Neural Mechanisms of Suboptimal Decisions

Bolton Ka Hung Chau

A thesis submitted in partial fulfilment of the requirements
for the degree of D.Phil. in University of Oxford.

Wolfson College,
Trinity Term 2014

Neural Mechanisms of Suboptimal Decisions

Bolton Ka Hung Chau

Abstract

Making good decisions and adapting flexibly to environmental change are critical to the survival of animals. In this thesis, I investigated neural mechanisms underlying suboptimal decision making in humans and underlying behavioural adaptation in monkeys with the use of functional magnetic resonance imaging (fMRI) in both species.

In recent decades, in the neuroscience of decision making, there has been a prominent focus on binary decisions. Whether the presence of an additional third option could have an impact on behaviour and neural signals has been largely overlooked. I designed an experiment in which decisions were made between two options in the presence of a third option. A biophysical model simulation made surprising predictions that *more suboptimal decisions* were made in the presence of a very *poor* third alternative. Subsequent human behavioural testing showed consistent results with these predictions. In the ventromedial prefrontal cortex (vmPFC), I found that a value comparison signal that is critical for decision making became weaker in the presence of a poor value third option. The effect contrasts with another prominent potential mechanism during multi-alternative decision making – divisive normalization – the signatures of which were observed in the posterior parietal cortex.

It has long been thought that the orbitofrontal cortex (OFC) and amygdala mediate reward-guided behavioural adaptation. However, this viewpoint has been recently challenged. I recorded whole brain activity in macaques using fMRI while they performed an object discrimination reversal task over multiple testing sessions. I identified a lateral OFC (lOFC) region in which activity predicted adaptive win-stay/lose-shift behaviour. In contrast, anterior cingulate cortex (ACC) activity predicted future exploratory decisions regardless of reward outcome. Amygdala and lOFC activity was more strongly coupled for adaptive choice shifting and decoupled for task irrelevant reward memory. Day-to-day fluctuations in signals and signal coupling were correlated with day-to-day fluctuations in performance. These data demonstrate OFC, ACC, and amygdala each make unique contributions to flexible behaviour and credit assignment.

A thesis submitted in partial fulfilment of the requirements
for the degree of D.Phil. in University of Oxford.
Approximate word count: 45,200 words
Wolfson College,
Trinity Term 2014

Acknowledgments

Over the four years of doctorate study, one question I have been asked repeatedly is that how much I like my study. I am fortunate enough to have an exceptional experience in Oxford because of two reasons. First, I have been working on a research topic that I am strongly interested in. Second, I have enjoyed a great research environment.

As a research student, perhaps to work in a lab with good funding and facilities are important, but actually it is the most important to have great supervisors. Matthew Rushworth and Mark Walton are definitely great supervisors to me. Matthew is a very kind and intelligent person. He has shown me a lot of patience and support and has selflessly spent vast amount of time supervising me. I still remember when we were working on the human experiment, Matthew stayed in the office till late on several occasions to discuss and analyze the fMRI data with me. To make my doctorate experience even better, he also introduced Mark to me as a co-supervisor. I would like to express my deepest gratitude for Matthew's exceptionally good supervision.

I feel very grateful that I also have Mark as my supervisor. He has opened up some invaluable opportunities to broaden my learning experience. He has often given useful advice to my works, especially when I was writing up my experiments and when I was preparing for presentations. I felt so grateful when he included me to discuss some experiments in his lab and to do some writing with him. I would like to thank Mark for adding me a big bonus to make my doctorate study even more pleasurable.

Jerome Sallet is always supportive, especially to my monkey works. I would like to thank Jerome for teaching me how to conduct behavioural training, to build equipment for the experiments and to operate the MRI scanner. I enjoyed the time when we talked about science and worked together to solve problems in the monkey experiments. He is the key person that has made our monkey fMRI experiment possible.

Nils Kolling has significant contributions to turn the human experiment into a nice paper. Nils was very generous for spending his time running various versions of the biophysical model; I know this was very tedious work. Moreover, I had great times with people in the Rushworth lab, especially with Jacquie Scholl, Marco Wittmann, Nils Kolling and Rei Akaishi. We often have lunch and coffee together and always show each other enormous peer support.

I would like to show my gratitude to two important people. Miss C.P. Yuen and Mr C.W. Sham are two great science teachers in my secondary school. They are both very passionate and caring teachers. They made science very interesting to me, this is definitely essential for inspiring me to pursue for a career as a scientist.

