-channel phased array head coil (see supplementary methods for details).

We estimated two different general linear models with AR(1) for each individual subject (see the supplementary methods for details). In each case we computed contrasts of interest at the individual level using linear combinations of the regressors and, to enable inference at the group level, we calculated second-level group contrasts using a one-sample t-test.

Whole brain inference was carried out at  $p < 0.001$  uncorrected. We also computed small volume correction (SVC) for multiple comparisons at the  $p < 0.05$  level in areas of a-priori interest (supplementary methods).

The structural T1 images were co-registered to the mean functional EPI images for each subject and normalized using the parameters derived from the EPI images. Anatomical localization was carried out by overlaying the t-maps on a normalized structural image averaged across subjects, and with reference to an anatomical atlas [97].

In order to insure the independence of the effect size analysis in Figure 2.2 and Figure 2.3 we randomly divided the data into two halves: the first half was used to define an ROI, the second half was used to measure the effect sizes (see the supplementary materials for details).

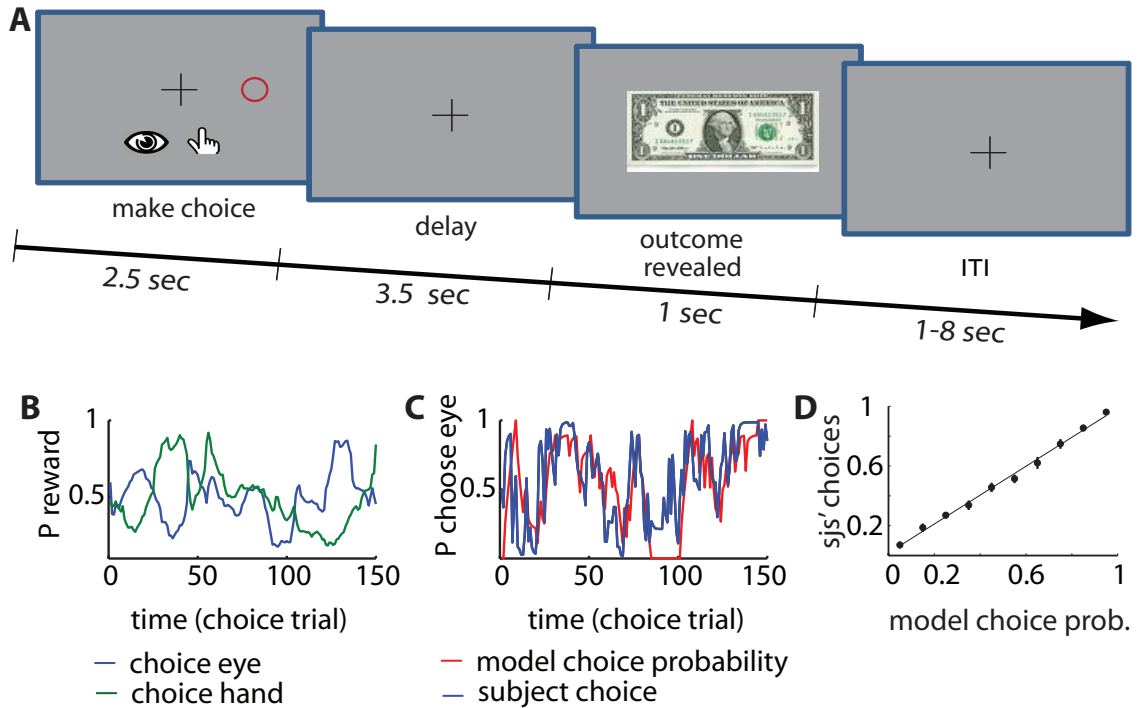


Figure 2.1 Experimental Design and Behavior

(A) Subjects were presented with a choice cue after which they had to respond within 2.5s by performing a saccade to the red target circle or a right handed button press. Once a response was registered the screen was immediately cleared for a short delay and subsequently the outcome was revealed (6 s after trial onset) indicating either receipt of reward or no reward. Inter-trial-intervals varied between 1 and 8 seconds. (B) Example reward probabilities for saccades and button presses as a function of the trial number. The probability of being rewarded following choice of either the hand or eye movement was varied across the experiment independently for each movement. (C) Fitted model choice probability (red) and actual choice behavior (blue) shown for a single subject. (D) Actual choice behavior versus model predicted choice probability. Data is pooled across subjects, the regression slope is shown as a line, vertical bars represent s.e.m.

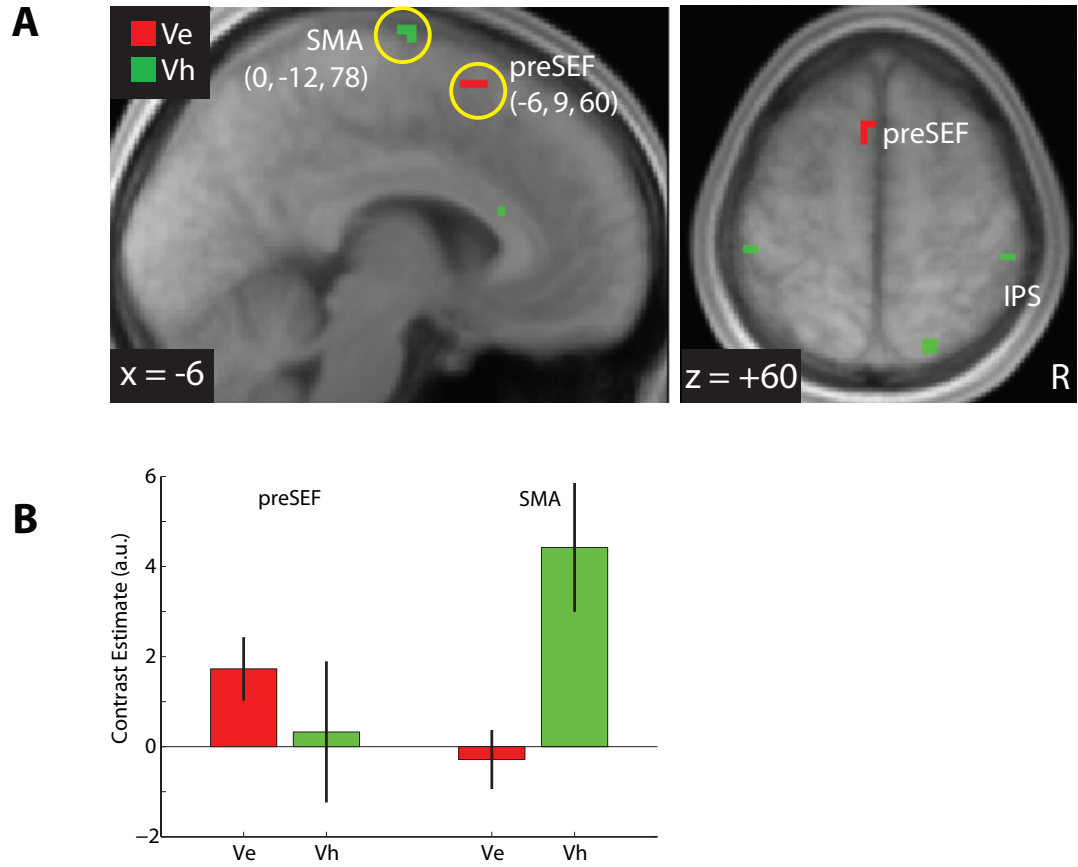


Figure 2.2 Action values

(A) Region of supplementary motor area showing correlations with action-values for hand movement (Vh/green) and a region of pre-SEF showing correlations with action-values for eye movements (Ve/red). T-maps are shown from a whole brain analysis thresholded at  $p < 0.001$  uncorrected (see Figure S 2.1 for a version with colorbars relating to t-stats). (B) Average effect sizes of Ve (red) and Vh (green) extracted from SEF and SMA. The effects shown here were calculated from trials independent of those used to functionally identify the ROI. Note that only Ve but not Vh modulate the signal in preSEF, and that activity in SMA shows the opposite pattern. Vertical lines depict s.e.m.



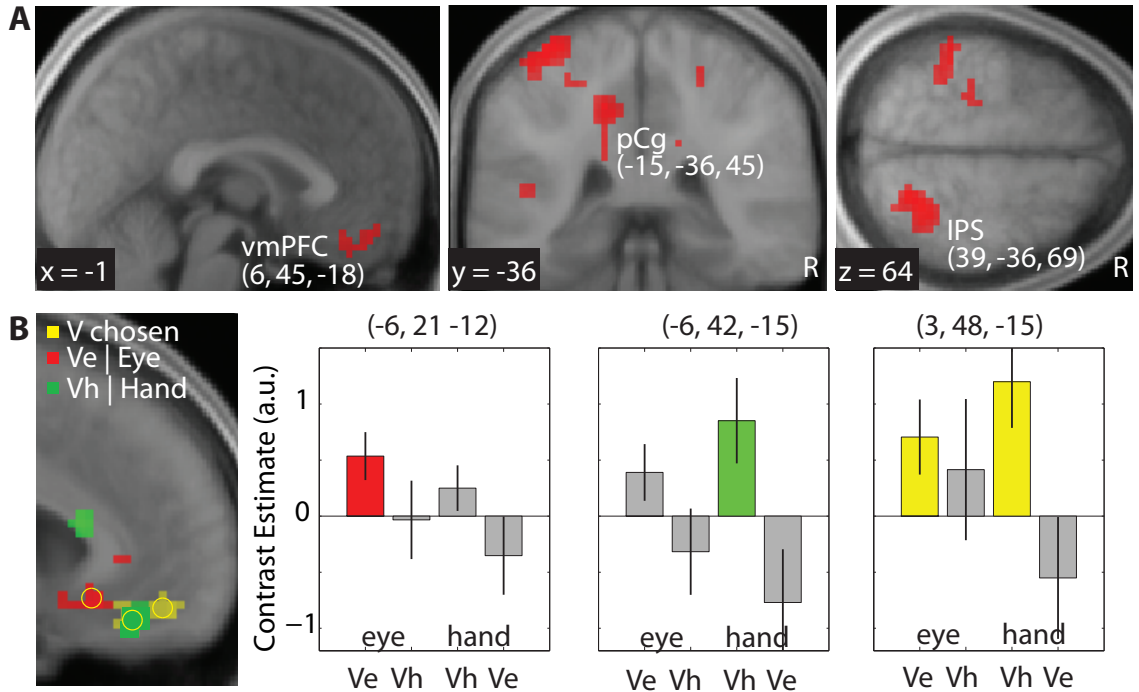


Figure 2.3 Chosen values

(A) Brain regions showing significant correlations with the value of the action chosen. Areas shown include vmPFC, intra-parietal sulcus and posterior cingulate cortex. Threshold is set at  $p < 0.001$ . (B) Distinct forms of the value chosen signal are present within vmPFC. The area depicted in yellow indicates voxels that correlate with the value of the chosen action irrespective of whether the action taken is a hand or an eye movement. The area depicted in green correlates only with the value chosen on trials when the hand movement is chosen but not when the eye movement is chosen. Finally the area depicted in red indicates voxels correlating with value chosen only on trials when the eye movement is selected but not the hand movement. The results suggest an anterior to posterior trend in the selectivity of voxels to these different types of value chosen signals. Barplots show effect sizes averaged across subjects for the action specific value chosen signals in the three areas (left: red area, middle: green area, right: yellow area). Bars shown in chromatic color are significantly different from zero ( $t$ -test,  $p < 0.05$ ). Similar to barplots in Figure 2.2B, effects were calculated from a data sample independent of the one used to functionally identify the ROI. Vertical lines denote s.e.m.

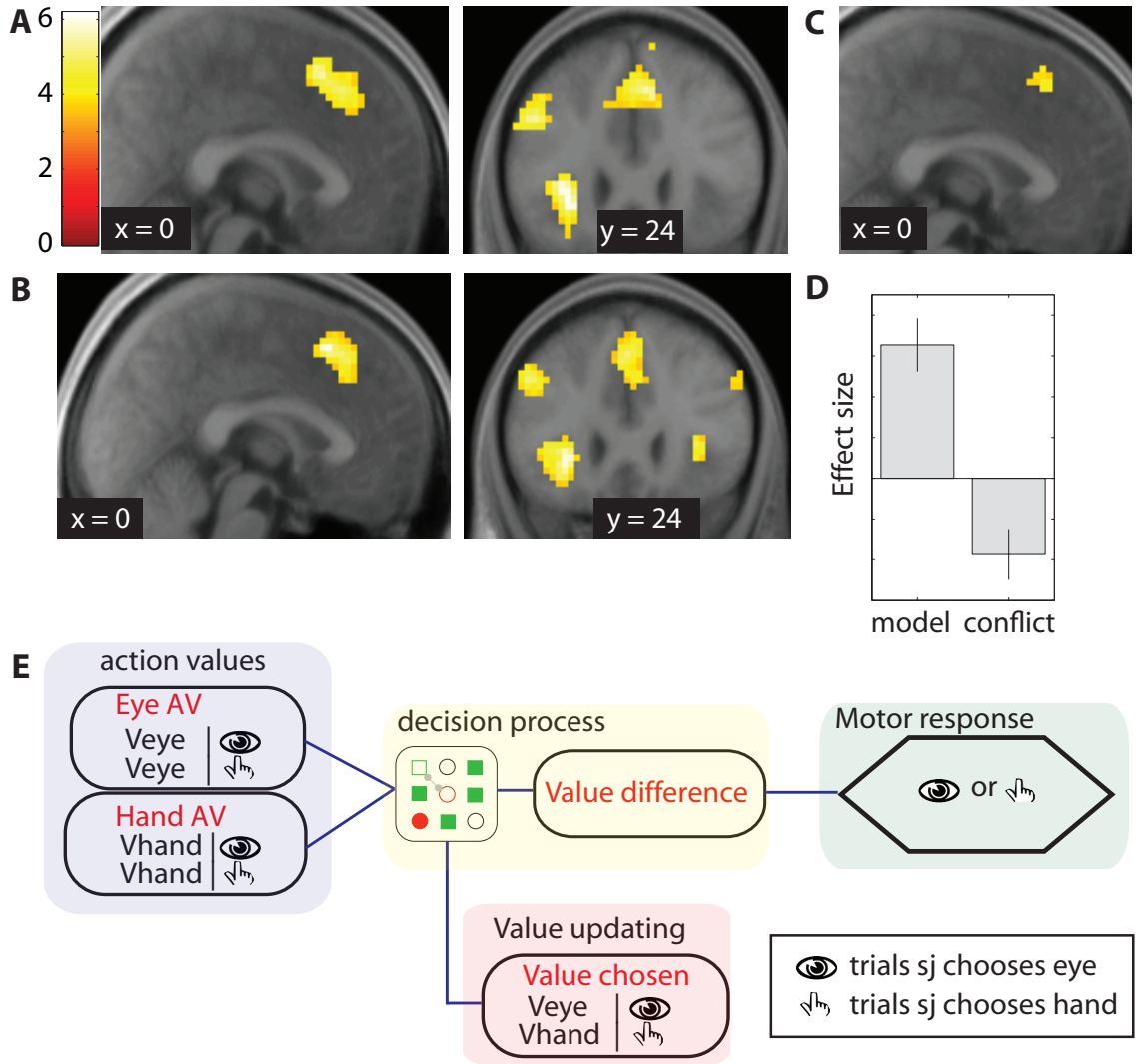


Figure 2.4 Value comparison

(A) Region of dmFC and adjacent ACC showing significant correlations with the  $V_{\text{unchosen}} - V_{\text{chosen}}$  value difference contrast. Additional areas correlating with this comparison signal are bilateral anterior insula and left dlFC. (B) Output of our stochastic decision model for the value comparison showing correlations with activity in the same brain regions. (C) The model explains activity in dmFC even on a subset of trials where subjects clearly choose the “correct” and not “erroneous” choice (where  $V_{\text{chosen}} - V_{\text{unchosen}} > 0.2$ ). This suggests that the result in (B) cannot be fully explained by error monitoring. (D) Average beta values in the random effects analysis of the model described in the text showing that neural activity in dmFC/ACC is explained better by the

output of our decision model than by a decision difficulty based index of decision conflict ( $p < 10^{-7}$ ). The vertical lines represent the s.e.m. (E) Illustration of the different stages involved in action based decision-making: action-based decision-making requires the computation of distinct value representations for both choice alternatives (purple box). These action values are compared against each other in a decision comparator (yellow box) in order to decide on a particular action. Such a comparator could yield a signal that approximately resembles the difference in the action values of the two actions. The output from this comparator could then be passed through a nonlinear function to inhibit a response of the unchosen action in primary motor areas (green box). The value of the chosen action is used to update future action values on the basis of experience, and to generate prediction errors (red box).

## Supplementary Methods

### *Reinforcement Learning Model*

Reinforcement Learning (RL) is concerned with learning the value of taking particular actions in different states of the world in which subjects do not have complete knowledge about the underlying reward generating process. Thus, it is ideally suited to model how subjects learn the value of taking different actions over time.

We used a version of RL called Q-learning, where action values are updated using a simple Rescorla-Wagner rule. If an action is not selected in a trial its value is not updated. In contrast, if action  $a$  is selected on trial  $t$ , its value is updated via a prediction error,  $d$ , as follows:  $V_a(t+1) = V_a(t) + \eta\delta(t)$ , where  $\eta$  is a learning rate between 0 and 1. The prediction error  $d(t)$  is calculated by comparing the actual reward received,  $r(t)$ , with the reward that the subject expected to receive from that action in that trial; that is,  $\delta(t) = r(t) - V_a(t)$ . Probabilistic rewards were delivered in free choice and forced choice trials and these trials were included in updating the value predictions. In null trials, subjects neither expected nor got any reward; hence no learning occurred and values were not updated.

To generate choices, we first used a soft-max procedure wherein every trial, the probability (P) of choosing action  $a$  is given by:  $P_{a,t} = \sigma(\beta(V_a(t) - V_b(t)) - \alpha)$ , where  $\sigma(z) = 1/(1 + e^{-z})$  is the Luce choice rule or logistic sigmoid,  $\alpha=0$  denotes the indecision point (at which both actions are selected with equal probability), and  $\beta$  determines the

degree of stochasticity involved in making decisions. In the paper we refer to  $V_e$  as the action value of the eye movement and  $V_h$  as the action value of the button press.

The model decision probabilities  $P_e$  and  $P_h$  were fitted against the discrete behavioral data  $B_e$  and  $B_h$  in order to estimate the free parameters ( $\eta$  and  $\beta$ ). This was done using maximum likelihood estimation and a log likelihood function given by:

$$\log L = \frac{\sum B_e \log P_e}{N_e} + \frac{\sum B_h \log P_h}{N_h},$$

where  $N_e$  and  $N_h$  denote, respectively, the number of trials in which *eye* and *hand* were chosen, and  $B_e$  ( $B_h$ ) equals one if *eye* (*hand*) was chosen in that trial, and zero otherwise.

We also fitted a model with an additional parameter that allowed the unchosen value to decay towards 0.5. However, we found that in our rather simple task with only two choice options the BIC corrected fit of this model was not significantly better than that of our simple learning rule.

### *The Competition Difference Model of the decision process*

Based on our finding of a robust correlation between activity in the anterior cingulate cortex and a variable equal to the difference between the value of the unchosen and chosen actions, we propose a simple conceptual model for how this value difference might be implemented in the brain in order to guide value-based choice.

Importantly, the model that we propose has two key properties: (1) it leads to stochastic choices, and (2) its output is sensitive to both the choice that is made and to the action values of the two alternatives.

The model consists of a neural network with  $N$  ‘neurons’. Each neuron could take on either an ON or OFF state at every particular instant. These neurons were split into two discrete populations of  $N/2$  neurons each: one population was associated with the value of an eye movement, the other with the value of a finger movement. In each trial the comparison process is initialized by turning ON a fraction of neurons in each population that is proportional to the action value of the associated action. Thus, for example, if  $V_e=0.56$  and  $N=200$ , then 56 out of the 100 eye neurons were set to the ON state (see Figure S 2.4 for an illustration).

The network was then allowed to evolve in discrete steps as follows:

1. In every step every active neuron for one of the two actions is paired with a randomly chosen neuron (with replacement) for the other action. Once the assignment is made for all neurons the following rule is implemented: if the matching unit is ON, the neuron is switched OFF, otherwise no change is made on the state of the neuron. (Note: this rule is implemented simultaneously for all of the neurons, so there are no order effects).
2. Noise is injected after every iteration as follows: the state of every unit in the network is flipped to its opposite state with a probability that is given by the product of a noise parameter  $\sigma$  and the number of active units encoding the value of the same action.

The basic idea behind the CDM is that the decision process works by virtue of a stochastic mutual inhibitory competition between the two distinct populations of neurons encoding the value of the two actions. A “winner” is declared when one of the two

populations reaches zero. At this point the population that has a positive number of ON neurons is declared the winner. Note that the model incorporates two desirable features: (1) higher value actions have a higher chance to win the competition process, which means that the better action is chosen with higher probability; and (2) the change in activity in every step scales with the amount of existing activity in the network.

We then added an additional layer (with constant positive input from which the previous result is subtracted) to the model to invert the output to the value difference between the action not chosen and the action chosen.

We simulated the model using a population of  $N=200$  as follows. First, we simulated the stochastic comparison process 1000 times for each possible value difference between the two actions. Second, after the model converged in each simulation (which always occurred in less than 50 steps) we computed the number of ON units in the population that won the competition. Note that, since the model is stochastic, in some simulations it converged to the action with the larger value, but in others it converged towards the action with the smaller value. Third, we averaged the 1000 simulations for each possible action value difference in order to estimate a reference output value for later use in the comparison regressor in the general linear models of the fMRI data described below. As depicted in Figure S 2.6, the averaging was done conditional on whether the optimal choice was made or not. We constructed a trial-by-trial parametric modulator by retrieving the stored values from this analysis in each trial for the current value difference and dependant on whether the subject chose optimally (action with the higher action

value) or the action with the lower action value from either the red or blue curve in Figure S 2.6.

To validate our model behaviorally, we determined for each possible value difference ( $V_e - V_h$ ) in 1000 model runs the fraction of runs in which the model settled on the eye choice. For this purpose, a noise parameter  $\sigma$  was estimated for each subject using the maximum likelihood procedure described in the RL section above. The resulting psychometric choice functions (probability of the model to choose eye dependant on  $V_e - V_h$ ) are compatible with subjects' observed behavior and model performance is very similar to the reinforcement learning soft-max procedure (Table S3).

Note a few things about the model. First, it leads to stochastic choices, consistent with the behavior in Figure 2.1D. Second, unlike other models such as the drift diffusion model (see the discussion in the next section), the output signal depends on the action value difference when the best item is chosen (and is constant otherwise).

### *The Drift Diffusion Model of the decision process*

A very popular model of how the comparison is made is called the Drift Diffusion Model (DDM, sometimes also called race-to-barrier model) [87-89]. This model has proven extremely useful in explaining the psychophysics of perceptual choice as well as some aspects of neural activity in areas such as LIP during perceptual decision tasks [90, 91].

The basic idea of the model, as applied to value-based decision-making, is illustrated in Figure S 2.5. The process computes a net value signal (say  $V_h - V_e$ ) that fluctuates between two barriers until a decision is made. A decision is reached when the net value signal crosses either of the two barriers. If the top barrier is crossed the hand action is



chosen. If the bottom barrier is crossed the eye action is chosen. The net value signal climbs to the hand barrier with a slope proportional to  $V_h - V_e$ , but it is also affected by white Gaussian noise. In the simple version of the model the net value signal commences the integration process mid-way between the two barriers, which implies that there is no bias between the two options (i.e., when  $V_h = V_e$  both options are chosen with equal probability).

Note that this is a “high-level” computational model, which is silent about how the brain might implement these computations. This question needs to be answered to be able to make predictions about how to identify areas that might implement this process using fMRI. Consider an extremely simple neural implementation of the DDM. There are two populations of neurons: one encodes for the net value of a hand movement ( $V_h - V_e$ ), the other encodes for the net value of an eye movement ( $V_e - V_h$ ). Both populations encode a signal with a dynamic range 0 to  $M$ . Both signals begin the competition process at  $M/2$  and the decision process stops when one of the signals reaches  $M$ . The signal in the two populations evolves in discrete time until a choice is made. Each of the populations is connected to an output signal that encodes the selected motor movement, triggered once the integration threshold  $M$  is reached. Note a few interesting properties of the neural implementation of the DDM. First, the sum of activity in all neurons at every instant during the comparison equals  $M$ . Second, the sum of activity in both output signals is also equal to a constant, call it  $B$ , independent of  $V_e$  and  $V_h$ .

These properties imply that an area implementing the comparison should have a level of neural activity equal to  $M$  (independent of  $V_e$  and  $V_h$ ) from the onset of the trial until a choice is made. They also imply that the output of the process is characterized by a

constant level of activity  $B$  (again, independent of  $V_e$  and  $V_h$ ) that is on from the moment the decision is made to the time the motor output is executed.

These properties mean that the comparator activity of the DDM should be modeled in the general linear models of BOLD activity described below as an unmodulated regressor that begins with the onset of a free trial and ends with the deployment of one of the two actions (i.e., it has a duration equal to the reaction time). In contrast, the output activity should be modeled as an unmodulated regressor at the time of (either) action execution with a duration of 0 seconds.

Although these regressors provide a full characterization of the neural activity associated with the DDM, and they are easily incorporated in the general linear models described below, they present a major problem for fMRI. Consider, for example, the regressor for the comparator process. The activity for this process is perfectly correlated with those of other processes that come on-line during the evaluation process, that are also unmodulated by value, and that also last until a choice is made. Given that a large number of such processes are likely to exist (and in fact a large number of distinct areas are robustly activated by this type of contrast in decision-making tasks) it is difficult to isolate the location of the DDM comparator process using fMRI, particularly for the range of reaction times taken for decisions in a standard fast-paced decision task such as the one featured here. A similar problem holds for the output signal of the DDM, since it is perfectly correlated with motor activity that is not modulated by action values.

Given these issues, we concluded that the neural signatures of the DDM cannot be identified using the fMRI methods deployed in the present study. It is important to

emphasize that these measurement problems are not present in single-unit electrophysiology since this technique permits the independent measurement of neural activity in a putative decision region with sufficiently high spatial and temporal resolution. Moreover, the output of the model does not resemble the value difference signal we observed in anterior cingulate cortex in the present study. Thus, while we cannot assess the relevance of the DDM model to value-based decision making in the present study, it is the case that such a model does not provide a good account for the value comparison signal we observed in anterior cingulate cortex.

#### *FMRI data acquisition*

Functional images were taken with a gradient echo T2\*-weighted echo-planar sequence (TR = 2.65 s, flip angle = 90°, TE = 30 ms, 64 × 64 matrix). Whole brain coverage was achieved by taking 45 slices (3 mm thickness, no gap, in-plane resolution 3 × 3 mm), tilted in an oblique orientation at 30deg to the AC-PC line to minimize signal dropout in OFC. Subjects' head was restrained with foam pads to limit head movement during acquisition. Functional imaging data were acquired in two separate 568-volume runs, each lasting about 24 min. A high-resolution T1-weighted anatomical scan of the whole brain (MPRAGE sequence, 1×1×1 mm resolution) was also acquired for each subject.

#### *FMRI data analysis*

Image analysis was performed using SPM5 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, U.K.). Images were first slice time corrected to TR/2, realigned to the first volume to correct for subject motion, spatially

normalized to a standard T2\* template with a voxel size of 3 mm, and spatially smoothed with a Gaussian kernel of 8 mm FWHM. Intensity normalization and high pass temporal filtering (using a filter width of 128 s) were also applied to the data.

We estimated several general linear models (GLM) for each individual.

*GLM 1.* Two events were modeled in each trial: the time of the choice cue, parametrically modulated by the trial-by-trial action values  $V_e$  and  $V_h$ , and the time of the presentation of the outcome, modulated by the prediction error  $d$ . Trials in which subjects chose the eye action and trials in which subjects chose the hand action were modeled as separate regressors. Trials were further split to build separate regressors for each trial type: free choice, forced choice and null trials. Choice and forced trials were modulated by the estimated action values to find neural representations of those signals. In null trials there were no modulators. The model also included an orthogonalized version of the parametric action value modulators described in the previous paragraph during the inter-trial interval. The rationale behind this last set of regressors was to allow for the possibility that participants might already be considering which option to choose next after receiving the feedback on the previous trial. In such a case, the ITI window would be part of the decision process. However, we did not find any correlates of value related signals during the ITI, which led us to focus our analysis on the time of the choice cue. All regressors were convolved with the canonical hemodynamic response function. In addition, the 6 scan-to-scan motion parameters produced during realignment and two session constants were included as additional regressors of no interest. We then computed

contrasts of interest at the individual level using linear combinations of the regressors:  
 Value chosen:  $V_e|_{eye\_chosen} + V_h|_{hand\_chosen}$ ; Value difference ( $V_{unchosen} - V_{chosen}$ ):  
 $V_e|_{hand\_chosen} + V_h|_{eye\_chosen} - V_e|_{eye\_chosen} - V_h|_{hand\_chosen}$ . This model was  
 used to generate the statistics reported in Figures 2.2, 2.3 and 2.4A.

*GLM 2.* This model was identical to the first GLM except for the addition of following  
 two regressors:

1. A Dirac delta function 700 ms into every trial (which is equal to the average response  
 across subjects) modulated by the estimated output signal of the DCM model given the  
 values of  $V_h$  and  $V_e$  and the optimality of the choice made. The values of these  
 modulators are depicted in Figure S 2.6, dependant on the action value difference and  
 whether the subject chose optimally (red curve) or non optimally (blue curve).

2. An indicator for the time of cue presentation modulated by a decision difficulty  
 measure given by  $-|V_e - V_h|$ . Note that this modulator takes a maximum value when  
 $V_e = V_h$ . We also tested alternatively for subject specific conflict by taking into account  
 individual choice biases in calculating the value difference  $-|V_e - V_h - \alpha|$ . For example,  
 if a participant had a slight overall bias towards saccading, the difficulty would be  
 centered on this subject's individual point of equilibrium. For this purpose we estimated a  
 subject specific indecision point  $\alpha$  by fitting the RL model with a third free parameter  
 that allowed for horizontal shifts of the sigmoidal choice function.

As before, note that the value of the output signal of the CDM model used in every trial  
 computed is obtained by averaging over 1000 simulations, and that, due to the  
 stochasticity of the CDM, it is a noisy measure of the actual activity during the trial.

The goal of this second GLM was to look for regions in which activity correlated with the output signal of the DCM. The results of this second GLM were used to generate the statistics reported in Figure 2.4B and 2.4C.

*GLM 3 & 4.* We carried out two further analyses to rule out the possibility that action-value signals observed in SMA and pre-SEF could be attributed to motor preparation. In the first such additional analysis we estimated a GLM in which trials involving hand or eye movements were entered as separate indicator variables, and reaction times (RTs) for those hand and eye movements were used as parametric modulators around those indicator variables. We then tested for areas correlating separately with RTs for eye and hand movements (as a proxy for motor preparation).

In a fourth GLM we re-ran the same analysis as in GLM 1, except this time with the inclusion of additional parametric modulators of RTs for hand and eye movements, in order to establish whether motor preparation as indexed by RTs could account even in part for the regions found to correlate with action-values.

To enable inference at the group level, we calculated second-level group contrasts using a one-sample t-test. Results are reported at  $p < 0.001$  uncorrected in the entire brain and tested in areas of interest at  $p < 0.05$  after small volume correction (SVC) for multiple comparisons.

*ROI analyses.* The effect size plots in Fig 2.2B were computed by averaging GLM's beta values across subjects. In order to ensure the independence of the data used to compute the effect sizes from the data used to select the ROI we performed the following steps. First, for each subject we randomly selected half of the choice trials across the entire

experiment and then created a new design matrix in which we modeled the selected 50% of the trials (T1) and the remaining 50% of the trials (T2) as separate regressors. Similar to the GLM 1 (described above), regressors T1 and T2 each consisted of an onset time regressor and parametric modulators for  $V_e$  and  $V_h$ . Second, to define our ROIs we performed a whole-brain SPM analysis similar to the one shown in Fig 2.1A, but this time restricted to the T1 trials only. Note that the activation map produced by this step (with a threshold set at  $p < 0.005$  unc.) looks very similar to the one shown in Fig 2.2A. Third, we defined the SMA/preSEF ROIs using a 6 mm sphere around the individual subject peak voxel within the activated cluster for T1. Finally, we extracted average effect sizes within these spheres from the remaining T2 regressor. Very similar results were obtained if instead of splitting the data into two groups of trials within subject, data from 50% of the subjects were used to define the ROIs, while data from the remaining 50% were used for extracting the effect sizes (from the co-ordinates defined in the first group of subjects). The effect size plots for vmPFC shown in Fig 2.3B were calculated using an identical procedure.

*Small volume corrections.* Seed region coordinates for small volume correction were defined by two alternative methods: Firstly, we used an anatomical definition for supplementary motor cortex provided by the AAL human brain atlas [98], and we corrected for small volume within the entire area of supplementary motor cortex defined by this atlas (comprising both SEF and SMA), superimposed on the normalized average structural scan from our study. Secondly, we took the average peak co-ordinate from 16 previous fMRI studies identifying activation in SMA and defined a sphere of 12 mm around that averaged peak co-ordinate in which to perform the small volume correction

[55, 99, 100]. The size of the sphere in the functionally defined seed region was set to 1.5 times the size of the smoothing kernel used during preprocessing of the fMRI dataset. Using each and every one of these criteria our effects survived correction for small volume with family wise error at  $p < 0.05$ .

The structural T1 images were co-registered to the mean functional EPI images for each subject and normalized using the parameters derived from the EPI images. Anatomical localization was carried out by overlaying the t-maps on a normalized structural image averaged across subjects, and with reference to an anatomical atlas [97].



Table 2.1 Different types of value signals

Characteristics of each value signal type in terms of the specific variable that activity in a given region should be correlated with as a function of the action performed (choice taken):

<b>Value signal categories</b>	<b>Action performed</b>	
	<i>Eye chosen</i>	<i>Hand chosen</i>
Action_value: eye	$V_{\text{eye}}$	$V_{\text{eye}}$
Action_value: hand	$V_{\text{hand}}$	$V_{\text{hand}}$
Value_chosen	$V_{\text{eye}}$	$V_{\text{hand}}$
Value_chosen: eye only	$V_{\text{eye}}$	-
Value_chosen: hand only	-	$V_{\text{hand}}$

Table 2.2 Activated regions

Locations of significant correlation with parametric contrasts in the fMRI analysis (threshold  $p < 0.001$ ) MNI coordinates denote the group peak voxel of each cluster.

\*  $p < 0.05$  SVC corrected.

*V<sub>h</sub>* (Fig 2.2A):

	<i>x</i>	<i>y</i>	<i>Z</i>	<i>T</i>	# voxels	
<b>01:</b>	-21	-90	-18	6.29	46	<i>left occipital cortex</i>
<b>02:</b>	57	-54	-09	5.68	65	<i>right inferior temporal gyrus</i>
<b>03:</b>	-30	-24	75	5.58	59	<i>left postcentral sulcus</i>
<b>04:</b>	<b>00</b>	<b>-12</b>	<b>78</b>	<b>4.61</b>	<b>33</b>	<b><i>SMA*</i></b>
<b>05:</b>	39	-78	39	4.34	32	<i>left intraparietal sulcus</i>

*V<sub>e</sub>* (Fig 2.2A):

<b>01:</b>	27	15	-03	4.27	5	<i>ventral striatum</i>
<b>02:</b>	<b>-06</b>	<b>09</b>	<b>60</b>	<b>4.16</b>	<b>4</b>	<b><i>preSMA*</i></b>

*V<sub>chosen</sub>* (Fig 2.3A):

<b>01</b>	-39	-36	69	6.37	142	<i>Left postcentral gyrus</i>
<b>02</b>	21	-48	60	5.84	133	<i>Right postcentral sulcus</i>
<b>03</b>	36	-21	39	5.59	24	<i>Right central sulcus</i>
<b>04</b>	-15	-36	45	5.37	34	<i>Left cingulate sulcus</i>
<b>05</b>	-48	-15	51	5.19	50	<i>Left central sulcus</i>
<b>06</b>	<b>06</b>	<b>45</b>	<b>-18</b>	<b>4.76</b>	<b>51</b>	<b><i>Ventromedial prefrontal cortex*</i></b>
<b>07</b>	-60	-09	-06	4.72	38	<i>Sup. Temporal sulcus</i>
<b>08</b>	-54	-30	03	4.43	64	<i>Planum temporale</i>

*V<sub>unchosen-V<sub>chosen</sub></sub>* (Fig 2.4A):

01:	-27	24	00	8.02	249	<i>Left anterior insula</i>
02:	36	24	06	7.0	176	<i>Right anterior insula</i>
<b>03:</b>	<b>03</b>	<b>24</b>	<b>51</b>	<b>5.64</b>	<b>268</b>	<b><i>Dorsomedial frontal cortex &amp; anterior cingulate*</i></b>
04:	-33	-54	42	5.47	48	<i>Intraparietal sulcus</i>
05:	-48	12	36	5.1	214	<i>Inferior frontal sulcus</i>

*Hand bias (Fig S 2.3):*

	<i>x</i>	<i>y</i>	<i>z</i>	<i>T</i>	<i># voxels</i>	
01:	-39	-21	51	7.61	558	<i>Left precentral gyrus</i>
02:	21	-45	-30	6.4	144	<i>Right cerebellum</i>
03:	06	21	66	6.16	158	<i>Right dorsal medial frontal cortex</i>
04:	-39	-66	-39	6.02	168	<i>Left cerebellum</i>
05:	66	-27	-21	5.93	132	<i>Right temporal lobe</i>
06:	21	-39	60	5.76	32	<i>Right central sulcus</i>
07:	-39	-66	57	5.54	83	<i>Left superior parietal gyrus</i>
08:	48	-51	54	5.26	320	<i>Right superior parietal gyrus</i>
09:	30	15	45	5.02	58	<i>Right middle frontal gyrus</i>
10:	33	-27	69	4.92	30	<i>Right precentral gyrus</i>
11:	-12	-42	30	4.57	126	<i>Left posterior cingulated gyrus</i>
12:	45	-72	-33	4.56	44	<i>Right cerebellum</i>

*Eye bias (Fig S 2.3):*

01:	12	-72	12	7.76	2734	<i>Occipital lobe (bilateral)</i>
02:	-27	12	00	5.26	25	<i>Left striatum</i>
03:	24	-42	45	5.08	20	<i>Right parietal cortex</i>

Table 2.3 Model performance

Model performance in predicting individual subject choices of the soft-max model and the CDM model

Subject	<i>soft-max model</i>		<i>CDM model</i>	
	R2	p<	R2	p<
1	0.57	0.000	0.54	0.000
2	0.65	0.000	0.65	0.000
3	0.20	0.000	0.15	0.000
4	0.20	0.000	0.20	0.000
5	0.18	0.000	0.15	0.000
6	0.04	0.019	0.02	0.118
7	0.13	0.000	0.11	0.000
8	0.54	0.000	0.54	0.000
9	0.04	0.011	0.05	0.009
10	0.18	0.000	0.17	0.000
11	0.64	0.000	0.62	0.000
12	0.06	0.003	0.05	0.004
13	0.36	0.000	0.35	0.000
14	0.25	0.000	0.22	0.000
15	0.56	0.000	0.53	0.000
16	0.58	0.000	0.57	0.000
17	0.31	0.000	0.27	0.000
18	0.38	0.000	0.39	0.000
19	0.46	0.000	0.44	0.000
20	0.56	0.000	0.57	0.000
21	0.61	0.000	0.62	0.000
22	0.19	0.000	0.18	0.000
23	0.06	0.003	0.06	0.004
	<b>0.34</b>		<b>0.32</b>	

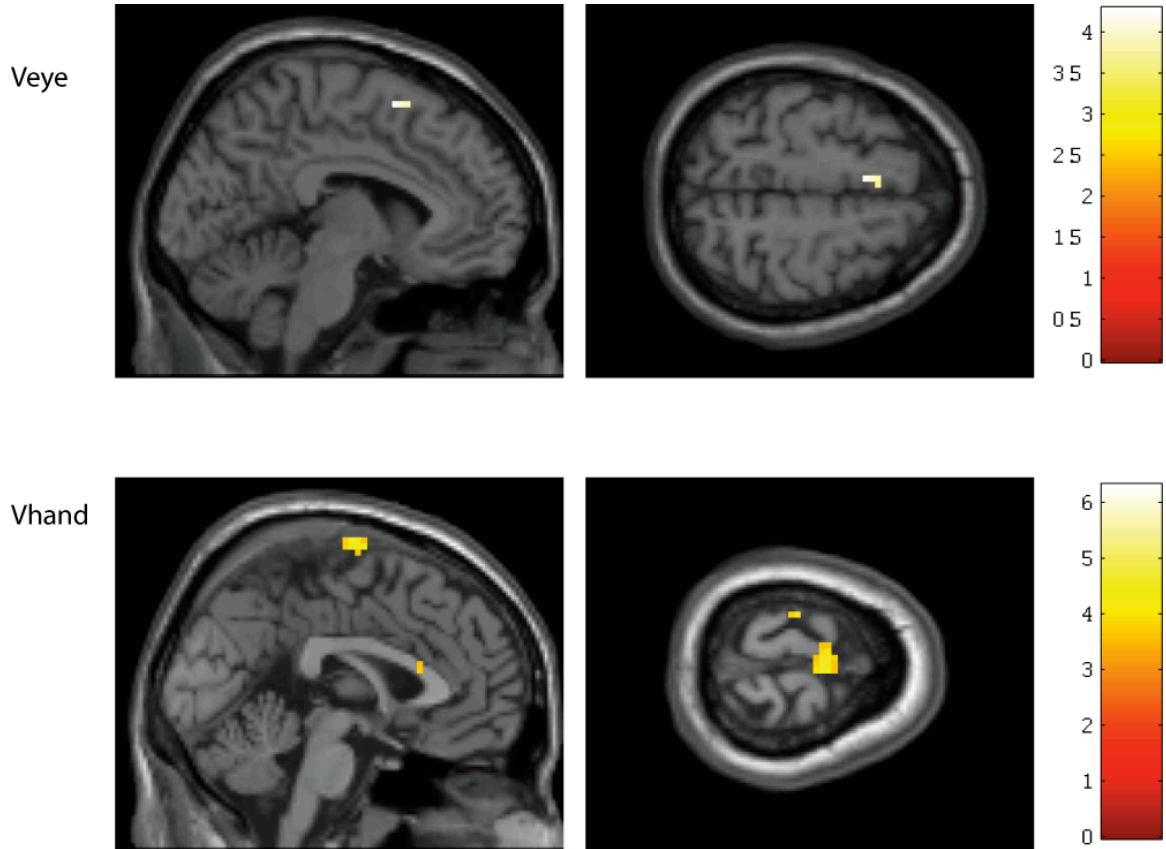


Figure S 2.1 Activations by action values with T-stats coloring

A region of pre-SEF showing correlations with action-values for eye movements (Ve, top) and a region of supplementary motor area showing correlations with action-values for hand movement (Vh, bottom). T-maps are shown from a whole brain analysis thresholded at  $p < 0.001$  uncorrected. The color bars indicate the magnitude of the t-scores. These same two contrasts are also shown combined in Figure 2.2A.

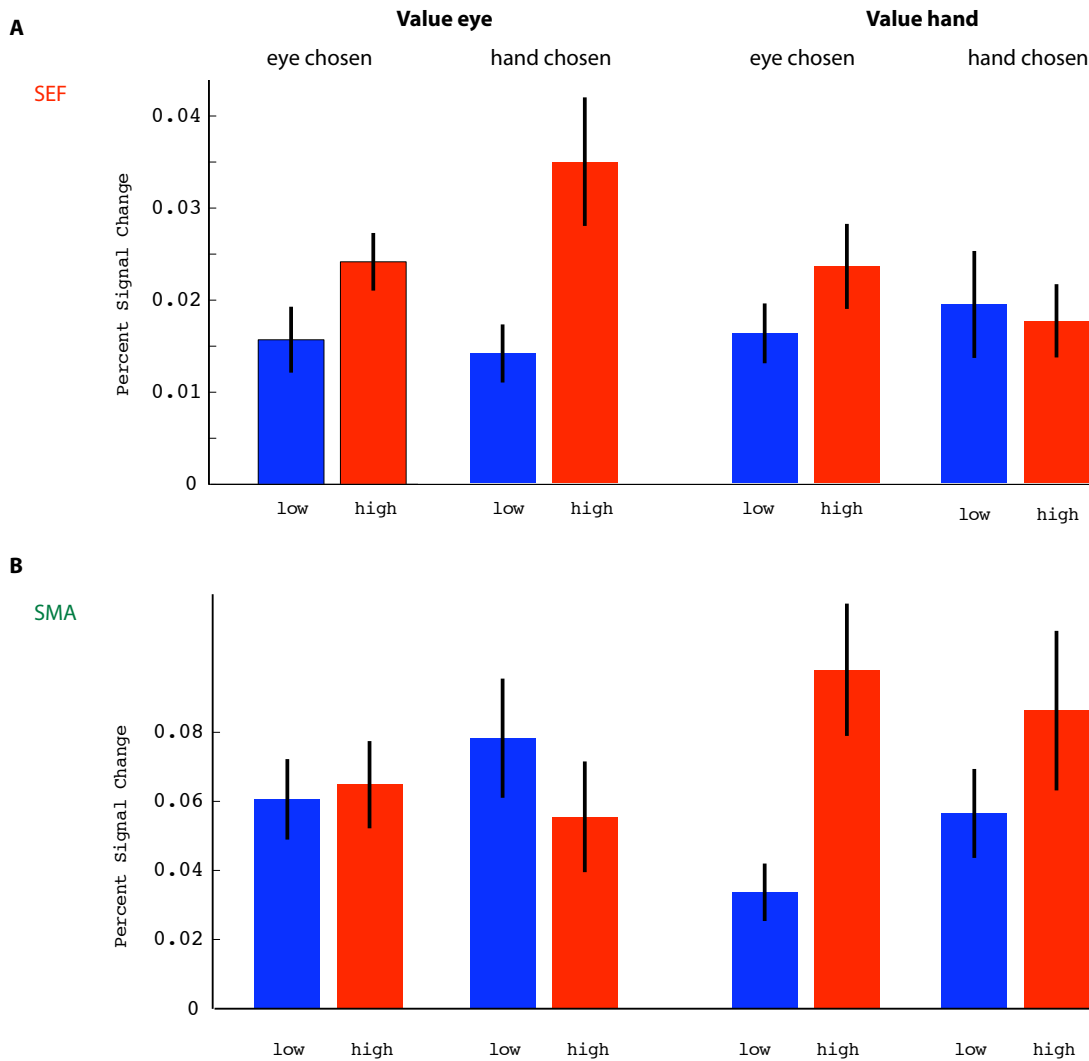


Figure S 2.2 Post-hoc effect sizes

Post-hoc plots of effect sizes (expressed as percent signal change) averaged across all voxels in the activated clusters at the group level for each subject and then averaged across subjects separately for the pre-SEF (shown in A) and SMA (shown in B). (A) The graph on the left hand shows average % signal change extracted from pre-SEF separately for trials in which the action value of the eye movement is low ( $V_{eye} \leq 0.5$  percentile) and high ( $V_{eye} > 0.5$  percentile), further separated by trials in which either the eye movement or hand movement was actually chosen. As expected (given this area correlates with action-values for eye movements), the % change plots discriminate high and low eye values irrespective of whether that action is chosen on that trial. Importantly

however, when activity within the same area is plotted as a function of the action value for the hand movement (shown on the right hand side), the signal change on high and low hand value trials does not discriminate between high and low hand values. Although for eye\_chosen trials the value hand signal does appear to separate in the direction of high and low hand values, this difference is not statistically significant (paired t-test  $t=1.8$ ;  $p<0.08$ ; note this post-hoc comparison for the value of hand is independent of the parametric contrast used to select the voxels (value\_eye)). **(B)** Similar plot for SMA showing that this region distinguishes high and low action\_values for hand movements irrespective of whether that action is chosen, but does not distinguish the value of eye movements.