I am also lucky enough to have a family that is always supportive to my ambitions. My dad is a very traditional Chinese father. He sometimes says to my sister and me "don't be bothered by other things, just stay focused on your studies". I think he actually meant to

encourage us to pursue what we want and not to get worried about life. Mom, sister and I sometimes tease him for saying that. I would like to thank my family for their tolerance to my enormously long education compared to most people in the society.

This thesis was written during an unusual time when my hometown Hong Kong, one of the world's most dynamic and orderly cities, became a very tense and unstable place. On 28th September 2014, the police attempted to crack down a peaceful protest for democracy by using tear gas and pepper spray. This was the strongest force used towards our own citizen that I have ever seen. This caused a larger scale protest with more than 100,000 people gathered in several downtown areas, it is sometimes known as the Umbrella Movement in the media. People in the protest wanted to voice out their anger towards violence and their desire for a true universal suffrage for electing the Chief Executive. Through modern technology, I was able to witness that Hong Kong people can be highly rational, peaceful and caring even though they were protesting to express their dissatisfaction. I thank those people who have contributed to make Hong Kong a more liveable place. I dedicate this thesis to the peace, democracy and freedom of Hong Kong.

List of abbreviations

ACC	anterior cingulate cortex
BOLD	blood oxygen level dependent
cOFC	central OFC
COpt	value of a chosen option on the contralateral side
D	distractor
dlPFC	dorsolateral prefrontal cortex
FEF	frontal eye field
FLAME	FMRIB's local analysis of mixed effects
fMRI	functional magnetic resonance imaging
FSL	FMRIB's Software Library
GLM	general linear model
HV	higher value option
I_D, I_{HV}, I_{LV}	inputs from D, HV and LV respectively
IPL	inferior parietal lobule
IPS	intraparietal sulcus
IV	total option value of the ipsilateral side
LCA	leaky, competing accumulator
LIP	lateral intraparietal cortex
lOFC	lateral orbitofrontal cortex
LOS	lateral orbital sulcus
LV	lower value option
MEG	magnetoencephalography
MIP	medial IPS region
MNI	Montreal Neurological Institute
mOFC	medial orbitofrontal cortex

MOS	medial orbital sulcus
ODR	object discrimination reversal
OFC	orbitofrontal cortex
P_D, P_{HV}, P_{LV}	excitatory neuronal populations for D, HV and LV respectively
P_i	inhibitory neuronal population
PFC	prefrontal cortex
PPC	posterior parietal cortex
RDM	random dot motion
RT	reaction time
SMH	somatic marker hypothesis
SPL	superior parietal lobule
vlPFC	ventrolateral prefrontal cortex
vmPFC	ventromedial prefrontal cortex
WSLS	win-stay/lose-shift

Contents

Abstract i

Acknowledgments ii

List of abbreviations iv

Contents vi

Chapter 1. General Introduction 1

1.1 Decision Making 1

1.2 Reinforcement Learning 13

1.3 Ventral Prefrontal Cortex 24

1.4 Posterior Parietal Cortex 37

1.5 The current thesis 44

Chapter 2. Suboptimal Decisions with Multiple Alternatives: Biophysical Model 46

2.1 Introduction 46

2.2 Methods 49

2.3 Results 53

2.4 Discussion 61

Chapter 3. Suboptimal Decisions with Multiple Alternatives: Human Behaviour 67

3.1 Introduction 67

3.2 Methods 70

3.3 Results 77

3.4 Discussion 91

Chapter 4. Suboptimal Decisions with Multiple Alternatives: Neural Mechanism 98

4.1 Introduction 98

4.2 Methods 100

4.3 Results 105

4.4 Discussion 117

Chapter 5. Contrasting Roles for Orbitofrontal Cortex, Anterior Cingulate Cortex, and Amygdala in Credit Assignment, Learning, and Exploration in Macaques 123