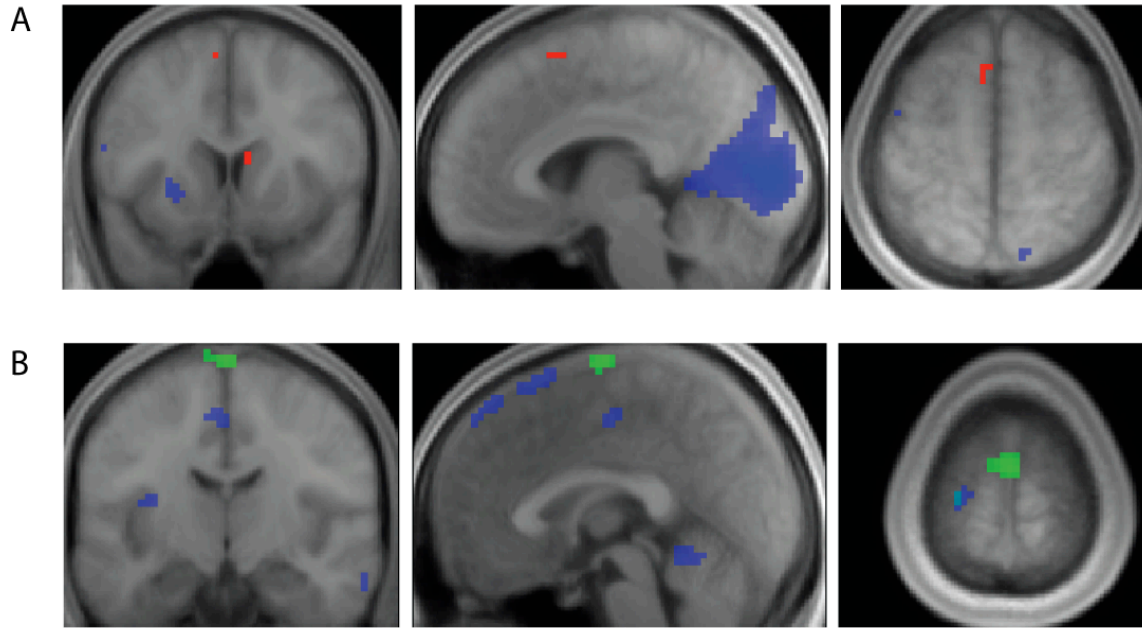


Figure S 2.3 Local predominance bias

(A) Correlations with eye action value  $V_{\text{eye}}$  (red, same as in Figure 2.2A) don't overlap with correlations of the local predominant choice bias for eye (blue). Areas in blue indicate regions that are significantly more active during periods in which the subject predominantly chooses eye. (B) Similarly, the correlation with hand action value  $V_{\text{hand}}$  in SMA (green, same as displayed in Figure 2.2A) is not overlapping with areas that are significantly more active at times when the subject predominantly chooses hand actions (blue).



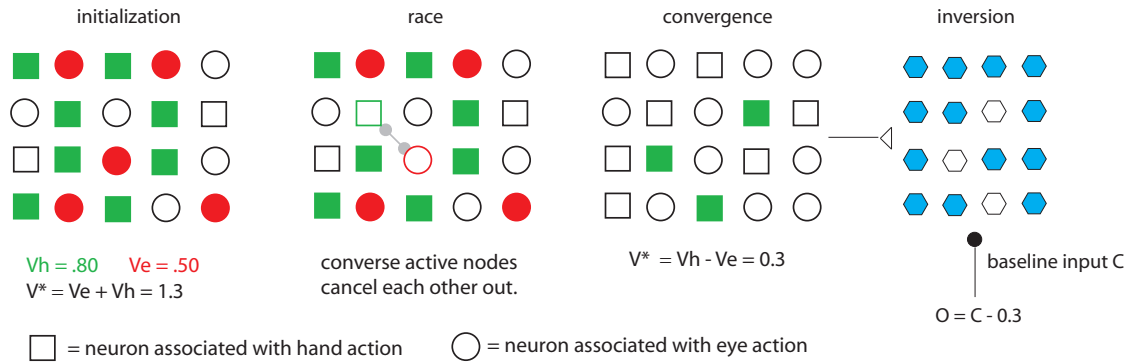


Figure S 2.4 Illustration of the competition difference model

The decision process was modeled as an iterative algorithm of mutual competition between the neuronal populations associated with the valuation of eye and hand actions. Model inputs were action-values for eye and hand movements. The evolution of the model output over time within a trial is illustrated in a hypothetical case with action-values of  $V_h=0.8$  and  $V_e=0.5$  at the time of model initialization (left), during competition between populations (middle), and after convergence (right). The model includes an additional final layer (with constant positive input from which the previous result is subtracted) that inverts the output of the network to compute the value difference between the action not chosen and the action chosen.

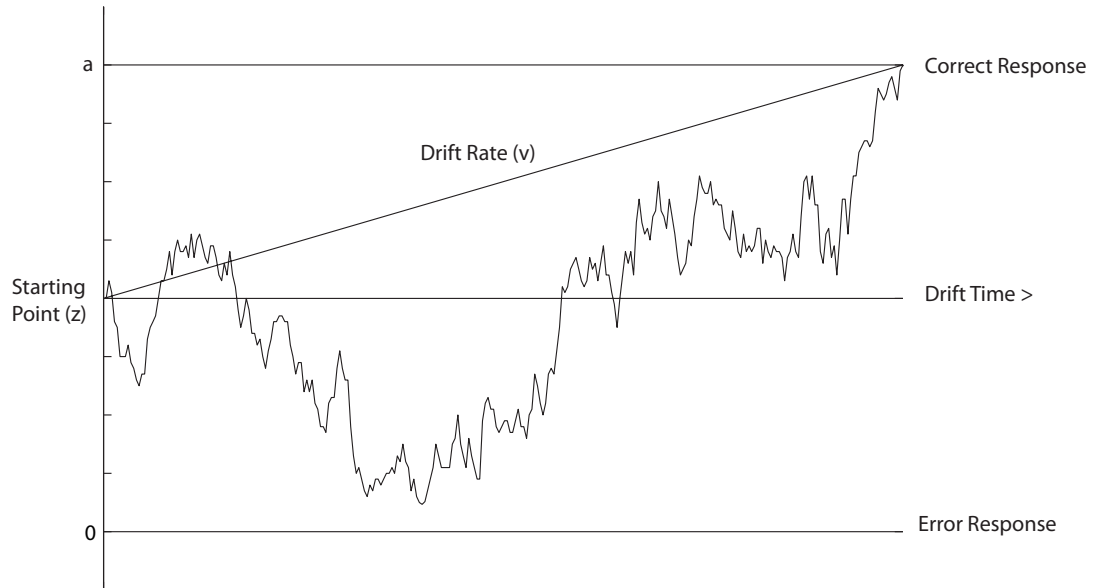


Figure S 2.5 Illustration of the Drift Diffusion Model

During the decision making process the model computes a net value signal (say  $V_h - V_e$ ) that fluctuates between two barriers. A decision is reached when the net value signal crosses either of the two barriers. If the top barrier is crossed the hand action is chosen. If the bottom barrier is crossed the eye action is chosen. The net value signal climbs to the hand barrier with a slope proportional to  $V_h - V_e$ , but it is also affected by white Gaussian noise. In the case depicted in the figure,  $V_h > V_e$  so that hand is the correct choice.

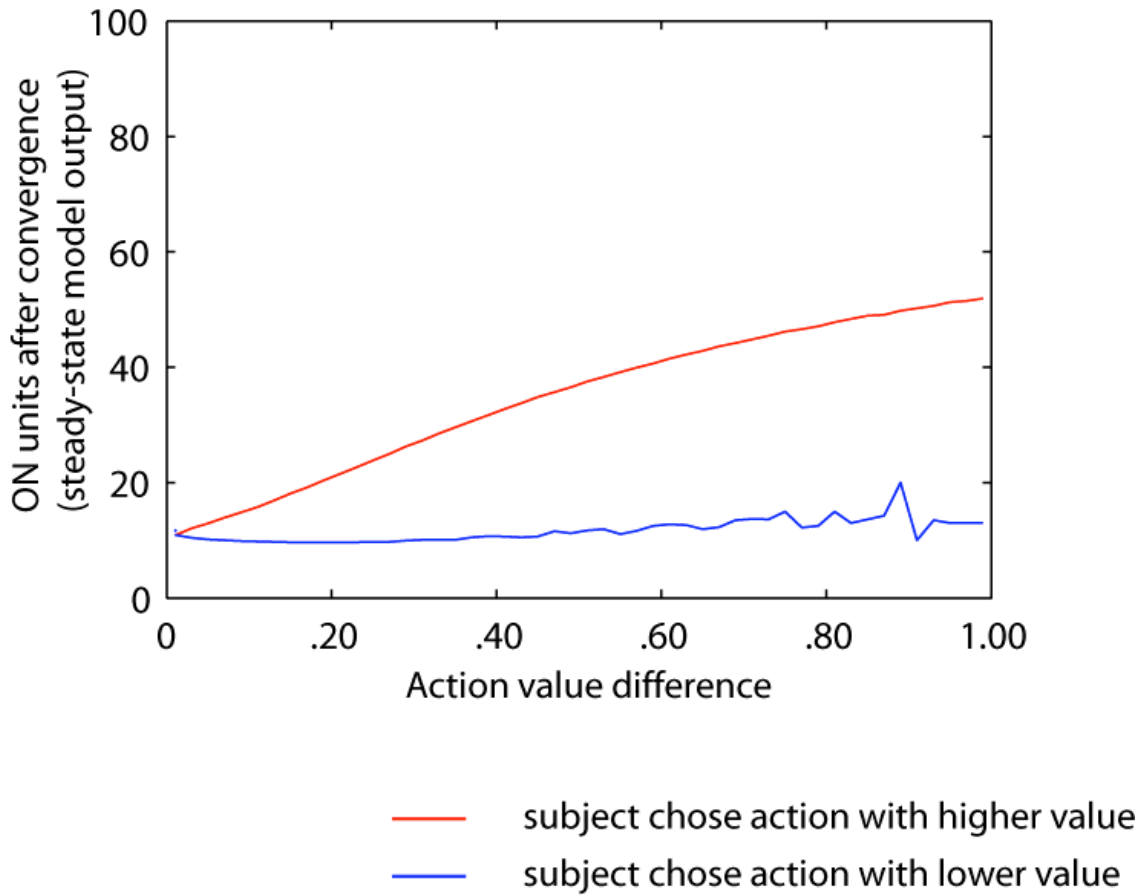


Figure S 2.6 CDM steady state output

Average steady state output of the computational decision model of the choice process as a function of value difference between the two action values. The remaining number of ON units after 50 iterations was averaged across 1000 model runs. The red curve displays the average remaining total ON units in the model in trial that the model converged towards the action with the larger action value. In contrast, the blue curve displays the same statistic for trials in which the model selected the action with the lower action value.

## Chapter 3. Economic choices<sup>ii</sup>

*Decision-making often involves choices between different goods, each of which is associated with a different physical action. A growing consensus suggests that the brain makes such decisions by assigning a value to each available option and then comparing them to make a choice. An open question in decision neuroscience is whether the brain computes these choices by comparing the stimulus values directly (goods space) or by assigning values to the associated actions and then comparing those action values (action-space). We used a novel fMRI experimental design in which human subjects made choices between different stimuli before and after knowing which actions were required to obtain the different stimuli. We found neural correlates of the value of the stimulus that is chosen in a trial (a post-decision signal) in vmPFC before the action pairing was revealed. These findings provide strong evidence for the hypothesis that the brain is capable of making choices completely within an abstract representation of stimuli.*

---

<sup>ii</sup> Adapted with permission from Klaus Wunderlich, Antonio Rangel, John P O’Doherty, “Economic choices can be made using only stimulus values”, (manuscript under review).

## Introduction

Imagine that you are thirsty and walk up to a vending machine that is serving a variety of soft drink beverages. On the machine you see the brand marks of the offered beverages, and since you had previously sampled them, you easily assign values to each drink based on their taste. In order to get the desired beverage you press the button that is distinctively associated with the preferred option. This situation exemplifies many of the decisions that humans and animals make in daily life. It is a well-established belief among economists, psychologists, and neuroscientists that the brain solves such choice problems by first computing a value for each alternative and then selecting the one that has the highest value [5, 6, 8, 9]. Neuroscientists have considered two possible alternative ways for how values might be compared to make a choice in these situations:

In the action-based model, choices are embedded in premotor processes of action selection: the values of goods are passed as action values to the motor plans required to obtain them and the decision is then made within the action space [17, 36, 101]. Several variants of this approach have been proposed [11, 102-104]. For example, Glimcher extended this theory to values that are acquired through experience by a reinforcement learning mechanism such that a value is attached to each possible course of action. Evidence for this view comes from the finding of action-value signals in several regions of the brain including the caudate nucleus [17, 36], supplementary motor cortex [105], and action-related value signals in lateral intraparietal cortex [31, 33].

In contrast, in the stimulus-based model, values of the available goods are compared to make a choice in the absence of any action information (i.e., the choice takes place in

stimulus-value space), and only after a stimulus is chosen are the necessary motor plans identified and executed. Thus, this view proposes a sequential choice process in which action selection is temporally separated from the actual process of choice. Support for this model comes from studies showing abstract value representations in the orbitofrontal cortex [10, 51, 52], and from lesion studies indicating double dissociations between lesions of orbitofrontal cortex and anterior cingulate cortex on learning of stimulus-reward and action-reward associations respectively [106, 107]. However, while there is considerable evidence for stimulus-values and stimulus-based learning, it is as yet unknown whether such signals can be actually used to compute choices, or whether by contrast such signals need to be converted into action space before choice signals can be computed.

The aim of the present study was to directly address this question. For this we used fMRI in human subjects while they performed a variant of a two-armed bandit task in order to obtain probabilistically delivered monetary rewards (Figure 3.1A). In every trial, subjects made a choice between two stimuli and selected one by executing the action that was randomly paired with the chosen stimulus (either button press or saccade). A critical feature of the task was that in half of the trials (stimulus condition, SC), subjects were first presented with the two stimuli alone in a horizontal arrangement that did not contain any information about the actions required to obtain them. The actions were only revealed after a variable interval by randomly flipping the stimuli in vertical alignment. At this stage subjects could choose the upper stimulus by making a saccade to a target in the right hemifield, and the lower stimulus by pressing a button with their right hand. In the other half of the trials (action condition, AC) the first screen was not shown and

instead the stimuli appeared immediately in the vertical action-pairing position. To avert the possibility that subjects had already formed a decision in the preceding inter-trial-interval, subjects were in each trial presented with a choice between two out of three possible stimuli (triangle, square, and circle) in pseudo-random appearance so that that they did not know until the trial onset which pair of stimuli would be presented. The probability of being rewarded on selecting each of the three stimuli drifted randomly over time and was independent of the probability of being rewarded on the others (Figure 3.1B). We estimated the value of taking each stimulus in every trial by calculating the stimulus values using a computational reinforcement-learning (RL) model in which the value of each stimulus,  $V_{\text{triangle}}$ ,  $V_{\text{square}}$ , and  $V_{\text{circle}}$ , was updated in proportion to a prediction error on each trial. The model also assumed that stimulus selection in every trial followed a soft-max probability rule based on the difference of the estimated values, which provided a good description of behavior (Figure 3.1C).

We reasoned that if choices can be computed in stimulus space, we would observe signals corresponding to the value of the option that was subsequently chosen in SC trials at the time of stimulus presentation, but before any action-related information was made available to the subjects, so that the choice could be implemented. This hypothesis is important because, if correct, it provides evidence that the brain can compute choices in stimulus-space. We also speculated that subjects would respond faster in SC trials than in AC trials because the time necessary to make a decision between the desired stimuli had already been provided before the action pairing.

## Results

Consistent with the reaction time hypothesis, we found that subjects responded significantly faster (paired t-test;  $p < 10^{-11}$ ) in the SC than the AC condition (Figure 3.1D).

In order to look for neural correlates of value signals we entered the trial-by-trial estimates of the values of the two stimuli under consideration into a regression analysis against the fMRI data. We focused our search for decision related value signals on the ventromedial prefrontal cortex (vmPFC), a region that has been found to encode the value of the chosen stimulus or action. We found that in SC trials neural activity in vmPFC ( $x=-3, y=27, z=-9; T=3.73$ ) correlated with the value of the stimulus that is subsequently chosen already before the stimulus-action pairing was revealed (Figure 2.2A, Table 3.1). Importantly, we also tested for a value chosen signal in SC trials at the time of action-pairing but did not find any significant correlation for this contrast. In AC trials the vmPFC ( $x=-6, y=39, z=-12; T=5.55$ ) also correlated significantly with the value chosen signal. Although the peak of the area encoding the chosen value in AC trials was found to be located slightly more anterior and ventral than the peak in SC trials (Figure 3.2A, Table 3.2), we did not find any significant activated area for the difference contrasts “SC value\_chosen during stimulus presentation – AC value chosen” at a liberal threshold of 0.01 uncorrected. Effect size plots (Figure 3.2B) and timecourse plots (Figure 3.2C) in the overlapping area (center at  $x=-9, y=42, z=-3$ ) confirmed that in SC trials activity in vmPFC was correlated with the chosen value only at the time of stimulus presentation but not at the succeeding time of stimulus-action pairing. All reported activations are significant at a height



threshold of  $p < 0.005$  and a cluster extent threshold of  $p < 0.05$  corrected for multiple comparisons estimated using alpha-sim. Note that the data used to calculate effect sizes were independent of the data used in the functional definition of the region-of-interest (see Supplementary Materials for details). Timecourses were averaged separately for trials in which the chosen value was large and small. Again, timecourses separate according to the value signal in SC trials already at the time of the stimulus presentation, which precedes the action-pairing screen by 3-5 seconds.

We also looked for representations of the individual stimulus values, because clearly such signals should be a precursor of choice in that these values need to be compared in order to work out which option is ultimately chosen. Due to spatial limitations of fMRI we assumed that it would likely not be possible to detect activity patterns encoding the value of the individual stimuli. Instead, we assumed that neurons encoding such values would be spatially intermixed within the same region, and that such intermixed neural signals would be reflected at the level of the BOLD signal as an average of the values of the two stimuli under consideration on a given trial. Consistent with our hypothesis, we found such an averaged stimulus-value signal in SC trials within a sub-region of vmPFC ( $x = -9$ ,  $y = 48$ ,  $z = -3$ ,  $T = 4.41$ ) (Figure 3.3, Table 3.3). Furthermore, consistent with the notion that such individual stimulus-values only need to be used by the brain at the time of decision making, we only found evidence for such an averaged stimulus-value signal at the time of stimulus presentation (and not once the action pairings were presented).

## Discussion

We used a novel fMRI experimental design in which human subjects made choices between different stimuli before and after knowing which actions were required to obtain the different stimuli. We found neural correlates of a post-decision signal, the value of the stimulus that is chosen in a trial, in vmPFC before the action pairing was revealed. These findings indicate that the brain is capable of computing choices completely within an abstract representation of stimuli.

One possible alternative explanation of our findings is that in the stimulus condition (at the time of presentation of the stimuli but before the action pairings are revealed), subjects make decisions by assigning temporary action pairings to the stimuli, and then comparing the temporary action pairings. These temporary action pairings could then be substituted for the real action assignments at the time of action presentation. Although we cannot completely rule out these explanations on the basis of our data alone, there are several reasons why this type of explanation is unlikely to account for our results. First, when subjects are in a situation where it is necessary to make decisions over actions (i.e. where there is no unique stimulus information to discriminate between different options), regions of the brain known to be involved in motor planning and initiation such as supplementary motor cortex, lateral intraparietal cortex, and anterior cingulate cortex have been found to contain action-related value signals in prior imaging studies [79, 105]. However, these regions did not show significant correlations with value signals in the present paradigm even at a liberal uncorrected threshold ( $p < 0.01$ ), suggesting that neural systems involved directly in action representations were not directly engaged during the decision process in the present study. Secondly, on a more conceptual level, while

encoding of conditional action pairings might be feasible in the present simplified experimental paradigm, such a mechanism is unlikely to scale well in many real-world sequential decision problems with large numbers of sequential conditional action pairings, because decisions in such contexts would require encoding of long strings of conditional action pairings that could rapidly become computationally intractable. By contrast, the parsimonious alternative proposed here whereby in such contexts a decision is made between the stimuli would not suffer from the same scaling problem.

Our findings provide the first direct evidence that the brain is capable of computing choices completely within an abstract representation of stimuli. It is important to emphasize that our data does not show that all decisions are made in stimulus-space, but rather that the brain is capable of computing a decision purely in stimulus space when action pairings are not available. There is ample experimental evidence that behavioral decisions can be and are made over actions in many contexts [7, 79, 105, 108-110]. Given the results in this paper, plus the presence of action value signals in caudate [17, 36] and supplementary motor system [105] it is natural to conjecture that both mechanisms may co-exist during certain types of choices, or that some types of choice may be better computed in stimulus-space and others in action-space. The current study provides direct evidence that the brain is capable of computing decisions in stimulus space even when the actions required to implement the choices are not available.

## Methods

### *Subjects*

24 healthy subjects (18–31 years old; right-handed, assessed by self-report with an adapted version of the Edinburgh handedness inventory [96]) with no history of neurological or psychiatric illness participated in the study. The Institutional Review Panel of the California Institute of Technology approved the study.

### *Task*

The task is a variant of a 2-armed bandit problem in which subjects make pair wise choices between subsets of two stimuli that were pseudo-randomly selected out of three stimuli used in the experiment: a green triangle, a blue square, and a yellow circle.

There were two conditions. In the first one (stimulus condition, SC), subjects were first presented with the stimuli in horizontal arrangement without the information of what action they had to perform to choose the stimuli. After a variable time (3, 4, or 5 seconds, uniform distribution) the stimuli flipped to vertical position that indicated the action associated with each stimulus. The assignment of stimuli to actions was made randomly in every trial. At this stage, subjects could press a button with their right index finger to choose the bottom stimulus or perform a saccade from a central fixation cross to a target located at 10 degrees of visual angle in the right hemifield to choose the top stimulus.

In the second condition (action condition, AC) the trials were identical except that the first screen was not shown and subjects were immediately presented with the stimuli in vertical arrangement at the beginning of the trial.

The probability ( $Q_{i,t}$ ) of stimulus  $i$  being rewarded in trial  $t$  evolved over time as a decaying Gaussian random walk process, with  $Q_{i,t+1} = \max(0, \min(1, \lambda Q_{i,t} + (1 - \lambda)\theta + \nu))$ ; where the decay parameter  $\lambda$  was 0.79836, the decay center  $\theta$  was .50, and the diffusion noise  $\nu$  was zero-mean Gaussian with standard deviation  $\sigma_d = .208$ . Five different probability trajectories were generated using this method and were assigned across subjects randomly. Figure 3.1B depicts one of the five probability paths used in the experiment. An important feature of this design is that the probability of being rewarded on one of the three stimuli is independent of the probability of being rewarded on the others. This feature is useful because it implies that the reinforcement learning-based estimates of the stimulus values are uncorrelated with each other, which increases our ability to dissociate the neural correlates.

The task consisted of four sessions of 75 trials each separated by a short break. Subjects had to choose between two actions within 2.5 seconds after onset of the stimulus-action pairing screen; otherwise the trial was counted as an invalid missed trial. Subjects very rarely failed to make a response within this time window: none of the subjects had more than two such events during the entire experiment, and most subjects did not miss any trials at all. After the response was registered the screen changed to a fixation cross until six seconds after trial onset. At this time the outcome was displayed for one second by

showing either an image of a dollar bill in rewarded trials, or a scrambled dollar bill in non-rewarded trials. Trials were separated by a fixation cross that lasted between 1 and 8 seconds (uniform distribution).

Prior to the experiment subjects received full instructions about the task and the two conditions, they were informed that the probabilities of being rewarded on each stimulus changed as a continuous function over time (but were not given details about the underlying stochastic process), and they were instructed to try to maximize their earnings which were paid to them at the end of the experiment. Subjects accumulated 25 cents in each rewarded trial. We did not reveal the exact reward per trial to subjects before the experiment but instead instructed them only that they would get a small amount of money for each rewarded trial. At the end of the experiment subjects were paid their accumulated earnings in addition to a flat amount of 20\$.

The task was presented to the subjects via back projection on a translucent screen, viewable through a headcoil mounted mirror. Subjects chose the hand action by pressing a button on a button box with their right index finger. Eye positions were monitored at 120 Hz with a long-range infrared eye-tracking device (ASL Model L6 with control unit ASL 6000, Applied Science Laboratories, Bedford, MA). An eye action during the choice period was registered when the median horizontal eye coordinate during the past 200 msec exceeded 8 degrees of visual angle to the right from fixation. Subjects were instructed to maintain central fixation during the entire experiment when not deliberately making a saccade.

### *Reinforcement Learning (RL) model*

RL is concerned with learning the value of taking particular actions in different states of the world in a model-free environment in which subjects do not have complete knowledge about the underlying reward generating process. Thus, it is ideally suited to model how subjects learn the value of taking the different actions over time.

In this study we used Q-learning, where action values are updated using a simple Rescorla-Wagner rule. If a stimulus is not selected in a trial its value is not updated. In contrast, if stimuli  $s_1$  and  $s_2$  are shown and  $s_1$  is selected on trial  $t$ , its value is updated via a prediction error,  $\delta$ , as follows:

$$V_{s_1}(t+1) = V_{s_1}(t) + \eta\delta(t),$$


where  $\eta$  is a learning rate between 0 and 1. The prediction error is given by

$$\delta(t) = r(t) - V_{s_1}(t).$$

Probabilistic rewards were delivered in free choice and forced choice trials and these trials were included in updating value prediction in the RL-model. The value of the stimulus that was not presented in a trial ( $s_3$ ) was not updated.

To generate choices, we first used a soft-max procedure where in every trial, the probability (P) of choosing stimulus  $s$  is given by:

$$P_{s_1,t} = \sigma(\beta(V_{s_1}(t) - V_{s_2}(t))) - \alpha$$

where  is the Luce choice rule or logistic sigmoid,  $\alpha=0$  denotes the indecision point (at which both actions are selected with equal probability), and  $\beta$  determines the degree of stochasticity involved in making decisions.

The model decision probabilities  $P_{s1}$  and  $P_{s2}$  were fitted against the discrete behavioral data  $B_{s1}$  and  $B_{s2}$  in order to estimate the free parameters ( $\eta$  and  $\beta$ ). This was done using maximum likelihood estimation and a log likelihood function given by:

$$\log L = \frac{\sum B_{s1} \log P_{s1}}{N_{s1}} + \frac{\sum B_{s2} \log P_{s2}}{N_{s2}},$$

where  $N_{s1}$  and  $N_{s2}$  denote, respectively, the number of trials in which  $s1$  and  $s2$  were chosen, and  $B_{s1}$  ( $B_{s2}$ ) equals one if  $s1$  ( $s2$ ) was chosen in that trial, and zero otherwise.

We compared the choice probabilities predicted by the RL model using the soft-max procedure to subjects' behavior by binning  $P_{s1}$  into 10 bins (bin size=0.1) and calculating for each bin the fraction of trials in which subjects chose  $s1$ . To test the fit between the model and the behavioral data we performed a linear regression subject-by-subject of the fraction of choices on the binned choice probability versus the predicted bin. Overall, the regression results suggest that the model captures actual action value estimation and choice behavior well (Figure 2.1C).

The results presented in the paper are based on analyses in which all subjects' behavior was restricted to be generated by a single learning rate for all subjects, but in which subject-specific heterogeneity was allowed in fitting the parameter  $\beta$  for controlling



choice stochasticity. We also performed the same analyses using a version of the model with fully individualized parameter fits. Although these alternative results support the same general conclusions, we focus on the case of a shared learning rate for several reasons. First, the examination of estimated Hessians of the likelihood at the optima suggested that the parameters were better identified in the restricted case. Second, the action values estimated using the fully individualized model correlated less strongly with fMRI measurements, an effect that has also been observed in a number of previous model-based fMRI studies, suggesting that individual subject parameter fits are prone to being over-fitted to individual behavior [38, 48, 49, 111].

#### *fMRI data acquisition*

Data were acquired with a 3T scanner (Trio, Siemens, Erlangen, Germany) using an eight channel phased array head coil. Functional images were taken with a gradient echo T2\*-weighted echo-planar sequence (TR = 2.65 s, flip angle = 90°, TE = 30 ms, 64 × 64 matrix). Whole brain coverage was achieved by taking 45 slices (3 mm thickness, no gap, in-plane resolution 3 × 3 mm), tilted in an oblique orientation at 30 deg to the AC-PC line to minimize signal dropout in OFC. Subject's head was restrained with foam pads to limit head movement during acquisition. Functional imaging data were acquired in four separate 370-volume runs, each lasting about 16 min. A high-resolution T1-weighted anatomical scan of the whole brain (MPRAGE sequence, 1x1x1 mm resolution) was also acquired for each subject.

*fMRI data analysis*

Image analysis was performed using SPM5 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, U.K.). Images were first slice time corrected to TR/2, realigned to the first volume to correct for subject motion, spatially normalized to a standard T2\* template with a voxel size of 3 mm, and spatially smoothed with a Gaussian kernel of 8 mm FWHM. Intensity normalization and high pass temporal filtering (using a filter width of 128 s) were also applied to the data.

First, we estimated a GLM with AR(1) for each individual subject. The following events were modeled in each trial:

- The time of the stimulus presentation in SC trials, parametrically modulated by the trial-by-trial stimulus values  $V_{s1}$  and  $V_{s2}$
- The time of the stimulus-action pairing in SC trials, parametrically modulated by the trial-by-trial stimulus values  $V_{s1}$  and  $V_{s2}$
- The time of the stimulus-action pairing in AC trials, parametrically modulated by the trial-by-trial stimulus values  $V_{s1}$  and  $V_{s2}$
- The time of the presentation of the outcome, modulated by the prediction error  $\delta$  and a binary function encoding whether a reward was given or not.

Trials in which subjects chose the eye action and trials in which subjects chose the hand action were modeled in separate regressors. All regressors were convolved with the canonical hemodynamic response function. In addition, the 6 scan-to-scan motion

parameters produced during realignment and two session constants were included as additional regressors of no interest.

Second, we computed contrasts of interest at the individual level using linear combinations of the regressors described above. Finally, to enable inference at the group level, we calculated second-level group contrasts using a one-sample t-test.

The activated voxels of all reported results are statistically significant at a threshold of  $p < 0.05$ , corrected for multiple comparisons, as stipulated by Monte Carlo simulations (AlphaSim within AFNI)[112]. AlphaSim generates an estimate of overall cluster size significance level by iteration of the process of random image generation, Gaussian filtering to simulate voxel correlation, thresholding, image masking, and tabulation of cluster size frequencies. In our simulation we generated a series of 10,000 random images, each having  $N$  (number of voxels in masked epi images) spatially uncorrelated voxels by filling the masked brain volume with independent normal random numbers. The effect of voxel correlation was simulated by convolving the random image with a Gaussian function of the size of our smoothing kernel (8 mm FWHM). The image was then scaled to provide our individual voxel probability threshold  $p_{thr} = 0.005$  by determining the value  $z_{thr}$  such that approximately  $p_{thr} * N$  voxels have intensity greater than  $z_{thr}$ . The thresholding was then accomplished by setting those voxels with intensity greater  $z_{thr}$  to 1 (activated voxels), voxels with intensity less than  $z_{thr}$  to 0. Finally, AlphaSim determined which activated voxels belong to clusters. Once all clusters had been found, the size of each cluster in voxels was recorded in a frequency table. This simulation estimated that in a 3D volume (entire brain as masked by the real epi-images)

a cluster size of  $> 51$  contiguous activated voxels would occur by chance with a probability of less than 0.05.

For visualization, results are reported in the figures at  $p < 0.005$  uncorrected in the entire brain.

The structural T1 images were co-registered to the mean functional EPI images for each subject and normalized using the parameters derived from the EPI images. Anatomical localization was carried out by overlaying the t-maps on a normalized structural image averaged across subjects, and with reference to an anatomical atlas [97].

The effect size / timecourse plots in Figure 3.2B, Figure 3.2C, and Figure 3.3B were computed by averaging the GLM's beta values / timecourse data across subjects. In order to ensure the independence of the data that we used to compute the effect sizes from the data used to select the ROI, we performed the following leave-one-out (LOO) analysis. First, we looped through all subjects and computed group averages for all but one subject. We then extracted the beta value from the LOO-group peak voxel of the subject that was excluded in this LOO-group. Finally, we averaged all extracted data.

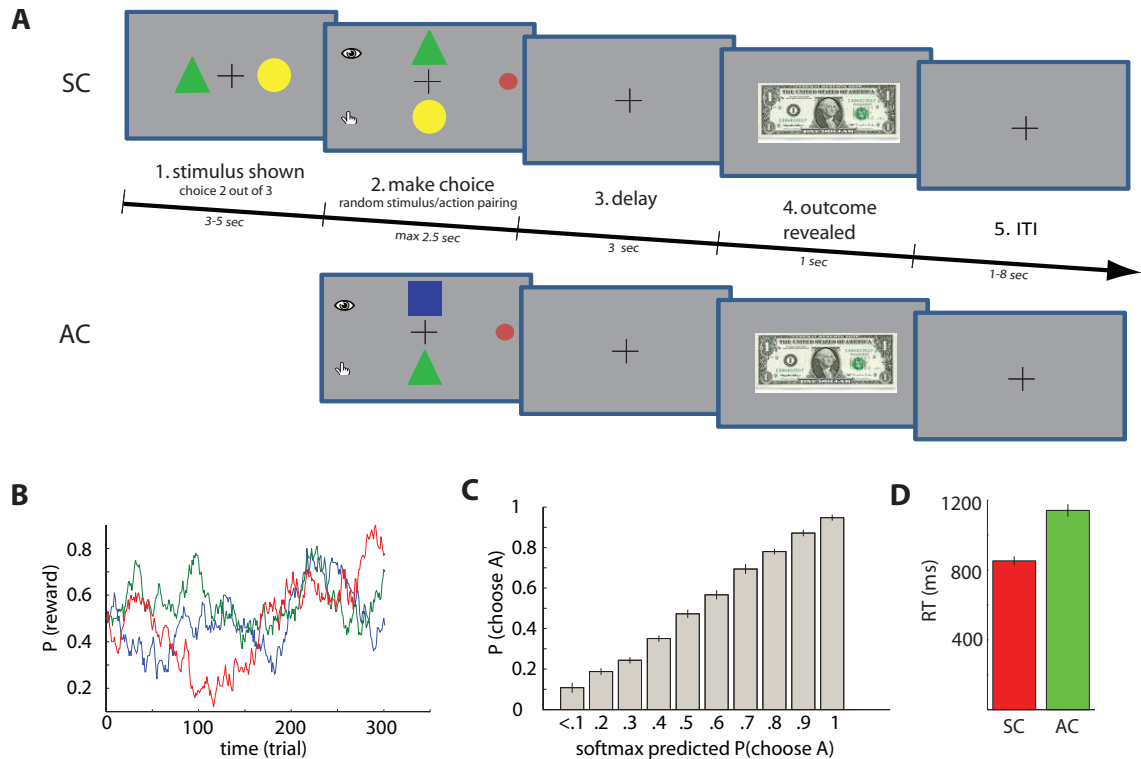


Figure 3.1 Experimental Design and Behavior

(A) Subjects were presented with two stimuli (in every trial pseudo-randomly selected out of 3 possible stimuli) in horizontal arrangement (screen 1, for a variable time between 3-5 s). Stimuli then flipped to a vertical arrangement indicating the actions required to obtain each stimulus (making a saccade or by pressing a button, screen 2). Once a response was registered the screen was immediately cleared for a short delay and subsequently the outcome was revealed (screen 4 at 6 s after screen 2), indicating either receipt of reward or no reward. There were two conditions: a stimulus condition (SC) as just described and action condition (AC), in which the first screen was not shown and subjects immediately saw the stimulus-action pairing. (B) Example reward probability paths for the 3 stimuli as a function of the trial number. The probabilities of being rewarded following choice fluctuated slowly and independently for each stimulus across the experiment. (C) Actual choice probability plotted against fitted model choice probability (binned in .1 wide), averaged across subjects (lines represents s.e.m.). (D) Reaction time (after the action pairing is revealed in screen 2) is significantly lower in SC trials than in AC trials (paired t-test,  $p < 10^{-11}$ , vertical lines represent s.e.m.).

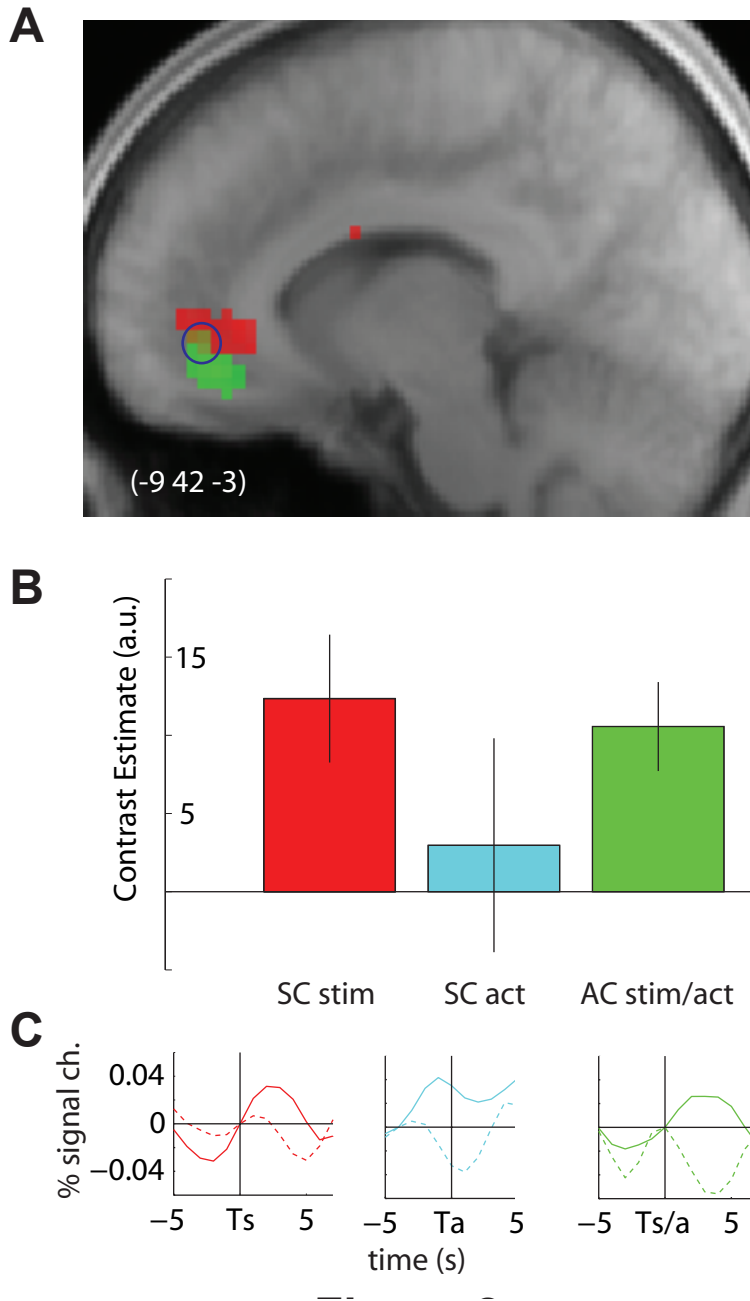


Figure 3.2 Neural correlates of chosen value

(A) Activity in vmPFC showed significant correlation with the value of the stimulus that was subsequently chosen before the stimulus-action pairing was revealed (SC trials, red). The value chosen signal in AC trials was represented slightly more ventrally (green). Activations survive correction for multiple comparisons as described in the methods. For visualization the threshold in this figure is set to  $p < 0.005$  unc. (B) A comparison of effect

size at the overlapping region confirms that in SC trials the value chosen is represented only at the time of the stimulus screen (red) but not at the time of the following stimulus-action pairing (cyan). In AC trials the value chosen is represented at the coinciding stimulus/action screen (green). Bars indicate standard error (C) Event-related BOLD responses in SC trials time locked to the stimulus presentation (left), the stimulus-action pairing (middle), and in AC trials time locked to the coinciding stimulus/action pairing (right). Time courses are plotted separately for trials in which the chosen values were small (dashed,  $V < 0.5$ ) and large (solid,  $V > 0.5$ ). Note that consistent with the effect sizes shown in (B), timecourses split in SC trials after stimulus presentation (Ts, left) but already approximately 4 s before the action-pairing (Ta, middle).

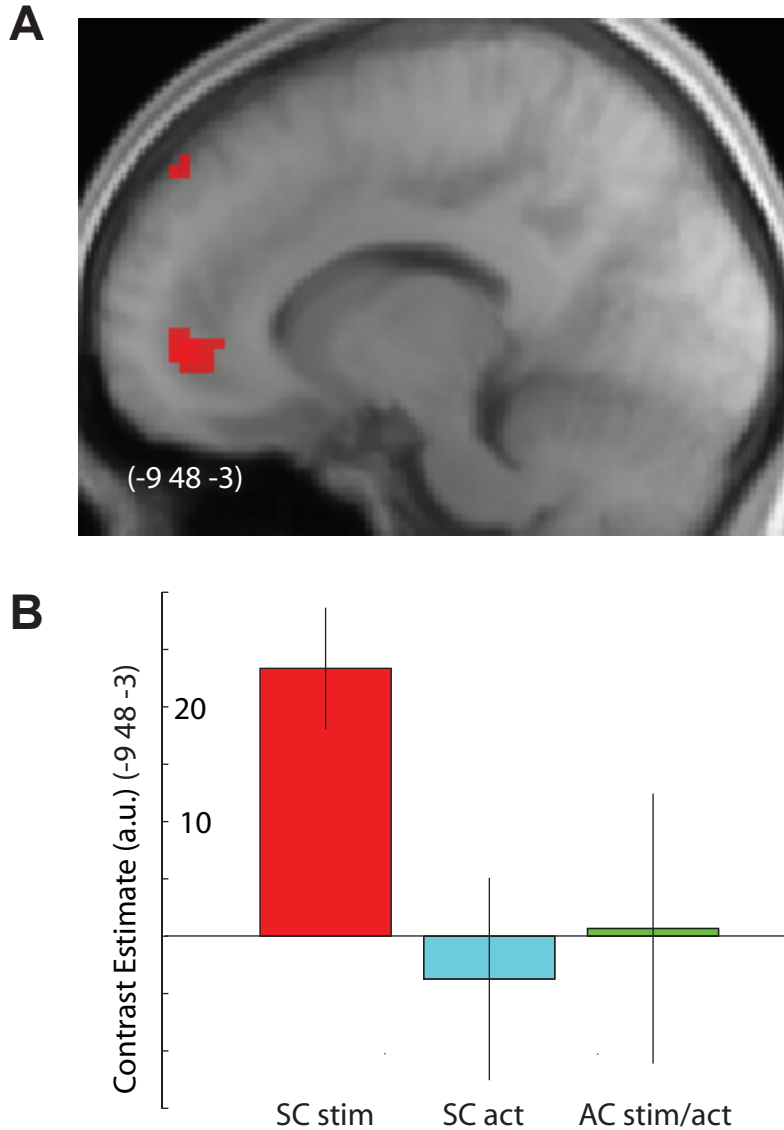


Figure 3.3 Neural correlates of stimulus value

(A) Activity in vmPFC showed significant correlation with the average value of the two stimuli that were presented in SC trials at the time of screen 1 (red). Activations survive correction for multiple comparisons as described in the methods. For visualization the threshold in this figure is set to  $p < 0.005$  unc. (B) Comparison of effect sizes at the peak region between conditions. Neural activity in this region correlated with stimulus values in SC trials but not in AC trials. Bars indicate standard errors.



Table 3.1 Activated regions value chosen SC

Locations of significant correlation with chosen value in SC trials during the stimulus screen (threshold  $p < 0.05$  corrected for multiple comparisons). MNI coordinates denote the group peak voxel of each cluster.