5.1 Introduction 123

5.2 Methods 126

5.3 Results 134

5.4 Discussion 153

Chapter 6. General Discussions 159

6.1 Multialternative Decision Making 159

6.2 Behavioural Adaptation 161

6.3 A Neural mechanism for Selecting Task Relevant Parameters 163

6.4 Future Directions 164

Bibliography 166

Chapter 1. General Introduction

1.1 Decision Making

Decision making generally involves estimating the value of options followed by value comparison in order to choose the best option, and finally response selection so that the choice is effected (Gold and Shadlen, 2007; Rangel et al., 2008; Rushworth and Behrens, 2008; Kable and Glimcher, 2009; Padoa-Schioppa, 2011; Shadlen and Kiani, 2013). In experimental settings, two types of decisions that challenge people in different ways– value-guided and perceptual decision making – have been the subject of particularly intensive investigation in recent years. Value-based decision making usually involves choosing between objects (such as food items or images that are symbolic or indicative of a particular reward) or between different actions (such as turning left or right at a cross road) that are associated with different expectations for immediate or future rewards. A key challenge in value-based decision making is assigning subjective values to options so that comparisons can be made between them. The estimation of subjective values depends on external features that are related to the properties of the objects/actions and on the decision-making agent's internal states. For instance, imagine that the menu of a restaurant offers a choice between a large steak for £25 and a small cheese sandwich for £8. To make such a decision, one needs to compare the external features of taste, energy levels and costs which are each associated with the two options. Moreover, internal states, such as hunger, recent consumption of a specific food type and financial plans, are also taken into account in order to assign values. Another example would be choosing between a good restaurant that is far away and a bad restaurant nearby – decision makers need to consider the external features of food quality and walking distance as well as the internal state of physical fatigue.

Knowledge of the external features of the options and about the decision-maker's internal states is acquired either by explicit instructions (e.g. the price listed on the menu) or by reinforcement learning (e.g. consumption of a menu option, such as a cheese sandwich, allows estimation of its quality, especially if it is done more than once).

Usually during many investigations of value-based decision making, the perceptual properties of the objects or actions are unambiguous; such paradigms mimic the situation when it is clear to decision makers that the triangular food item is the sandwich and the numbers on the menu are the prices. In contrast *perceptual* decision making involves discrimination of the physical, and often ambiguous, features of stimuli (Gold and Shadlen, 2007; Summerfield and Tsetsos, 2012). Such experimental paradigms might be thought of as modelling the situation when the diner in our example attempts to identify the ingredients while tasting the steak or attempts to read the menu hanging on a wall that is far away.

In laboratories, one of the most widely used paradigms to study perceptual decisions is the random dot motion (RDM) task (Shadlen and Newsome, 1996; Gold and Shadlen, 2002). In most cases, the task involves a proportion of the dots moving in one coherent direction while the other perceptually identical dots move in random directions. All the dots are presented within the same area on the computer screen. Experimental subjects are required to report the motion direction of the coherent dots. The difficulty of the task can be manipulated by adjusting the relative proportion of dots that move coherently or in random directions. This paradigm captures two key components of perceptual decision making. First of all, it involves decisions about the physical characteristic of stimuli (whether the dots are moving coherently in the left or the right direction). Second, the stimuli presented are often ambiguous or noisy such that it is necessary to accumulate evidence favouring one decision against another over time (it is only after sampling the

movement direction of the dots over a period of time that one can establish whether they are moving coherently in one direction).

In making perceptual choices, decision makers are generally challenged at a perceptual level to make a choice that is externally correct; there is a clear external criterion for establishing which direction the dots are moving in the RDM task. However, in value-based choices people make decisions to maximize their reward in a more subjective way, for example there is no clear answer to whether it is better to have a cheese sandwich or a ham sandwich for lunch and choices are based on subjective preference.

It is not entirely clear how the neural substrates for value-based decisions and perceptual decisions are different. For example, the lateral intraparietal (LIP) cortex in monkeys has a strong relationship with the evidence accumulation process for making decisions in the RDM task (Shadlen and Newsome, 1996). In this paradigm the monkeys are trained to make saccades to one of two spatial positions as a function of the direction in which most dots are moving. For example, in one set of trials a monkey might observe random dot kinetograms in which the coherent motion direction was either to the right or the left. In response, on each trial, the animal would make an eye movement or saccade to a target to the left or the right of the kinetogram. Neurons in the LIP are tuned to respond preferentially to one location in space (Goldberg et al., 1990; Sereno et al., 2001; Sereno and Huang, 2014); they become more active when an animal is about to make a saccade to that position and the firing rate of these neurons ramps up when the dots are moving coherently to the tuned direction (Shadlen and Newsome, 1996, 2001; Gold and Shadlen, 2007). The rate at which the neurons firing rate increases, however, is proportional to the strength of evidence about which direction of motion is the better one; it is proportional to the coherency of the RDM stimulus. The high firing state is sustained even when the dots are turned off and before a choice response is made (Shadlen and Newsome, 1996, 2001;

Churchland et al., 2008). Neurons in the same region also encode the inherent reward value of visual stimuli for value-based decisions – they fire more rapidly in response to stimuli that are more likely to lead to a primary reward (Platt and Glimcher, 1999).