	<i>x</i>	<i>y</i>	<i>Z</i>	<i>T</i>	<i># voxels</i>	
01:	51	-3	-	4.91	71	<i>Right posterior insula cortex</i>
			3			
02:	-	6	6	4.87	110	<i>Left posterior insula cortex</i>
	36					
03:	-3	27	-	3.73	112	<i>Rostral ACC</i>
			9			

Table 3.2 Activated regions value chosen AC

Locations of significant correlation with chosen value in AC condition during the stimulus/action screen (threshold  $p < 0.05$  corrected for multiple comparisons). MNI coordinates denote the group peak voxel of each cluster.

	<i>x</i>	<i>y</i>	<i>z</i>	<i>T</i>	<i># voxels</i>	
01:	-	39	-	5.55	105	<i>vmPFC</i>
	6		12			
02:	-	-	24	4.23	128	
	2	48				
	4					

Table 3.3 Activated regions stimulus values

Locations of significant correlation with the average stimulus value during the stimulus screen of the SC condition (threshold  $p < 0.05$  corrected for multiple comparisons). MNI coordinates denote the group peak voxel of each cluster.

	<i>x</i>	<i>y</i>	<i>z</i>	<i>T</i>	<i># voxels</i>	
<b>01:</b>	-	48	-	4.41	53	<i>vmPFC</i>
	9		3			

## Chapter 4. Optimal integration of multiple evidences.<sup>iii</sup>

*When trying to understand the causal nature of events in the world, an individual may consider a number of candidate theories, yet optimal decisions depend crucially on identifying the correct theory. Here, we present behavioral and neuroimaging evidence that humans tend to solve such problems not by first picking the theory that is most likely correct and then choosing accordingly (the “attention-gated” approach), but by considering all possible explanations simultaneously. We used model-based fMRI analysis in a hierarchical reversal decision task and found that decision variables based on the integrated approach were better able to explain activity in prefrontal cortex than those generated by the attention-gated approach. Furthermore, between-subject variance in the degree to which subjects’ deployed an integration strategy was correlated with the strength of the integration decision variables in prefrontal cortex. Our results demonstrate that the human brain and the prefrontal cortex in particular is capable of integrating information in an optimal manner, similar to that of an ideal Bayesian observer.*

---

<sup>iii</sup> Adapted with permission from Klaus Wunderlich, Ulrik Beierholm, Peter Bossaerts, John P O’Doherty, “The human prefrontal cortex mediates optimal integration during inference about multiple causes”, (manuscript under review).

## Report

In numerous real-life situations we face unknown causal relationships and have to find out which of multiple, and sometimes contradictory, causes explain the observed phenomena. One solution to deal with this problem is to first choose the most likely explanation and handle the situation accordingly until enough evidence accumulates to reject the theory as invalid. This approach has been the core of scientific deductive reasoning and formalized in hypothesis testing of classical statistics. By looking at many real life situations across a large variety of fields from medical practice to economics to monetary policy making, one observes that humans often follow an "attention-gated" approach of classical hypothesis testing, in that at any one time, attention is focused on one possible explanation to the relative exclusion of all others. Also at a much higher level, most dogmas in religions are only occasionally revised and scientific theories are accepted until proven false, even if there are concurrent observationally equivalent propositions – "science as falsification", as Karl Popper puts it [113].

However, an alternative problem-solving strategy is to simultaneously consider multiple theories and use a weighting of all the possible causes. No theory, even the least likely, is thereby ever excluded. Such an integrative approach is the mathematically optimal strategy and typifies Bayesian analysis [114].

Using a hierarchical reversal learning task, we tested whether humans indeed follow an attention-gated approach to problem solving or rather the integrative procedure whereby less likely explanations for observed phenomena still influence decisions. We did so in two ways: first, we studied to what extent participants' choices revealed sensitivity to

more than one theory at a time; second, using fMRI, we analyzed brain activation, to verify and localize neural correlates of decision variables that are needed to implement the attention or integration model.

To study these questions we used a variant of a hierarchical reversal learning task with two stimulus dimensions. The task is a modified version of classical neuropsychological tests of hierarchical decision making, such as the Wisconsin-Card-Sorting Task [115] and its modern variants such as the Intra-Extra Dimensional Set Shift Task [116]. In these tasks, only one stimulus dimension out of several is causally linked to reinforcement in any given trial, and subjects are rewarded if they select the correct exemplar based on the currently relevant dimension. The task is rendered more difficult because the contingencies at each level of the hierarchy change frequently and subjects have to constantly relearn the correct response in order to achieve rewards.

The ability to select actions in relation to internal goals is a cardinal function of the prefrontal cortex [117]. Lesions of prefrontal cortex in both humans and other animals are known to dramatically impair performance on such hierarchical decision tasks [118-121], and imaging studies have revealed activity in a number of different regions of prefrontal cortex during performance of such tasks in healthy volunteers [122, 123]. It should be emphasized, however, that such studies have as of yet neither addressed the computational mechanisms underlying the process of solving this type of problem, nor their encoding in the brain.

Here, subjects had to choose between two compound stimuli that were presented simultaneously on the screen (Figure 4.1A). Each stimulus had two dimensions (color

and motion) and within each dimension there were two exemplars of each stimulus category (either red or green for color and leftward or rightward for motion). The exemplars for the upper stimulus were assigned pseudo-randomly in each trial and converse exemplars were assigned to the lower stimulus. A constraint ensured that identical pairings did not occur more than two times in a row. Only one dimension was relevant for determining reward at any given time and within that dimension a particular exemplar was correct. Choosing the stimulus that contained the correct exemplar resulted in a high reward probability (80%) while choosing the other stimulus resulted in a low reward probability (20%). For example, at a given time “color” may be the relevant dimension and within color, “green” may be the correct exemplar. In our illustration in Figure 4.1A, choice of the upper stimulus is correct and will lead to a reward with 80% probability. After subjects consistently chose the stimulus with the correct exemplar three times in a row (indicating that they had formed a hypothesis about the relevant dimension and correct exemplar) the correct exemplar switched in each further trial with a 50% probability. Furthermore, after a variable number of such within-dimension switches (between 2 and 4), the relevant dimension switched.

As in many real-life problems, participants in our experiment have more than one theory (“which is the relevant dimension?”) to go by in any trial, and corresponding characteristics (“which is the correct exemplar?”) on which to base their decisions between the upper or lower stimulus. It is important to note that participants never got direct feedback about whether their theory was currently correct, but only whether their choice proved successful or not.

To analyze subjects' behavior, we developed two alternative computational models that could be used to guide choice in our task, corresponding to the two generic strategies. (1) A model implementing the "gated-attention" strategy, and (2) a model based on evidence integration. The first model computes a decision in a two-step procedure by first evaluating which dimension is currently relevant before subsequently working out which exemplar within that dimensional category is currently reinforced (Figure 4.1B). It thereby allocates attention to only that dimension which is deemed relevant. The second model integrates information based on weights that reflect the likelihood that each dimension is relevant (Figure 4.1C). This model fully integrates over all of the available probabilistic information: even if, say, color is deemed highly likely to be the relevant dimension, the model not only takes into account the probability that red or green is correct, but to a lesser degree also uses the information it has from the motion dimension for which movement direction would indicate the correct choice.

We used standard reinforcement learning [5] to model subjects' learning process, combined with the gated-attention or integration model to generate model-predicted estimates for subjects' choices.

Overall, we found behavioral evidence that our subjects tend to rely more on the mathematically optimal integration strategy than on attention-gating when solving our task. The BIC corrected log-likelihood fit was better for the integration model than the gated-attention model in all subjects (Figure 4.1D). We further compared the fitted  $\beta$ -parameters for the two models in individual subjects and found that  $\beta$ -parameters of the

integration model were significantly larger than the  $\beta$ -parameters of the attention model (one-sided paired t-test,  $p < 10^{-7}$ ). The  $\beta$ -parameter is also an indicator for the behavioral fit of the model as a high  $\beta$  indicates a steep softmax decision function, accommodating behavior even when the data are best described by a nearly perfect mapping from desirability to choice. Altogether, this suggests that subjects tended to utilize information from the characteristics of both dimensions, rather than the alternative of concentrating only on the dimension most likely to be correct and then choosing accordingly.

In order to identify neural correlates of the valuation process we separately regressed neural activity onto trial-by-trial value signals for the attention-gating and integration model. Specifically, we were most interested in the full-value signal on each trial as this is the key output variable of the decision making process. In the case of the attention model, this value corresponds to the RL-value of selecting the better stimulus within the dimension that the model predicts is currently relevant. In the integration model, this value is a linear combination of RL-values for the color and motion exemplars, weighted by the model predicted across dimension likelihood. We also tested for correlations between confidence signals of the exemplars across and within each dimension and the fMRI data. The across-dimension confidence reflects how likely it is that one of the complementary dimensions is the correct one. The within-dimensions confidence measures the likelihood that one gets the exemplars for the two dimensions right.

We found BOLD correlates of decision variables for the integration model in specific sub-regions of prefrontal cortex. The full value signal correlated most strongly with activity in vmPFC extending dorsally and medially along PFC (Figure 4.2A). This region was previously found to encode expected value signals of the actions or stimuli that were

chosen on a trial [124], a signal correlated to but not identical with our full value ( $R^2=0.50$  across subjects). We also found value chosen signals in vmPFC (Figure S 4.1), though the effect size of the full value was higher than that of the value chosen signal at the peak of the activation. The within dimension certainty also correlated with activity located in mPFC (Figure 4.2B). We found a negative correlation with the within-dimension confidence in the anterior cingulate cortex and bilaterally in the frontal poles (Figure 4.2C). No significant correlation was found between the BOLD signal and our measure of confidence across dimensions at  $p<0.001$  uncorrected. A full list of all activated regions is shown in Table 4.3.

We found a similar activation pattern correlating with decision variables of the attention-gated model, though with a weaker effect size and smaller extent. Neural activity that correlated significantly with each contrast of the attention model was located at approximately the same location as neural activity correlating with the same contrast of the integration model (see supplementary materials for more details).

In particular, neural activity in medial prefrontal cortex, averaged across all subjects, correlated both with the full value of the attention and the integration decision model. The area activated by the attention gated model was thereby entirely contained within the larger area activated by the integration model at a threshold of  $p<0.001$  uncorrected (Figure 4.3A).

In order to quantitatively compare the value signals from both models within vmPFC we determined the relative probability for the models to explain the measured neural signal in every subject by means of a Bayesian model comparison approach [125] (see



supplementary methods for details). Consistent with our behavioral results, we found that overall the integrating model was more likely to be the underlying cause of the neural variability in vmPFC (Dirichlet alpha = 13.34 integration vs. 4.66 attention; posterior probability = 0.74, exceedance probability = 0.98; Figure 4.3B).

We also looked at posterior model probabilities for individual subjects in this area and compared them to a measure of the extent to which the integration model explained each individual's behavior better than the attention model. We found a significant correlation between the posterior probability for the integration model and the difference in softmax  $\beta$ -parameters between the integration and attention model ( $\beta_{\text{integration}} - \beta_{\text{attention}}$ ). This allowed us to measure inter-subject differences in the strategy that a subject employed and differentiate those subjects that are high integrators from those that lean more towards a gated attention strategy. We found a significant positive correlation ( $r=0.58$ ,  $p<0.02$ ) implying that subjects who behave closer to the integrating ideal had activity in vmPFC that also conformed closer to the value function from such a model (Figure 4.3C).

Here we investigated whether humans approach decision problems by first determining the most likely theory that explains how rewards are tied to stimuli and then choose on the basis of that theory, or by deciding on the basis of all possible explanations and weighing each by the likelihood that it is correct. Overall, the integration decision model explained choices better than the attention-gated model. Neural activity in vmPFC correlated more strongly with trial-by-trial valuations according to the integration model

than the attention-gated model. We also found neural signals correlating with key computational components, such as the confidence that one has identified the right choice for all theories simultaneously, while signals irrelevant to the integrating model, such as the confidence that one has identified the right theory, were absent.

In the integration approach, trust in one's decision is best measured by our within-dimension confidence measure, which *combines* the likelihood that one has identified the right exemplar for both color and motion dimensions. Accordingly, we discovered strong positive correlations between the within-dimension confidence measure and activation in PFC, and a negative correlation (in which case the activation reflected lack of trust, or risk) in the frontal poles. This means that the frontal poles are more active in those trials in which subjects are unsure about the correct choice, supporting the suggestion that the frontopolar cortex is associated with exploring multiple behavioral alternatives in search of optimal behavior [126]. Across-dimension confidence, the trust that one has identified the right dimension, is irrelevant in the integration approach, because both dimensions will be taken into account in the final decision anyways. It is, however, a crucial variable in the gated-attention approach, because one dimension will be chosen to guide the final decision, and hence, trust in that decision depends critically on the confidence that one is following the correct dimension.

Both our behavioral and neuroimaging data support the idea that humans are able to solve complex hierarchical decision problems in an integrated way, whereas the usual description of those problems, in terms of a hierarchy, suggests a two-step procedure. In the integration approach, even the least likely explanation of observed phenomena is taken into account in order to generate a decision and its weight is commensurate with its

likelihood. Such integration by likelihood has been formalized in Bayesian analysis and is equivalent to the issue of how to treat hyper-parameters in the machine learning literature [127]. Therefore, our evidence underscores the utility of framing human cognitive processes, whether at the perceptual level [128-130], at the level of cognition [131], or at the level of action selection [132, 133], in a Bayesian framework. Although fully Bayesian approaches have recently been promoted as a mechanism to account for some human learning processes such as causal inference [134, 135], our evidence suggests that integration of the possible explanations of observed phenomena may be the key ingredient. Indeed, notice that our decision models do not differ at the learning level: both use the same reinforcement learning principles. This ensured that gated-attention and integration approaches to decision making were compared in a clean fashion. When we fit a fully Bayesian model (which combines integration with Bayesian learning of likelihoods) to behavioral and brain data, results were marginally inferior, although the fully Bayesian model still performed notably better than the gated-attention model.

There is prior evidence indicating that prefrontal cortex plays a critical role in hierarchical decision making [116, 118-121, 136-138]. Our results indicate that at the level of inference – when working out what choice to take next – the prefrontal cortex uses probability information in an integrated fashion, and this is reflected in actual choices. It remains possible, that if humans are faced with a hierarchical problem of sufficient complexity (for example requiring integration over many more than two dimensions), keeping track of all theories simultaneously becomes both cognitively too challenging and normatively ill-advised [139]. In those instances, subjects might switch to a simpler strategy like attention gating, which employs fewer resources. Nevertheless,

our results indicate that at least for some classes of decision problem, the human brain and the prefrontal cortex in particular, is capable of integrating information in an optimal manner, similar to that of an ideal Bayesian observer.

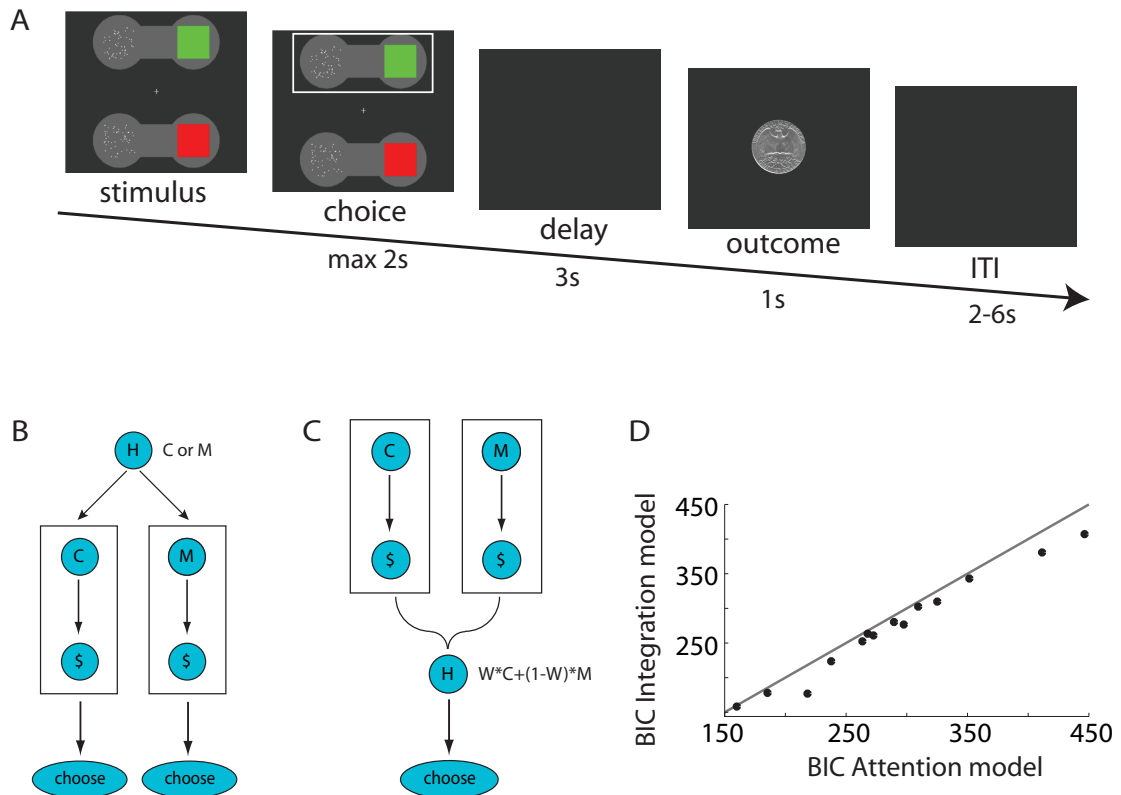


Figure 4.1 Task and Behavior

(A) Subjects choose one of two items, of which each had a color (red or green) and a motion (left or rightwards moving dots) attribute. The features were randomly assigned to both items. Once the subject selects an item, a box is placed around the target and remains on the screen until 2 s after stimulus onset. After a 3 s delay they either received a 25 cent reward or a subtraction of 25 cents from their payout. One feature is designated the correct feature, and the choice of the item carrying that feature leads to a reward on 80% of the occasions and a loss 20% of the time. Consequently, by choosing this correct item subjects accumulate monetary gain. The other item is incorrect, and choosing it leads to a reward 20% of the time and a punishment 80% of the time, leading to a cumulative monetary loss. After subjects choose the correct item on four consecutive occasions, the contingencies reversed with a probability of 50% in every consecutive trial. After two to four of such within-dimension reversals the relevant dimension changed (extra-dimensional switch). The inter-trial-interval was variable. (B)

Hierarchical decision model based on attentional-shifts. Stimulus-outcome associations are learned for color (C) and motion direction (M). Subjects first form a hypothesis (H) about which dimension is relevant (either C or M) and then base their reward expectation (\$) and choice exclusively on the information learnt about that dimension. (C) In the integration model, the available information from both dimensions (C and M) is integrated as a weighted sum and the decision is based on a linear combination of evidence from both dimensions. Subjects form a hypothesis (H) about the likelihoods that each dimension is relevant, corresponding to weights in the linear combination. Weights are updated on every trial. (D) Behavioral fit (BIC) of the two decision models. Smaller values indicate a better fit. The integration model fits better to subjects' behavior in every single subject. The variability in BIC across subjects is mainly due to the variable number of trials per subject (BIC depends on n).

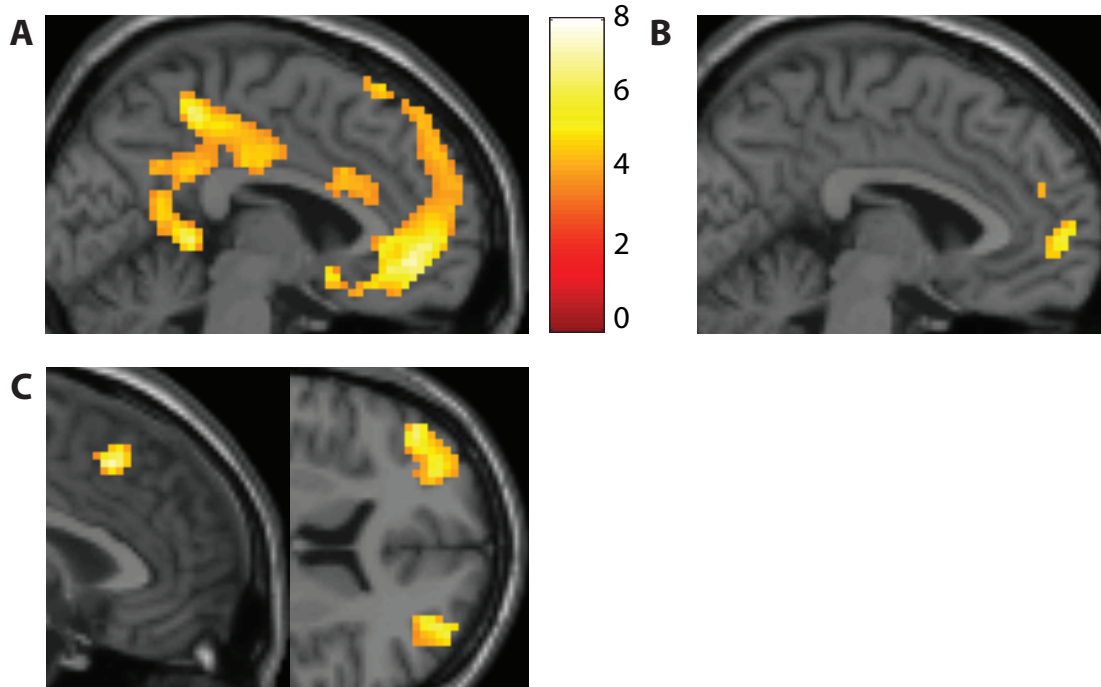


Figure 4.2 Activity reflecting the Integration model

(A) BOLD responses in mPFC correlate significantly with the full-value signal from the integration model. (B) BOLD responses in a sub-cluster of medial PFC correlate significantly with the within-dimension certainty (averaged across color and motion) from the integration model. (C) The frontal poles and anterior cingulate show negative correlations with the within-dimension certainty (thus indicating a positive correlation with risk). All data are shown at a height threshold of  $p < 0.001$  and corrected for multiple comparisons at the cluster level (AlphaSim extent threshold  $p < 0.001$ ).

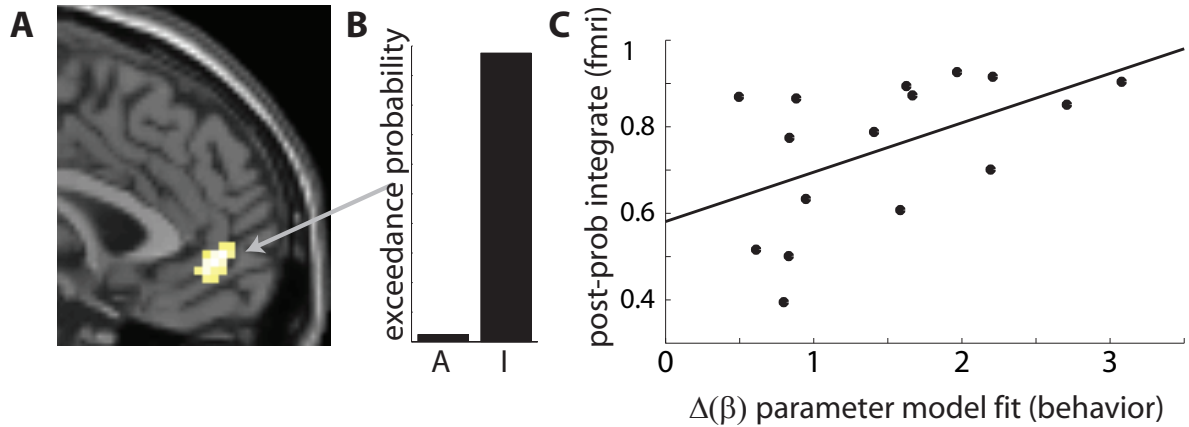


Figure 4.3 Bayesian model comparison

(A) Neural activity in vmPFC correlates with the trial-by-trial full value signal. Shown is the area that is commonly activated by both the integration and the attention-gated model at  $p < 0.001$ . (B) We used a Bayesian model comparison to identify the model that can better explain neural activity in this area. Overall, the integration model explains activity in this vmPFC area better than the attention-gated model (exceedance probability for the integration model is 0.98). The exceedance probability is the probability that one model is more likely than the other one, i.e., that the posterior probability for the integration model is larger than 0.5. (C) Within vmPFC, posterior model probability of individual subjects correlate with the softmax beta parameter difference between the integration and attention-gated model ( $r = 0.58$ ,  $p = 0.02$ ).



## Supplementary Materials

### *Optimal decision making*

The optimal decision strategy for an ideal observer in our task would be to follow a generative Bayesian model, which has the following properties:

A certain stimulus  $stim_t$  is always considered to be the correct one, as determined by the following algorithm:

1. Draw from a binomial distribution (with parameter  $\phi$ ), whether to change modality,  $modal_t$ .
2. Given the modality, draw from a binomial distribution (with parameter  $\varphi$ ), whether to change stimulus,  $stim_t$ .
3. Once a subject makes a choice, if the choice includes  $stim_t$ , reward with 80 percent probability, otherwise by 20 percent.

The goal for an optimal observer is to maximize his discounted expected utility at each round [114], but as each round can be considered independent the observer merely needs to maximize the expected utility in the current round.

In order to do this the subject needs to calculate which of the two options (*UP* or *DOWN*) is most likely to generate a payout, i.e., whether the relevant characteristic is most likely to be one of the two stimuli in the *UP* option or the two in the *DOWN* option.

The probability of *UP* having the relevant characteristic ( $stim_t$ ) is

$$P(UP \text{ is } rel) = P(UP_{stim1} = stim_t | modal_t = colour) * P(modal_t = colour) + P(UP_{stim2} = stim_t | modal_t = motion) * P(modal_t = motion)$$

Notice, it is here optimal to take a weighted average of the two stimuli (marginalize over  $P(modal_t)$ ). That is, probabilities of the stimuli are *integrated* across the two modalities. A different (suboptimal) approach would be to first determine which modality is likely to be correct, and then determine within that modality which stimuli is more likely (i.e. take the maximum over  $P(modal_t)$ ). In this alternative approach, attention is *gated* to the most likely modality.

Given the learned probabilities,  $P(modal_t)$ ,  $P(stim_1 | colour)$  and  $P(stim_2 | motion)$  each of these two approaches lead to a prediction about which outcome should be chosen. The actual learning of the probabilities can be done either using standard Bayesian inference or through reinforcement learning (used in this study and described in the Methods).

#### *Neural activity correlating with the attention-gated model*

After lowering our threshold to 0.001 uncorrected, we found BOLD correlates of decision variables for the attention-gated model in the same areas as those of the corresponding signals of the integration model. As anticipated by the inferior behavioral fit of the attention-gated model, the effect size of attention-gated modulated activity was also smaller than that of the corresponding contrast of the integration model (see the main text for a quantitative Bayesian model comparison in this region).

The full value signal correlated with activity in vmPFC, extending dorsally and medially along PFC (Figure S 4.2A). The within dimension certainty also correlated with activity

located in mPFC (Figure S 4.2B) and negatively with the frontal poles bilaterally (Figure S 4.2C). Interestingly, we found an additional area in ventral-medial OFC correlating with the across dimension certainty of the attention model. Across-dimension confidence, the trust that one has identified the right dimension, is irrelevant in the integration approach because both dimensions will be taken into account in the final decision anyways. It is, however, a crucial variable in the gated-attention approach, because one dimension will be chosen to guide the final decision, and hence, trust in that decision depends critically on the confidence that one is following the correct dimension. Though we identified neural activity correlating with the across dimension certainty of the attention-gated model only at  $p < 0.001$  uncorrected, we did not find any such correlates with the integration model at the same threshold.

We ran a Bayesian model comparison on the across dimension effect, similar to the model comparison of the full value described in the main text. Not surprisingly, activity in this area can be explained exclusively by the attention model (exceedance probability 0.99, Figure S 4.3A/B). We observed a negative correlation ( $r = -0.47$ ,  $p = 0.055$ ) of individual subjects' posterior probability in this area and the beta parameter difference (Figure S 4.3C). Interestingly, those subjects who scored high on integration posterior-probability in vmPFC (Figure 4.3C), did indeed score low on attention posterior-probability in OFC and vice versa ( $r = -.74$ ,  $p = 0.001$ ). These results suggest that, though subjects tend to rely in general in our study more on the integration approach, inter-subject differences exist in the employed strategies and some people are better integrators than others.

## Methods

### *Subjects*

16 healthy subjects (6 female; 18–28 years old; right-handed) with no history of neurological or psychiatric illness participated in the study. The study was approved by the Institutional Review Board of the California Institute of Technology.

### *Experimental design and task*

The task is a variant of an intra/extra-dimensional (ID/ED) shift task. In every trial, subjects chose between two stimuli that were presented simultaneously on the screen. Each stimulus had two dimensions (color and motion) and within each dimension there were two exemplars of each stimulus category (either red or green for color and leftward or rightward for motion). The exemplars for the upper stimulus were assigned pseudo-randomly in each trial and converse exemplars were assigned to the lower stimulus. We implemented a constraint that identical pairings did not occur more than two times in a row to avoid trials in which subjects cannot associate the outcome unambiguously to a chosen motion or color exemplar. At any given time, one dimension was “relevant”, and within that dimension a particular exemplar was correct. For example, “color” may be the relevant dimension and within color, “green” may be correct. Choice of the stimulus that has the correct exemplar will yield monetary rewards on a probabilistic basis with 80% probability, whereas selection of the other stimulus will yield reward with only 20% probability. After subjects chose the stimulus with the correct exemplar three times in a row (indicating that they learned the relevant dimension and correct exemplar) the correct

exemplar will switch with a 50% probability in each further trial. Furthermore, after a variable number of such within-category switches (2, 3, or 4; rectangular distribution), the relevant dimension is also switched. This design imposes a hierarchical structure on the task with exemplar reversals occurring on a faster timescale than the dimensional switches. The total number of trials varied across subjects due to the above described rules; however one experiment always contained 9 dimensional switches (5 motion & 5 color blocks). Rewarded trials yielded a prize of 25 cents, while unrewarded trials resulted in a loss of 25 cents. At the end of the experiment subjects were paid their accumulated earnings in addition to a flat amount of \$25.

The task was presented via back projection on a translucent screen, viewable through a head-coil mounted mirror. Subjects chose the upper or lower stimulus by pressing one of two distinct buttons on a button box with their right thumb. Eye positions were monitored at 120 Hz with a long-range infrared eye-tracking device (ASL Model L6 with control unit ASL 6000, Applied Science Laboratories, Bedford, MA).

### *Reinforcement Learning (RL) model*

For modeling the data we assumed that subjects learned the relevant values assigned to the two actions on the basis of trial-by-trial experience using a simplified version of Q-learning, the Rescorla-Wagner rule [5]. Our implementation of this rule can be seen as an approximation to the full Bayesian approach (described below) and generates choices and values highly correlated with the Bayesian model in this task.

If action  $a$  is selected on trial  $t$ , its value is updated via a prediction error,  $\delta$ , as follows:  $V_a(t+1) = V_a(t) + \eta\delta(t)$ , where  $\eta$  is a learning rate between 0 and 1. The prediction error  $\delta(t)$  is calculated by comparing the actual reward received,  $r(t)$ , with the reward that the subject expected to receive from that action in that trial; that is,  $\delta(t) = r(t) - V_a(t)$ . Specifically for this task, three variables had to be learned and updated in each round:  $V_{green}$ ,  $V_{right}$  and  $V_{colour}$ , keeping track, respectively, of the value of the green stimulus (versus red stimulus), rightwards motion stimulus (versus left stimulus) and the color vs motion modality. We assume that  $V_{red} = -V_{green}$ ,  $V_{left} = -V_{right}$ , and  $V_{motion} = -V_{colour}$ . Hence, the updating of a value can be done with the inverse prediction error ( $-\delta(t)$ ) if the complementary action is chosen.

In each round, a subject chooses either UP or DOWN, and thus selects a combination of one color (red or green) and one motion (left or right). The values of these ‘intermodality’ choices are updated in each round using the following standard RL scheme:

$$\delta_i(t) = r_i - V_i(t), \text{ where } i = \text{red (green) if red (green) was chosen.}$$

$$\delta_j(t) = r_j - V_j(t), \text{ where } j = \text{right (left) if right (left) was chosen.}$$

The extramodality value is updated according to:

$V_c(t+1) = V_c(t) + \gamma\delta_c(t)$ , with  $\delta_c(t) = r_i(V_i - V_j) - V_c(t)$ , where  $\gamma$  is the learning rate for the extramodal value. The extramodal value is increased if the difference in expected value for each of the two modalities was larger than the expected extramodality value. In terms

of the typical Rescorla-Wagner model the extramodal signal is tracking the ability of one modality to predict reward, relative to the other.

Using these learned values we compared two ways to generate the choices:

1. *Gated-attention*: Determine which modality has the highest expected value,  $V_C$  or  $V_M = -V_C$ . For that modality choose which of the two stimuli has the highest value  $V_i$ . Given choice between  $UP = \{V_{i1}, V_{j1}\}$  and  $DOWN = \{V_{i2}, V_{j2}\}$  calculate  $V_{UP} = V_{k1}$  and  $V_{DOWN} = V_{k2}$ , where  $k = \text{argmax}(V_C, V_M)$ . Or  $V_{UP} = V_{i1} * H(V_C) + V_{j1} * H(-V_C)$  where  $H$  is the Heaviside operator,  $H(+)=1$ ,  $H(-)=0$ . Choose action  $a = \text{argmax}(V_{UP}, V_{DOWN})$ .
2. *Integration*: Calculate (similar to the marginalizing example for Bayesian model) the value for UP as  $V_{UP} = V_{i1} * P(\text{color}) + V_{j1} * P(\text{motion})$  where  $P(\text{colour}) = P_o^k / (P_o^k + (1 - P_o)^k)$ ,  $P(\text{motion}) = 1 - P(\text{color})$  and  $P_o = (V_C + 1)/2$ . Choose action  $a = \text{argmax}(V_{UP}, V_{DOWN})$ .

We used a soft-max procedure to generate choices, where in every trial, the probability (P) of choosing action  $a \in \{UP, DOWN\}$  over  $b$  is given by:  $P_{a,t} = \sigma(\beta(V_a(t) - V_b(t)))$ , where  $\sigma(z) = 1/(1 + e^{-z})$  is the Luce choice rule or logistic sigmoid, and  $\beta$  determines the degree of stochasticity involved in making decisions. We fit the parameters (learning

rates  $\eta$ ,  $\gamma$ , soft-max  $\beta$ , and  $\kappa$  for the integrating model) such that the model best explained subjects' choices (maximum likelihood).

### *Behavioral model comparison*

Given the learned variables, we have specified two ways to generate choices. One is based on focusing on the most likely modality, the other on optimally integrating the evidence for each modality. To compare the values for models with higher complexity we report the Bayesian Information Criterion (BIC) [140] in Table S1, which corrects for the number of parameters,  $k$ , in a model based on the number of data points  $n$ :  $BIC = -2 \cdot \log(P_s) + k \cdot \log(n)$ . The model with the lower  $BIC$  explains the subjects' behavior better. Comparing the two RL models we find that the integrating RL model performs better for all subjects.

The attention-gated model can be seen as a special case of the integrated model (i.e. a nested model). This happens when the transformation from value to probability approximates a Heaviside function (when  $\kappa \rightarrow \infty$ ), as implemented in the attention-gated model. The fitted parameter  $\kappa$  is shown in Figure S 4.4 for all subjects. Notice that the function is very far from being a Heaviside for most subjects. Although the extra parameter gives more flexibility in fitting for the integrating model, the BIC correction takes this into account.



Alternatively we can compare the models through cross validation, by which the data is split in two halves, the model parameters are fitted on the first half (training data), and the models are then compared on the second half of the data (test data). This method confirms the above analysis in that the integrating model performs better on the log-likelihood of the test data-set (see Table 4.2).

### *Other alternative models*

In the analysis above we assumed that subjects base their decisions on the values learned through a two-layered RL model. RL models had been shown to closely mimic the type of behavior elicited by human subjects for a large number of similar learning tasks [38, 48]. In many of these tasks the RL models can be seen as approximations to the optimally Bayesian model ([114, 141]), which are often too computationally cumbersome to actually implement. However, for completeness we tested whether subjects would better follow the behavior predicted by a fully optimal Bayesian model. We implemented such a model (for details see section Bayesian model below) and compared the BIC values with those of our RL models. The fully Bayesian model did not describe subjects' performance as well as the integrating RL model (BIC values of all models are shown in Table 4.1).

On the other hand, we also tested whether the two-layer RL models would perhaps be too complex for our subjects to follow. We created two simpler one-layer variants of the RL model: (1) a one-layer version of the model above where  $V_C$  is always kept at 0 and hence there is never any information with regard to which modality is more likely to be correct,

as well as (2) a 4-option model where each of the 4 options (red, green, left, right) are treated as a separate options for the RL model to learn about (i.e.,  $V_{red}$  and  $V_{green}$  are not assumed anti-correlated) and furthermore  $V_C$  is also kept at 0. When comparing the likelihood of the subject responses we again find that the Integrating RL model does much better at describing subjects' performance, even after taking into account the lower complexity of the 1-layer models (2 instead of 4 parameters) (Table 4.1).

### *Bayesian model*

We will refer to the dimension as  $D=\{\text{color,motion}\}$ , color as  $C=\{\text{red,green}\}$ , and motion as  $M=\{\text{left,right}\}$ . Assume that  $UP=\{\text{green, left}\}$  and  $DOWN=\{\text{red,right}\}$ . Our goal is to be able to calculate the probability of the monetary reward being given for the Up versus DOWN option, given the previous reward history  $rew_{1:t}$ :

$$P(UP_t | rew_{1:t}) = P(C_t = \text{green} | D_t = \text{color}, rew_{1:t})P(M_t = \text{left} | D_t = \text{motion}, rew_{1:t}) + P(M_t = \text{left} | D_t = \text{color}, rew_{1:t})P(D_t = \text{motion} | rew_{1:t})$$

Hence we need to know  $P(C_t = \text{green} | D_t = \text{color}, rew_t)$ ,  $P(M_t = \text{left} | D_t = \text{motion}, rew_t)$ , and  $P(D_t = \text{color} | rew_t)$ .

After a reward at time  $t+1$ , the updating of each of these is done according to Bayes rule:

$$P(C_{t+1} | D_{t+1} = \text{color}, rew_{1:t+1}) = P(rew_{t+1} | C_{t+1}, D_{t+1} = \text{color})P(C_{t+1} | D_{t+1} = \text{color}, rew_{1:t}) / Z \quad \text{where } Z$$

stands for normalization and where

$$P(C_{t+1} | D_{t+1} = \text{color}, rew_{1:t}) = \sum_{D_t} \sum_{C_t} P(C_{t+1}, C_t, D_t | D_{t+1} = \text{color}, rew_{1:t})$$

$$\begin{aligned}
&= P(D_t | rew_{1:t}) \sum_{C_t} P(C_{t+1} | C_t, D_t = color, D_{t+1} = color) P(C_t | D_t, rew_{1:t}) + \frac{1}{2}(1 - P(D_t | rew_{1:t})) \\
&= [\varphi P(C_t \neq C_{t+1} | D_t, rew_t) + (1 - \varphi) P(C_t = C_{t+1} | D_t, rew_{1:t})] P(D_t | rew_{1:t}) + \frac{1}{2}(1 - P(D_t | rew_{1:t})).
\end{aligned}$$

$P(rew_{t+1} | C_{t+1}, D_{t+1} = color) = P(rew_{t+1} | M_{t+1}, D_{t+1} = motion)$  is 0.8 for a positive reward, 0.2 for a punishment, and  $\varphi$  and  $\phi$  represent the probability of a switch happening within, and across modality, respectively.

Hence the posterior for round t+1 can be expressed in terms of the posteriors from the previous round t. Similarly for  $P(M_{t+1} | D_{t+1} = motion, rew_{1:t+1})$ .

For across dimensions:

$$P(D_{t+1} | rew_{1:t+1}) = P(rew_{t+1} | D_{t+1}) P(D_{t+1} | rew_{1:t}) / Z, \text{ where}$$

$$P(rew_{t+1} | D_{t+1}) = \sum_{C_t} P(rew_t | C_{t+1}) P(C_{t+1} | D_t) \text{ if } D_t = color$$

$$P(rew_{t+1} | D_{t+1}) = \sum_{M_t} P(rew_t | M_{t+1}) P(M_{t+1} | D_t) \text{ if } D_t = motion$$

$$\text{and } P(D_{t+1} | rew_t) = \sum_{D_{t+1}} P(D_{t+1} | D_t) P(D_t | rew_t) = (1 - \phi) P(D_t = D_{t+1} | rew_t) + \phi P(D_t \neq D_{t+1} | rew_t).$$

$$\text{Hence, } P(D_{t+1} | rew_{1:t+1}) = P(rew_{t+1} | D_{t+1}) [(1 - \phi) P(D_t = D_{t+1} | rew_t) + \phi P(D_t \neq D_{t+1} | rew_t)] / Z.$$

*FMRI data acquisition*

Data were acquired with a 3T scanner (Trio, Siemens, Erlangen, Germany) using an eight-channel phased array head coil. Functional images were taken with a gradient echo T2\*-weighted echo-planar sequence (TR = 2.65 s, flip angle = 90°, TE = 30 ms, 64 × 64 matrix). Whole brain coverage was achieved by taking 45 slices (3 mm thickness, no gap, in-plane resolution 3 × 3 mm), tilted in an oblique orientation at 30 deg to the AC-PC line to minimize signal dropout in OFC. Subject's head was restrained with foam pads to limit head movement during acquisition. A high-resolution T1-weighted anatomical scan of the whole brain (MPRAGE sequence, 1x1x1 mm resolution) was also acquired for each subject.

*FMRI data analysis*

Image analysis was performed using SPM5 (Wellcome Trust Centre for Neuroimaging, Institute of Neurology, London, U.K.). Images were first slice time corrected to TR/2, realigned to the first volume to correct for subject motion, spatially normalized to a standard T2\* template with a voxel size of 3 mm, and spatially smoothed with a Gaussian kernel of 8 mm FWHM. Intensity normalization and high pass temporal filtering (using a filter width of 128 s) were also applied to the data.

First, we estimated for each individual subject a GLM for the attention-gated model and separately another GLM for the integration model, differing only in the model-predicted parametric modulator values. Two events were modeled in each trial: the time of the

stimulus presentation, parametrically modulated by four variables  $M_x$ , and the time of the presentation of the outcome, modulated by the binary outcome (+1/-1).

The four parametric modulators contained values from the respective model (attention / integration):

M1: Full Integrated value of the model ( $\text{argmax}(V_{UP}, V_{DOWN})$ )

M2: Intradimensional confidence for color ( $\text{argmax}(V_{red}, V_{green})$ )

M3: Intradimensional confidence for motion ( $\text{argmax}(V_{right}, V_{left})$ )

M4: Extradimensional confidence ( $\text{argmax}(P_{MOTION}, P_{COLOR})$ )

GLMs were with AR(1). All regressors were convolved with the canonical hemodynamic response function. In addition, the 6 scan-to-scan motion parameters produced during realignment and a session constant were included as additional regressors of no interest.

Second, we computed contrasts of interest at the individual level using the regressors described above. The within-dimension confidence contrast shown in Figure 4.2B and Figure 4.2C are an equally weighted linear combination of within-dimension confidence regressors M2 and M3. We also looked at correlations with M2 and M3 separately and found that areas activated by M2 and M3 overlap and are both exclusively located at the same region as the combined contrast shown in Figure 4.2.

To enable inference at the group level, we calculated second-level group contrasts using a one-sample t-test. The structural T1 images were co-registered to the mean functional EPI images for each subject and normalized using the parameters derived from the EPI images. Anatomical localization was carried out by overlaying the t-maps on a normalized structural image averaged across subjects, and with reference to an anatomical atlas [97].

The activated voxels of all reported results are statistically significant at a threshold of  $p < 0.001$ , corrected for multiple comparisons, as stipulated by Monte Carlo simulations (AlphaSim within AFNI)[112]. AlphaSim generates an estimate of overall cluster size significance level by iteration of the process of random image generation, Gaussian filtering to simulate voxel correlation, thresholding, image masking, and tabulation of cluster size frequencies. In our simulation we generated a series of 10,000 random images, each having  $N$  (number of voxels in masked epi images) spatially uncorrelated voxels by filling the masked brain volume with independent normal random numbers. The effect of voxel correlation was simulated by convolving the random image with a Gaussian function of the size of our smoothing kernel (8 mm FWHM). The image was then scaled to provide our individual voxel probability threshold  $p_{thr} = 0.001$  by determining the value  $z_{thr}$  such that approximately  $p_{thr} * N$  voxels have intensity greater than  $z_{thr}$ . The thresholding was then accomplished by setting those voxels with intensity greater  $z_{thr}$  to 1 (activated voxels), voxels with intensity less than  $z_{thr}$  to 0. Finally, AlphaSim determined which activated voxels belong to clusters. Once all clusters had been found, the size of each cluster in voxels was recorded in a frequency table. This

simulation estimated that in a 3D volume (entire brain as masked by the real epi-images) a cluster size of  $>38$  contiguous activated voxels would occur by chance with a probability of less than 0.001.

We used a Bayesian model comparison [125] to determine which model (GLM\_attention or GLM\_integration) better explained the neural activity in vmPFC. First, we extracted and averaged effect sizes of the full value regressor (beta values) from within a 12mm sphere (1.5x smoothing kernel size) in vmPFC. Since the attention model activated cluster was completely contained by the integration model activated area, we centered the sphere on the group peak of the weaker attention-gated model (any selection bias towards the stronger correlating model would then be working against us). Next, we calculated posterior model probabilities in this region for every subject and the group of subjects. In brief, the procedure by Stephan et al. rests on treating the model as a random variable and estimating the parameters of a Dirichlet distribution, which describes the probabilities for all models considered. These probabilities then define a multinomial distribution over model space, allowing one to compute how likely it is that a model generated the subjects' data. To decide which model is more likely, we use the conditional model probabilities to quantify an exceedance probability, i.e. a belief that a particular model is more likely than the other model, given the group data.