One brain region which is particularly important to value-based decision making, perhaps not so much for perceptual decision making, is the ventromedial prefrontal cortex (vmPFC). Patients with lesion in this region and adjacent areas exhibit more choice inconsistencies when they are asked to make judgements on the basis of subjective preferences (Fellows and Farah, 2007). For example if a person prefers apple to orange and orange to banana then usually their preference between apple and banana is consistent; usually they prefer apple over banana. However vmPFC lesion patient preferences appear irrational because their preferences cannot be ordered in the normal way. However, when the same patients were given perceptual decision-making tasks to judge the size or colour of stimuli, their performances were as good as the healthy controls (Fellows and Farah, 2007). Various other studies in humans (Bechara et al., 2000; Kable and Glimcher, 2007; Plassmann et al., 2007; Boorman et al., 2009; Hare et al., 2009; Hunt et al., 2012), non-human primates (Roberts, 2006; Bouret and Richmond, 2010; Noonan et al., 2010; Strait et al., 2014) have shown that the vmPFC plays a crucial role in value-based decision making. There is also an important body of research on value and reward representations in the adjacent orbitofrontal cortex of rodents (Schoenbaum et al., 2003) but because it is not clear if this is the same brain region as the primate vmPFC and because it is difficult to reconcile findings made in primate vmPFC and in rodent OFC here in this thesis I will focus on the human and non-human primate vmPFC. The research reported in this thesis has a strong focus on value-based decision making both in relation to computational model predictions (Chapter 2), behaviour (Chapter 3), and neural mechanisms (Chapters 4 and 5).

Irrational choice in multi-alternative decision making

Over the last few decades of decision neuroscience development, most studies have focused on binary choices (Churchland and Ditterich, 2012). While making decisions between two options, people draw on a valuation process so as to estimate the value of the options via explicit cues and/or previous experience. Values of the two options are compared in order to choose an option with the higher value (Rangel et al., 2008). Using this framework for binary decision making allows simple but controlled investigation of the neural mechanism of value-based decision making.

In economics, the traditional rational choice theory makes assumption that people's choices are *independent of irrelevant alternatives* during multi-alternative decision making (Luce, 1959). In other words, when there are three options, decisions should not be influenced by the presence of a third option that is unchooseable or has a poorer value relative to the other two options. The choice between a lobster and a steak in a dinner menu should not be influenced by the presence of a poor tiny cheese sandwich. If this is true, arguably, most knowledge about binary decisions would be transferable to multi-alternative decisions. For example, the vmPFC, which is important for making binary choices, should exhibit the same kind of activity for both two-option and three-option decisions. However, whether our understanding for binary decisions can be applied to decisions involving multiple options is not clear and is one of the subjects of this thesis. In this thesis, "multiple alternative" decision making usually refers to three options decision making – it provides a simple starting point to investigate the commonalities and differences between binary and multi-alternative decision making.

There is a large body of evidence in behavioural economics showing that people make decisions that violate the notion of independence from irrelevant alternatives (Tversky, 1972; Huber et al., 1982; Simonson, 1989; Tversky and Simonson, 1993; Usher and

McClelland, 2004). Although it is still not well-understood exactly how an "irrelevant" third option affects decisions between the "relevant alternatives", there are two major approaches to study the phenomenon – by investigating the integrated value of the third option or by investigating the decision attributes of the third option. Next, I will discuss computational models and behavioural evidence for how the integrated value and the decision attributes of a third option affect rational choice.