Table 4.1 BIC model fit

Model comparison with BIC corrected model fit

Subject	RL Attention	RL Integrate	Bayes	RL 1-layer	RL 4-options
01:	237.73	223.63	235.61	261.30	349.25
02:	272.61	260.82	275.77	285.10	409.36
03:	446.46	407.10	395.42	460.37	557.49
04:	350.66	345.12	335.02	350.85	381.59
05:	267.80	263.36	270.01	267.77	436.69
06:	289.44	280.19	276.16	298.83	346.96
07:	299.04	276.91	331.02	339.86	408.95
08:	411.40	380.72	380.00	417.91	556.94
09:	263.44	252.34	299.14	302.12	413.76
10:	351.56	343.00	337.25	343.41	374.46
11:	309.42	302.49	303.31	314.34	375.98
12:	297.55	276.77	279.34	309.81	361.27
13:	325.14	309.76	308.10	323.34	430.55
14:	185.31	178.00	186.34	189.28	272.15
15:	218.25	176.91	224.69	262.44	389.75
16:	159.97	158.40	157.46	164.09	226.97
<b>Mean</b>	<b>292.87</b>	<b>277.23</b>	<b>287.16</b>	<b>305.68</b>	<b>393.26</b>



Table 4.2 Model cross-validation

Model comparison with cross-validation, given in log-likelihoods. The smaller the number the better the fit.

<b>Subject</b>	<b>RL Attention</b>	<b>RL Integrate</b>
01:	69.41	52.32
02:	74.55	68.26
03:	109.75	99.30
04:	90.23	85.82
05:	67.96	64.92
06:	67.40	66.33
07:	86.28	82.28
08:	95.18	88.68
09:	75.10	69.14
10:	81.65	80.95
11:	80.11	78.39
12:	76.14	68.26
13:	80.86	71.31
14:	46.49	39.77
15:	45.65	31.47
16:	46.56	46.15
<b>Mean</b>	<b>74.58</b>	<b>68.33</b>

Table 4.3 Activated regions

Locations of significant correlation with value signals of the integration model (threshold  $p < 0.001$  corrected for multiple comparisons). MNI coordinates denote the group peak voxel of each cluster.

*Full-value signal:*

	<i>x</i>	<i>y</i>	<i>z</i>	<i>Z</i>	# voxels	
01:	0	48	-3	4.71	1069	<i>Medial PFC</i>
02:	0	-39	54	4.50	1133	<i>Posterior CG</i>
03:	-54	-63	-6	4.44	92	<i>Left inf. temporal sulcus</i>
04:	45	-3	-18	4.24	35	<i>Right circ insular sulcus</i>
05:	42	-78	27	4.14	165	<i>Right angular gyrus</i>
06:	-48	-66	30	4.02	111	<i>Left angular gyrus</i>
07:	-60	-12	-27	3.89	51	<i>Left inf temporal sulcus</i>
08:	63	-54	-3	3.83	92	<i>Right inf temporal sulcus</i>
09:	24	-45	-15	3.79	87	<i>Hippocampus</i>

*Intra-certainty:*

	<i>x</i>	<i>y</i>	<i>z</i>	<i>Z</i>	# voxels	
01:	-3	57	0	3.84	101	<i>Medial PFC</i>

*Neg. intra-certainty:*

	<i>x</i>	<i>y</i>	<i>z</i>	<i>Z</i>	# voxels	
01:	51	15	45	5.30	537	<i>Right inf frontal sulcus</i>
02:	33	51	15	4.65	213	<i>Right frontal pole</i>
03:	3	21	48	4.64	43	<i>Anterior cingulated cortex</i>
04:	48	-48	51	4.50	340	<i>Right Intraparietal sulcus</i>
05:	-42	45	12	4.41	169	<i>Left frontal pole</i>
06:	-48	24	36	4.10	131	<i>Left inf frontal sulcus</i>
07:	-42	-51	54	3.86	118	<i>Left Intraparietal sulcus</i>

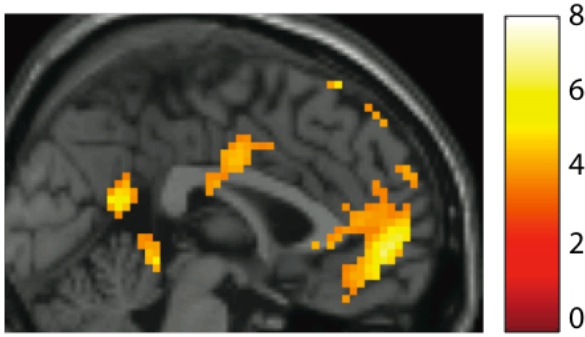


Figure S 4.1 Value chosen of the integration model

We estimated an additional GLM for the integration model in which we replaced the full value (Figure 4.2A) regressor with a value chosen regressor. Shown are areas that correlate with the value chosen in a trial (threshold  $p < 0.001$  corrected). Value chosen and full value are highly correlated and the activated areas are similar (the full value contrast has a larger extent in mPFC).

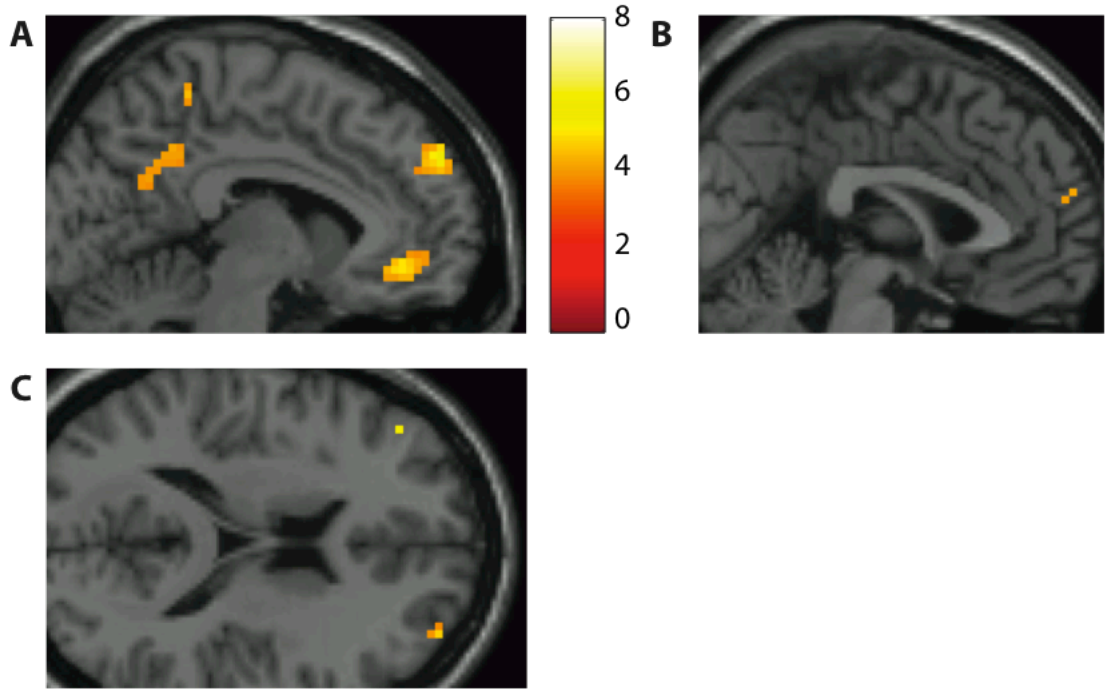


Figure S 4.2 Activity modulated by the attention-gated model

(A) Neural activity in mPFC correlate with the full-value signal of the integration model. (B) Neural activity in a sub-cluster of medial PFC correlate with the within-dimension certainty (averaged across color and motion). (C) The frontal poles and anterior cingulate correlate negatively with the within-dimension certainty. All data are shown at a threshold of  $p < 0.001$  uncorrected.

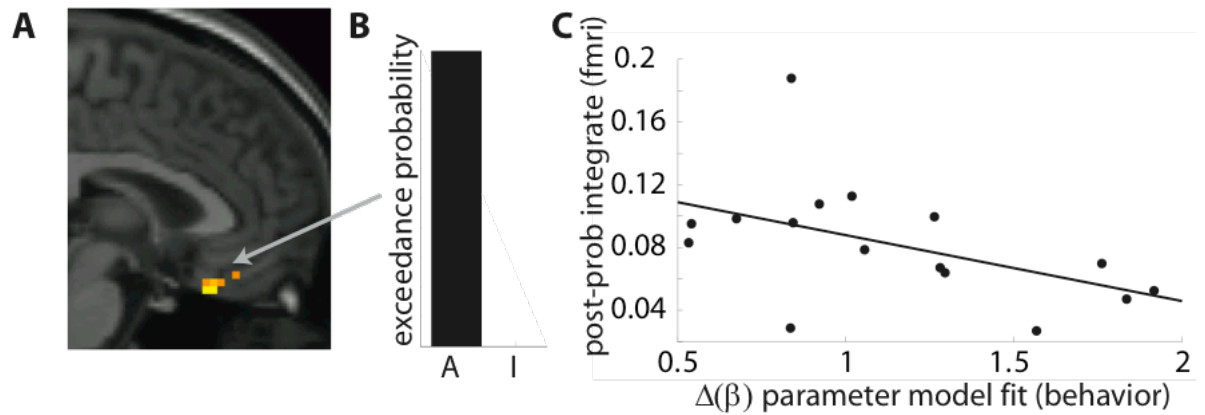


Figure S 4.3 Across-dimension certainty signal of the attention model

(A) Activity in ventral-medial OFC correlates with the trial-by-trial across-dimension certainty signal. (B) Activity in this area is exclusively explained by the attention-gated model (as compared to the integration model), shown by an exceedance probability of 0.99 in the subject group. (C) Posterior model probability of individual subjects correlate negatively with the softmax beta parameter difference between the integration and attention-gated model ( $r = 0.47$ ,  $p=0.055$ ).

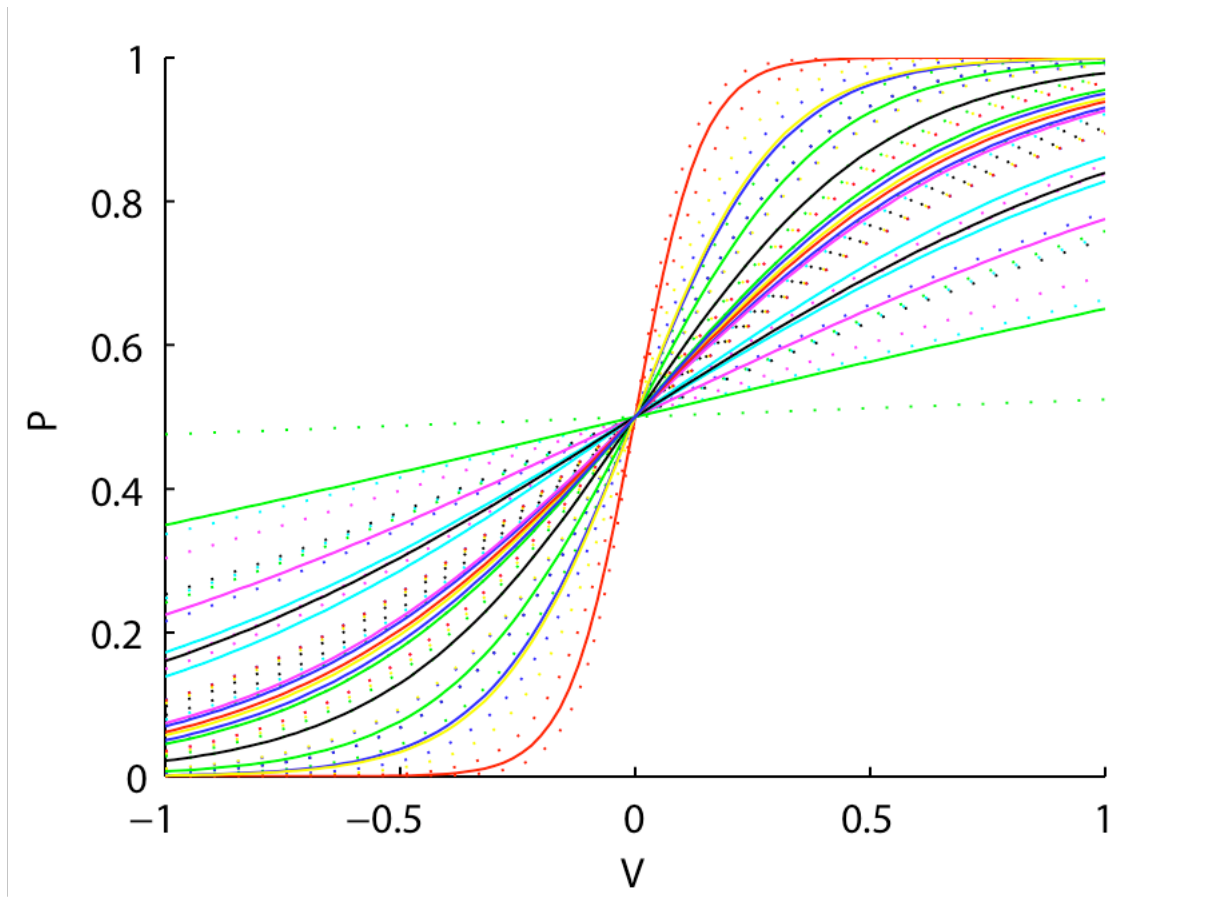


Figure S 4.4 Value – probability transformation

The transformation from value  $V_c$  to probability  $P_c$ , as found by fitting the hybrid function. Solid lines indicate the fitted function shape with dotted lines indicating error bars as found through a Laplace approximation [142].

## Chapter 5. Summary

### *Action Decisions*

In our first study we addressed two very fundamental issues in the neural basis of decision making that have remained largely unaddressed to date.

(1) While there is considerable consensus that most decisions between actions are computed by taking into account the expected future reward (or utility) of the available actions, typically neural signals that have been found in the brain correspond to the value of the action that is ultimately chosen. These signals are therefore post-choice, and by definition cannot contribute to the choice process itself. The key signals necessary for guiding subsequent choice are so-called action-values, signals that code separately for the value of each available action in the choice set, independently of what action is ultimately taken. Although there is preliminary evidence of action-value like signals in the striatum in monkeys, such signals have never been found in humans, nor reported at all in the cortex. A problem with trying to find such signals in humans using techniques such as fMRI is that limits in the spatial resolution of this technique preclude detecting such signals if they are mediated by separate but spatially intermingled neuronal populations. Here we use a novel experimental procedure to allow us to detect action values in spite of this problem: instead of having subjects make choices between actions in the same modality (such as between finger or hand movements, or between left or right eye movements, as is typically done), here subjects are making choices between different physical action modalities. That is, they choose to either make a hand movement or an eye movement. Using this method, we have uncovered evidence of separate spatially

distinct action-value signals for the two movements in a region of supplementary motor cortex. Thus, we show for the first time the existence of signals in the human brain that are the likely precursors of choice, rather than merely representing the consequences of choice as has been shown before in human imaging studies.

(2) An even more critical question that has not been addressed at all in the area of value-based decision making is: how are action-values for different actions ultimately compared against each other in order to decide what action to take, and where in the brain does this comparison process take place? This is the core of the decision process, but as of yet direct evidence for such a mechanism in the brain has proved elusive. Here we take a simple computational model of how such a process might take place (by mutual inhibitory competition), and test for brain regions exhibiting neural responses consistent with such a model. Using this approach we provide evidence to implicate the anterior cingulate cortex in mediating this function. Intriguingly, this area appears to mediate the decision process via an inhibitory rather than an excitatory mechanism. We further show that our decision model signal captures activity in this region better than does a decision conflict signal per se, ruling out that potential alternative explanation for our data. Thus, we have identified for the first time a possible locus for the actual decision-making process itself and also advanced a putative computational model for this function.

### *Economic choices*

Another fundamental debate in the neuroscience of decision-making is whether all decisions are computed by making comparisons between the physical actions required to obtain particular outcomes, such as by choosing between the hand movements necessary



to obtain different stimuli, or whether they can be made over abstract stimuli associated with those goals. In other words, if faced with a decision to go for dinner, does the brain make a choice between the available options, such as sushi or pizza, and then implement the motor output necessary to obtain the chosen stimulus? Or does it instead compare directly the physical actions needed to obtain the food (i.e., drive to the next town to one's favorite Sushi bar vs. walk down the road to the neighborhood Italian)? While the latter way of making decisions may seem strange and convoluted, in fact this is the predominant view among many decision neuroscientists who have found value signals in areas of the brain known to be involved in representing and planning movements such as lateral parietal and pre-motor cortices, and have therefore concluded that decisions are computed by comparing between actions not goals or stimuli.

We addressed this question in our second study by dissociating the process of making a decision over stimuli from the selection of the actions that are needed to implement the decision. Human volunteers were scanned with fMRI while they were presented with pairs of stimuli associated with different levels of reward. The key manipulation is that in half of the trials the stimuli were initially paired with the actions required to achieve them (as is the case in most existing experiments), but in the other half of the trials the concomitant actions were not shown until later in the trial, which precluded the subjects from initially making choices by comparing motor plans. In spite of this dissociation between stimulus presentation and action choice, we found neural signals at the time of stimulus presentation pertaining to the choice that was ultimately taken, providing evidence that the decision itself was taken at the point of stimulus presentation rather than at the point of action presentation.

It is important to note that while previous studies have shown evidence for stimulus-related values in the brain, no human or animal study to date has shown that decisions can be computed before the actions associated with the options are known. Thus, our findings indicate for the first time that at least some types of decisions are made entirely in stimulus space and not in action space.

The separate representation of action and stimulus values in different areas of the brain and the finding that the brain can compute stimulus decisions based only on stimulus values suggest that two separate brain systems are involved in making decisions among actions and stimuli. This resonates with previous findings that lesions in different parts of frontal cortex specifically impair learning about stimuli and actions.[64, 83]

### *Optimal integration*

In the real world, causal relationships between multiple possible predictors and outcomes are often not at all obvious. In our third study, we addressed the question of how the human brain tests and evaluates hypotheses about the underlying causal state of events in the world. One solution to deal with this problem is to first choose the most likely explanation and use this hypothesis until enough evidence accumulates to reject the theory as invalid. This approach has been the core of scientific deductive reasoning and formalized in hypothesis testing of classical statistics. However, an alternative problem solving strategy is to simultaneously consider multiple theories and use a weighting of all the possible causes. No theory, even the least likely, is thereby ever excluded. Such an integrative approach is the mathematically optimal strategy and typifies Bayesian analysis.

The question we asked in this study is how do humans actually solve causal inference problems – do they use an optimal integration approach (as would a full Bayesian), or instead do they focus only on the most likely solution to the exclusion of all others (an attentionally focused account). Current notions based on neuropsychology studies of human patients with prefrontal lesions support the idea that humans are attentionally focused hypothesis testers – that is they choose one possible cause and focus all of their attention on that until an alternative cause beckons. Using a combination of computational modeling with fMRI and behavioral data we show that in contrast to existing notions, human prefrontal cortex in fact acts as an optimal integrator – that is every possible cause is taken into account weighted by its likelihood and this cumulative information is then used to choose optimally. This shows for the first time that the human brain and prefrontal cortex in particular uses a mathematically optimal strategy for driving decisions rather than the less computationally intensive but sub-optimal attentionally gated approach.

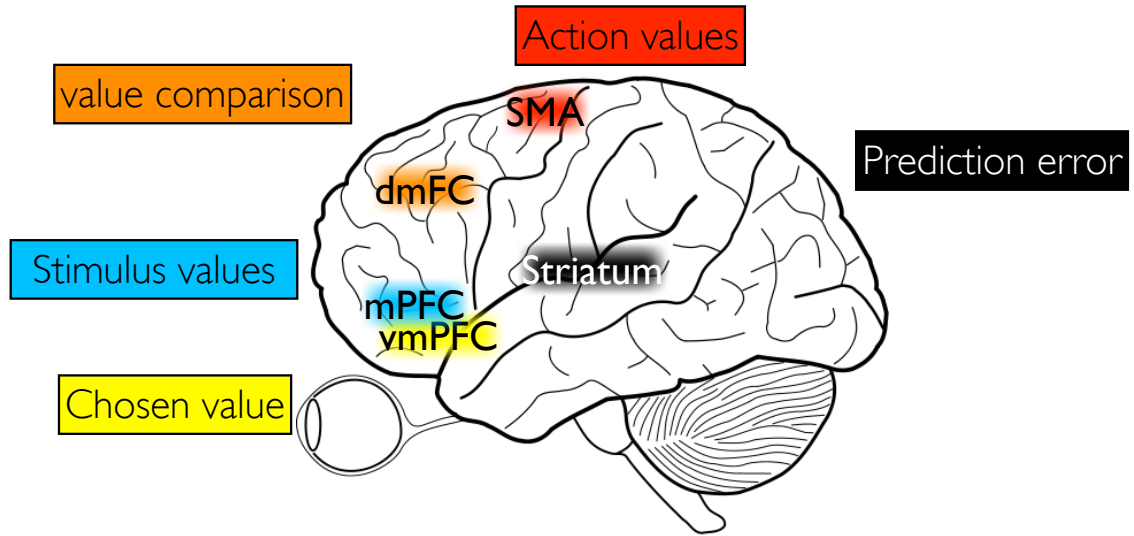


Figure 5.1 Summary of value signals in the human brain

We found various types of value signals in the human brain. Pre-choice value signals: action values for hand and eye movements are encoded in the supplementary motor system. Stimulus values were found in medial prefrontal cortex. Post-choice value signals: the value of the chosen action or stimulus was found in ventromedial prefrontal cortex. Anterior cingulate cortex / dorsomedial frontal cortex encoded a signal based on the value difference between the two available actions (a signal potentially relating to the output of a value comparison process). Finally, the prediction error signal was found in striatum.

# Bibliography

1. Quine, W.O., *Natural kinds*, in *Ontological relativity and other essays*. 1969, Columbia University Press: New York. p. 114-138.
2. Skinner, B.F., *Science and Human Behavior*. 1953, New York: Macmillan.
3. Barto, A.G., R.S. Sutton, and C.W. Anderson, *Neuronlike adaptive elements that can solve difficult learning control problems*. IEEE Trans. Systems Man Cyber., 1983. **13**: p. 834-846.
4. Rescorla, R.A. and A.R. Wagner, *A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement*, *Classical Conditioning II*, ed. A.H. Black and W.F. Prokasy. 1972: Appleton-Century-Crofts. .
5. Sutton, R.S. and A.G. Barto, *Reinforcement Learning: An Introduction*. 1998: MIT Press, Cambridge, MA.
6. Rangel, A., C. Camerer, and P.R. Montague, *A framework for studying the neurobiology of value-based decision making*. Nat Rev Neurosci, 2008. **9**(7): p. 545-56.
7. Valentin, V.V., A. Dickinson, and J.P. O'Doherty, *Determining the neural substrates of goal-directed learning in the human brain*. J Neurosci, 2007. **27**(15): p. 4019-26.
8. von Neumann, J. and O. Morgenstern, *Theory of Games and Economic Behavior*. 1944: Princeton University Press.
9. Dayan, P. and L.F. Abbott, *Theoretical Neuroscience*. 2001: MIT Press, Cambridge, MA.

10. Padoa-Schioppa, C. and J.A. Assad, *Neurons in the orbitofrontal cortex encode economic value*. Nature, 2006. **441**(7090): p. 223-6.
11. Glimcher, P.W., M.C. Dorris, and H.M. Bayer, *Physiological utility theory and the neuroeconomics of choice*. Games Econ Behav, 2005. **52**(2): p. 213-256.
12. Colby, C.L. and C.R. Olsen, *Spatial cognition*, in *Fundamental Neuroscience*, M.J. Zigmond, et al., Editors. 1999, Academic Press: London. p. 1363-1383.
13. Romo, R. and W. Schultz, *Dopamine neurons of the monkey midbrain: contingencies of responses to active touch during self-initiated arm movements*. J Neurophysiol, 1990. **63**(3): p. 592-606.
14. Schultz, W., *Predictive Reward Signal of Dopamine Neurons*. J Neurophysiol, 1998. **80**(1): p. 1-27.
15. Preusschoff, K. and P. Bossaerts, *Adding prediction risk to the theory of reward learning*. Ann N Y Acad Sci, 2007. **1104**: p. 135-46.
16. Samejima, K., et al., *Action value in the striatum and reinforcement-learning model of cortico-basal ganglia network*. Neuroscience Research, 2007. **58**(Supplement 1): p. S22.
17. Samejima, K., et al., *Representation of action-specific reward values in the striatum*. Science, 2005. **310**(5752): p. 1337-40.
18. Lau, B. and P.W. Glimcher, *Action and outcome encoding in the primate caudate nucleus*. J Neurosci, 2007. **27**(52): p. 14502-14.
19. Pasquier, F. and H. Petit, *Frontotemporal dementia: its rediscovery*. Eur Neurol, 1997. **38**(1): p. 1-6.
20. Bechara, A., et al., *Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex*. Cereb Cortex, 1996. **6**(2): p. 215-25.

21. Rahman, S., et al., *Specific cognitive deficits in mild frontal variant frontotemporal dementia*. Brain, 1999. **122 ( Pt 8)**: p. 1469-93.
22. Fellows, L.K. and M.J. Farah, *The role of ventromedial prefrontal cortex in decision making: judgment under uncertainty or judgment per se?* Cereb Cortex, 2007. **17(11)**: p. 2669-74.
23. Koenigs, M. and D. Tranel, *Irrational economic decision-making after ventromedial prefrontal damage: evidence from the Ultimatum Game*. J Neurosci, 2007. **27(4)**: p. 951-6.
24. Damasio, H., et al., *The return of Phineas Gage: clues about the brain from the skull of a famous patient*. Science, 1994. **264(5162)**: p. 1102-5.
25. O'Doherty, J.P., *Reward representations and reward-related learning in the human brain: insights from neuroimaging*. Curr Opin Neurobiol, 2004. **14(6)**: p. 769-76.
26. Wallis, J.D., *Orbitofrontal cortex and its contribution to decision-making*. Annu Rev Neurosci, 2007. **30**: p. 31-56.
27. Padoa-Schioppa, C. and J.A. Assad, *The representation of economic value in the orbitofrontal cortex is invariant for changes of menu*. Nat Neurosci, 2008. **11(1)**: p. 95-102.
28. Colby, C.L., J.R. Duhamel, and M.E. Goldberg, *Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area*. J Neurophysiol, 1996. **76(5)**: p. 2841-52.
29. Gnadt, J.W. and R.A. Andersen, *Memory related motor planning activity in posterior parietal cortex of macaque*. Exp Brain Res, 1988. **70(1)**: p. 216-20.
30. Platt, M.L. and P.W. Glimcher, *Responses of intraparietal neurons to saccadic targets and visual distractors*. J Neurophysiol, 1997. **78(3)**: p. 1574-89.

31. Platt, M.L. and P.W. Glimcher, *Neural correlates of decision variables in parietal cortex*. Nature, 1999. **400**(6741): p. 233-8.
32. Dorris, M.C. and P.W. Glimcher, *Activity in posterior parietal cortex is correlated with the relative subjective desirability of action*. Neuron, 2004. **44**(2): p. 365-78.
33. Sugrue, L.P., G.S. Corrado, and W.T. Newsome, *Matching Behavior and the Representation of Value in the Parietal Cortex*. Science, 2004. **304**(5678): p. 1782-1787.
34. Sugrue, L.P., G.S. Corrado, and W.T. Newsome, *Choosing the greater of two goods: neural currencies for valuation and decision making*. Nat Rev Neurosci, 2005. **6**(5): p. 363-375.
35. Kable, J.W. and P.W. Glimcher, *The neural correlates of subjective value during intertemporal choice*. Nat Neurosci, 2007. **10**(12): p. 1625-1633.
36. Lau, B. and P.W. Glimcher, *Value representations in the primate striatum during matching behavior*. Neuron, 2008. **58**(3): p. 451-63.
37. Roesch, M.R., et al., *Ventral striatal neurons encode the value of the chosen action in rats deciding between differently delayed or sized rewards*. J Neurosci, 2009. **29**(42): p. 13365-76.
38. O'Doherty, J., et al., *Dissociable roles of ventral and dorsal striatum in instrumental conditioning*. Science, 2004. **304**(5669): p. 452-4.
39. Schultz, W., P. Dayan, and P.R. Montague, *A neural substrate of prediction and reward*. Science, 1997. **275**(5306): p. 1593-9.
40. McClure, S.M., G.S. Berns, and P.R. Montague, *Temporal prediction errors in a passive learning task activate human striatum*. Neuron, 2003. **38**(2): p. 339-46.



41. Tobler, P.N., et al., *Human neural learning depends on reward prediction errors in the blocking paradigm*. J Neurophysiol, 2006. **95**(1): p. 301-10.
42. Bray, S. and J. O'Doherty, *Neural coding of reward-prediction error signals during classical conditioning with attractive faces*. J Neurophysiol, 2007. **97**(4): p. 3036-45.
43. Schonberg, T., et al., *Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making*. J Neurosci, 2007. **27**(47): p. 12860-7.
44. Schoenbaum, G., A.A. Chiba, and M. Gallagher, *Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning*. Nat Neurosci, 1998. **1**(2): p. 155-9.
45. Paton, J.J., et al., *The primate amygdala represents the positive and negative value of visual stimuli during learning*. Nature, 2006. **439**(7078): p. 865-70.
46. Gottfried, J.A., J. O'Doherty, and R.J. Dolan, *Encoding predictive reward value in human amygdala and orbitofrontal cortex*. Science, 2003. **301**(5636): p. 1104-7.
47. Schoenbaum, G., et al., *Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala*. Neuron, 2003. **39**(5): p. 855-67.
48. Daw, N.D., et al., *Cortical substrates for exploratory decisions in humans*. Nature, 2006. **441**(7095): p. 876-9.
49. Hampton, A.N., P. Bossaerts, and J.P. O'Doherty, *The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans*. J Neurosci, 2006. **26**(32): p. 8360-7.
50. Hommer, D.W., et al., *Amygdalar Recruitment during Anticipation of Monetary Rewards: An Event-Related fMRI Study*. Ann NY Acad Sci, 2003. **985**(1): p. 476-478.

51. Plassmann, H., J. O'Doherty, and A. Rangel, *Orbitofrontal cortex encodes willingness to pay in everyday economic transactions*. J Neurosci, 2007. **27**(37): p. 9984-8.
52. Hare, T.A., et al., *Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors*. J Neurosci, 2008. **28**(22): p. 5623-30.
53. Fujii, N., H. Mushiake, and J. Tanji, *Distribution of eye- and arm-movement-related neuronal activity in the SEF and in the SMA and Pre-SMA of monkeys*. J Neurophysiol, 2002. **87**(4): p. 2158-66.
54. Campos, M., et al., *Supplementary Motor Area Encodes Reward Expectancy in Eye-Movement Tasks*. J Neurophysiol, 2005. **94**(2): p. 1325-1335.
55. Lau, H.C., et al., *Willed action and attention to the selection of action*. Neuroimage, 2004. **21**(4): p. 1407-15.
56. Thaler, D., et al., *The functions of the medial premotor cortex. I. Simple learned movements*. Exp Brain Res, 1995. **102**(3): p. 445-60.
57. Seo, H., D.J. Barraclough, and D. Lee, *Dynamic Signals Related to Choices and Outcomes in the Dorsolateral Prefrontal Cortex*. Cereb. Cortex, 2007. **17**(suppl\_1): p. i110-117.
58. Seo, H. and D. Lee, *Temporal Filtering of Reward Signals in the Dorsal Anterior Cingulate Cortex during a Mixed-Strategy Game*. J. Neurosci., 2007. **27**(31): p. 8366-8377.
59. Cui, H. and R.A. Andersen, *Posterior parietal cortex encodes autonomously selected motor plans*. Neuron, 2007. **56**(3): p. 552-9.
60. Gold, J.I. and M.N. Shadlen, *The influence of behavioral context on the representation of a perceptual decision in developing oculomotor commands*. J Neurosci, 2003. **23**(2): p. 632-51.

61. Romo, R., A. Hernandez, and A. Zainos, *Neuronal correlates of a perceptual decision in ventral premotor cortex*. *Neuron*, 2004. **41**(1): p. 165-73.
62. Hernandez, A., A. Zainos, and R. Romo, *Temporal Evolution of a Decision-Making Process in Medial Premotor Cortex*. *Neuron*, 2002. **33**(6): p. 959-972.
63. Walton, M.E., J.T. Devlin, and M.F. Rushworth, *Interactions between decision making and performance monitoring within prefrontal cortex*. *Nat Neurosci*, 2004. **7**(11): p. 1259-65.
64. Kennerley, S.W., et al., *Optimal decision making and the anterior cingulate cortex*. *Nat Neurosci*, 2006. **9**(7): p. 940-7.
65. Yin, H.H., B.J. Knowlton, and B.W. Balleine, *Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning*. *Eur J Neurosci*, 2004. **19**(1): p. 181-9.
66. Yin, H.H., et al., *The role of the dorsomedial striatum in instrumental conditioning*. *European Journal of Neuroscience*, 2005. **22**(2): p. 513-523.
67. Haruno, M., et al., *A neural correlate of reward-based behavioral learning in caudate nucleus: a functional magnetic resonance imaging study of a stochastic decision task*. *J Neurosci*, 2004. **24**(7): p. 1660-5.
68. O'Doherty, J., et al., *Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices*. *J Neurosci*, 2003. **23**(21): p. 7931-9.
69. Tricomi, E.M., M.R. Delgado, and J.A. Fiez, *Modulation of caudate activity by action contingency*. *Neuron*, 2004. **41**(2): p. 281-92.
70. Carter, C.S., et al., *Anterior Cingulate Cortex, Error Detection, and the Online Monitoring of Performance*. *Science*, 1998. **280**(5364): p. 747-749.

71. Kerns, J.G., et al., *Anterior Cingulate Conflict Monitoring and Adjustments in Control*. Science, 2004. **303**(5660): p. 1023-1026.
72. Hoshi, E. and J. Tanji, *Differential Roles of Neuronal Activity in the Supplementary and Presupplementary Motor Areas: From Information Retrieval to Motor Planning and Execution*. Journal of Neurophysiology, 2004. **92**(6): p. 3482-3499.
73. Rathelot, J.A. and P.L. Strick, *Muscle representation in the macaque motor cortex: an anatomical perspective*. Proc Natl Acad Sci U S A, 2006. **103**(21): p. 8257-62.
74. Montague, P.R., P. Dayan, and T.J. Sejnowski, *A framework for mesencephalic dopamine systems based on predictive Hebbian learning*. J Neurosci, 1996. **16**(5): p. 1936-47.
75. Knutson, B., et al., *Dissociation of reward anticipation and outcome with event-related fMRI*. Neuroreport, 2001. **12**(17): p. 3683-7.
76. O'Doherty, J., et al., *Abstract reward and punishment representations in the human orbitofrontal cortex*. Nat Neurosci, 2001. **4**(1): p. 95-102.
77. Bayer, H.M. and P.W. Glimcher, *Midbrain dopamine neurons encode a quantitative reward prediction error signal*. Neuron, 2005. **47**(1): p. 129-41.
78. Roesch, M.R., D.J. Calu, and G. Schoenbaum, *Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards*. Nat Neurosci, 2007. **10**(12): p. 1615-24.
79. Glascher, J., A.N. Hampton, and J.P. O'Doherty, *Determining a Role for Ventromedial Prefrontal Cortex in Encoding Action-Based Value Signals During Reward-Related Decision Making*. Cereb Cortex, 2008.
80. Gerardin, E., et al., *Foot, Hand, Face and Eye Representation in the Human Striatum*. Cerebral Cortex, 2003. **13**(2): p. 162-169.

81. Matsumoto, K., W. Suzuki, and K. Tanaka, *Neuronal Correlates of Goal-Based Motor Selection in the Prefrontal Cortex*. *Science*, 2003. **301**(5630): p. 229-232.
82. Shima, K. and J. Tanji, *Role for Cingulate Motor Area Cells in Voluntary Movement Selection Based on Reward*. *Science*, 1998. **282**(5392): p. 1335-1338.
83. Hadland, K.A., et al., *The Anterior Cingulate and Reward-Guided Selection of Actions*. *J Neurophysiol*, 2003. **89**(2): p. 1161-1164.
84. Heekeren, H.R., et al., *Involvement of human left dorsolateral prefrontal cortex in perceptual decision making is independent of response modality*. *Proc Natl Acad Sci U S A*, 2006. **103**(26): p. 10023-8.
85. Heekeren, H.R., S. Marrett, and L.G. Ungerleider, *The neural systems that mediate human perceptual decision making*. *Nat Rev Neurosci*, 2008. **9**(6): p. 467-79.
86. Rorie, A.E. and W.T. Newsome, *A general mechanism for decision-making in the human brain?* *Trends Cogn Sci*, 2005. **9**(2): p. 41-3.
87. Usher, M. and J.L. McClelland, *The time course of perceptual choice: the leaky, competing accumulator model*. *Psychol Rev*, 2001. **108**(3): p. 550-92.
88. Smith, P.L. and R. Ratcliff, *Psychology and neurobiology of simple decisions*. *Trends Neurosci*, 2004. **27**(3): p. 161-8.
89. Busemeyer, J.R. and J.G. Johnson, *Computational models of decision making*, in *Handbook of Judgement and Decision Making*, D. Koehler and N. Narvey, Editors. 2004, Blackwell Publishing Co: New York. p. 133-154.
90. Shadlen, M.N. and W.T. Newsome, *Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey*. *Journal of Neurophysiology*, 2001. **86**(4): p. 1916-1936.

91. Shadlen, M.N., et al., *A computational analysis of the relationship between neuronal and behavioral responses to visual motion*. J Neurosci, 1996. **16**(4): p. 1486-510.
92. Brass, M. and P. Haggard, *To do or not to do: the neural signature of self-control*. J Neurosci, 2007. **27**(34): p. 9141-5.
93. Logothetis, N.K., *The ins and outs of fMRI signals*. Nat Neurosci, 2007. **10**(10): p. 1230-2.
94. Libet, B., *Unconscious cerebral initiative and the role of conscious will in voluntary action*. Behavioral and Brain Sciences, 1985. **8**: p. 529-566.
95. Sumner, P., et al., *Human medial frontal cortex mediates unconscious inhibition of voluntary action*. Neuron, 2007. **54**(5): p. 697-711.
96. Oldfield, R.C., *The assessment and analysis of handedness: the Edinburgh inventory*. Neuropsychologia, 1971. **9**(1): p. 97-113.
97. Duvernoy, H.M., *The Human Brain. Surface, Blood Supply and Three-Dimensional Section Anatomy*. 2nd ed. 1999, New York: Springer.
98. Tzourio-Mazoyer, N., et al., *Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain*. Neuroimage, 2002. **15**(1): p. 273-89.
99. Lau, H.C., et al., *Attention to intention*. Science, 2004. **303**(5661): p. 1208-10.
100. Picard, N. and P.L. Strick, *Activation of the supplementary motor area (SMA) during performance of visually guided movements*. Cereb Cortex, 2003. **13**(9): p. 977-86.
101. Glimcher, P., et al., *Neuroeconomics: Decision Making and the Brain*. 2008: Academic Press.

102. Doya, K. and M. Kimura, *The Basal Ganglia and the Encoding of Value*, in *Neuroeconomics: Decision Making and the Brain*. 2008. p. 407-416.
103. Lee, D. and X.-J. Wang, *Mechanisms for Stochastic Decision Making in the Primate Frontal Cortex: Single-neuron Recording and Circuit Modeling*, in *Neuroeconomics: Decision Making and the Brain*. 2008. p. 481-502.
104. Dickinson, A. and B.W. Balleine, *The role of learning in the operation of motivational systems*, in *Stevens' handbook of Experimental Psychology*, H. Pashler and R. Gallistel, Editors. 2002, John Wiley & Sons: New York. p. 497-533.
105. Wunderlich, K., A. Rangel, and J.P. O'Doherty, *Neural computations underlying action-based decision making in the human brain*. Proc Natl Acad Sci U S A, 2009. **106**(40): p. 17199-17204.
106. Rudebeck, P.H., D.M. Bannerman, and M.F. Rushworth, *The contribution of distinct subregions of the ventromedial frontal cortex to emotion, social behavior, and decision making*. Cogn Affect Behav Neurosci, 2008. **8**(4): p. 485-97.
107. Rudebeck, P.H., et al., *Frontal cortex subregions play distinct roles in choices between actions and stimuli*. J Neurosci, 2008. **28**(51): p. 13775-85.
108. Balleine, B.W. and A. Dickinson, *Goal-directed instrumental action: contingency and incentive learning and their cortical substrates*. Neuropharmacology, 1998. **37**(4-5): p. 407-419.
109. Ostlund, S.B., N.E. Winterbauer, and B.W. Balleine, *Evidence of action sequence chunking in goal-directed instrumental conditioning and its dependence on the dorsomedial prefrontal cortex*. J Neurosci, 2009. **29**(25): p. 8280-7.
110. Balleine, B.W., M. Liljeholm, and S.B. Ostlund, *The integrative function of the basal ganglia in instrumental conditioning*. Behav Brain Res, 2009. **199**(1): p. 43-52.

111. Hampton, A.N. and P. O'Doherty J, *Decoding the neural substrates of reward-related decision making with functional MRI*. Proc Natl Acad Sci U S A, 2007. **104**(4): p. 1377-82.
112. Ward, B.D., *Simultaneous Inference for FMRI Data*. 2000: Wisconsin.
113. Popper, K., *Logik der Forschung*. 1934, Vienna: Springer.
114. Berger, J., *Statistical Decision Theory and Bayesian Analysis*. 1980, New York: Springer.
115. Grant, D.A. and E.A. Berg, *A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem*. J Exp Psychol, 1948. **38**(4): p. 404-11.
116. Downes, J.J., et al., *Impaired extra-dimensional shift performance in medicated and unmedicated Parkinson's disease: Evidence for a specific attentional dysfunction*. Neuropsychologia, 1989. **27**: p. 1329-1343.
117. Koechlin, E., C. Ody, and F. Kouneiher, *The architecture of cognitive control in the human prefrontal cortex*. Science, 2003. **302**(5648): p. 1181-5.
118. Drewe, E.A., *The effect of type and area of brain lesion on Wisconsin card sorting test performance*. Cortex, 1974. **10**(2): p. 159-70.
119. Milner, B., *Effects of different brain lesions on card sorting*. Archives of Neurology, 1963. **9**: p. 100-110.
120. Robinson, A.L., et al., *The utility of the Wisconsin Card Sorting Test in detecting and localizing frontal lobe lesions*. J Consult Clin Psychol, 1980. **48**(5): p. 605-14.
121. Dias, R., T.W. Robbins, and A.C. Roberts, *Dissociation in prefrontal cortex of affective and attentional shifts*. Nature, 1996. **380**(6569): p. 69-72.



122. Monchi, O., et al., *Wisconsin Card Sorting Revisited: Distinct Neural Circuits Participating in Different Stages of the Task Identified by Event-Related Functional Magnetic Resonance Imaging*. J. Neurosci., 2001. **21**(19): p. 7733-7741.
123. Rogers, R.D., et al., *Contrasting cortical and subcortical activations produced by attentional-set shifting and reversal learning in humans*. J Cogn Neurosci, 2000. **12**(1): p. 142-62.
124. Wunderlich, K., A. Rangel, and J.P. O'Doherty, *Neural computations underlying action-based decision making in the human brain*. Proceedings of the National Academy of Sciences, 2009. **106**(40): p. 17199-17204.
125. Stephan, K.E., et al., *Bayesian model selection for group studies*. Neuroimage, 2009. **46**(4): p. 1004-17.
126. Koechlin, E. and A. Hyafil, *Anterior prefrontal function and the limits of human decision-making*. Science, 2007. **318**(5850): p. 594-8.
127. MacKay, D.J.C., *Comparison of Approximate Methods for Handling Hyperparameters*. Neural Computation, 1999. **11**(5): p. 1035-1068.
128. Jacobs, R.A., *Optimal integration of texture and motion cues to depth*. Vision Res, 1999. **39**(21): p. 3621-9.
129. Körding, K.P., et al., *Causal Inference in Multisensory Perception*. PLoS ONE, 2007. **2**(9): p. e943.
130. Ernst, M.O. and M.S. Banks, *Humans integrate visual and haptic information in a statistically optimal fashion*. Nature, 2002. **415**(6870): p. 429-33.
131. Tenenbaum, J.B., *Bayesian modeling of human concept learning*, in *Advances in Neural Information Processing Systems II*, M.S. Kearns, S.A. Solla, and D.A. Cohn, Editors. 1999, MIT Press: Cambridge, MA. p. 59-65.

132. Trommershauser, J., L.T. Maloney, and M.S. Landy, *Statistical decision theory and the selection of rapid, goal-directed movements*. J Opt Soc Am A Opt Image Sci Vis, 2003. **20**(7): p. 1419-33.
133. Kording, K.P. and D.M. Wolpert, *Bayesian integration in sensorimotor learning*. Nature, 2004. **427**(6971): p. 244-7.
134. Griffiths, T.L. and J.B. Tenenbaum, *Structure and strength in causal induction*. Cognit Psychol, 2005. **51**(4): p. 334-84.
135. Gopnik, A., et al., *A theory of causal learning in children: Causal maps and Bayes nets*. Psychological Review, 2004. **111**: p. 1-31.
136. Lawrence, B.J., et al., *Executive and mnemonic functions in early Huntington's disease*. Brain, 1996. **119**: p. 1633-1645.
137. Owen, A.M., et al., *Extra-dimensional versus intra-dimensional set shifting performance following frontal lobe excisions, temporal lobe excisions or amygdalo-hippocampectomy in man*. Neuropsychologia, 1991. **29**(10): p. 993-1006.
138. Owen, A.M., et al., *Contrasting mechanisms of impaired attentional set-shifting in patients with frontal lobe damage or Parkinson's disease*. Brain, 1993. **116** ( Pt 5): p. 1159-75.
139. Diaconis, P. and D. Freedman, *On the Consistency of Bayes Estimates*. The Annals of Statistics, 1986. **14**(1): p. 1-26.
140. Burnham, K.P. and D. Anderson, *Model Selection and Multi-Model Inference*. 2nd ed. 2002: Springer.
141. Dearden, R., et al. *Bayesian Q-learning*. in *15th National Conference on Artificial Intelligence (AAAI 98) / 10th Conference on Innovative Applications of Artificial Intelligence (IAAI 98)*. 1998. Madison, Wi: Amer Assoc Artificial Intelligence.

142. MacKay, D.J.C., *Choice of Basis for Laplace Approximation*. Machine Learning, 1998. **33**(1): p. 77-86.