Integrated value approach: behaviour

Recently, there have been investigations of how a model based on "divisive normalization" could potentially explain the effect of the value of a third option on decision making. The divisive normalization model was originally developed in neurophysiology to predict how the sensitivities of primary visual cortex neurons are adjusted to respond to stimuli under different environmental conditions (Heeger, 1992). It argues that neurons fire more rapidly as a function of the intensity of stimuli inside their receptive field divided by (normalized by) the sum of intensity of all stimuli presented inside and outside their receptive field. In visual perception, a divisive normalization mechanism is adaptive because it allows neurons to flexibly change their sensitive response range for detecting meaningful changes in the environment (Reynolds and Heeger, 2009; Carandini and Heeger, 2012). For example, in the dark it is important to be more sensitive to small changes in brightness to obtain meaningful information, such as the ignition of a candle. In a bright room, it is less meaningful to detect minute fluctuations of the brightness of a lamp even though the *absolute* change in brightness could be comparable to the candle light.

It is possible that a similar model could be applied to decision making so that the sensitive range could be adjusted for making decisions between different categories of

options with different value range. For instance, a small increment of cost is meaningful in lower cost items such as food and stationeries (e.g. a notebook of £1 compared to a similar notebook of £2), but not as much in higher cost items such as cars and houses (e.g. a car of £30,001 and a similar car of £30,002). Furthermore, interesting predictions could be made when divisive normalization is applied to multi-alternative decision making (Louie and Glimcher, 2012; Louie et al., 2013). The model predicts that the preference between two options becomes weaker and weaker when the value of a third option becomes larger, because the normalization term (the sum of value of all options) varies as a function of the third option. For example, a preference of a lobster over a steak in the presence of a tiny sandwich choice on the dinner menu will become weaker if the size of the sandwich becomes bigger.

Louie and colleagues (2013) tested empirically whether normalized pattern of decision making were exhibited in humans and monkeys when choices were made between three options. In trinary choices, an option with the lowest value out of the three is thought to be irrelevant. When the value of a *very poor* irrelevant third option became larger, both species showed weaker preference between the relevant options that is consistent with the divisive normalization prediction. Intriguingly in humans, when the value of the third option was even larger and gradually approaches the value of the relevant options the effect eventually inverted (opposite to divisive normalization prediction). In other words, the preference for the lobster over the steak becomes weaker when the tiny sandwich was substituted by a large sandwich on the dinner menu, but surprisingly the preference becomes stronger again when the large sandwich is substituted by an extra large sandwich.

Adding to the confusion is that the normalized preference observed in the monkeys was not found in a group of healthy monkeys in an earlier study by Noonan and colleagues (2010). When these monkeys were choosing manually between three options associated

with different reward probabilities, their decisions were not influenced by different values of the third option. Normalized preference was instead observed in monkeys with a lesion that included part of vmPFC – the medial orbitofrontal cortex (mOFC). This suggests that the pattern of decision making predicted by divisive normalization can only be observed in certain contexts in healthy human or monkey subjects. This issue will be addressed in Chapter 3 that reports the findings of a series of experiments in multi-alternative decision making in humans.

Integrated value approach: connectionist model

Another way to study the value effect of an irrelevant option on decision making is by performing neural network modelling that simulates actual brain activity and investigate the conditions under which "irrational choices" are made by the model. Wang (2002) developed a biophysically plausible connectionist model that could be used to simulate neuronal activity in a decision making network while making perceptual decisions in a RDM task. Specifically, the model involves two populations of pyramidal cells (excitatory neurons) each of which receives input that relates to evidence favouring decisions for one of the two motion directions. Pyramidal cells are connected within the same population to allow evidence to be *"remembered" via recurrent excitation*. To generate decisions, both pyramidal populations have excitatory connections to a common population of interneurons (inhibitory neurons) so that the pyramidal populations *compete via mutual inhibition*. In other words, when a coherent rightward dot movement is observed, the pyramidal population for making a rightward movement decision received strong input and fired rapidly. It exerted a strong excitatory influence on the interneurons, which in turn exerted a strong inhibitory influence on the pyramidal population for left movement decisions. The mutually inhibitory interactions between the leftward and rightward motion populations

lead to just one of them remaining in a high firing attractor state while the other population moves to a low firing attractor state. In this particular example it is likely that the rightward motion population would become increasingly active and the leftward motion population would become less and less active. When this happens the network has completed its decision.

This class of model is compelling because it attempts to simulate a decision network that is biophysically plausible by using parameters obtained from previous neurophysiological studies, such as cell type-specific assignment of membrane capacitance and conductance and receptor type-specific assignment of current time constants to allow different kinds of excitatory activity. For instance, on pyramidal cells the AMPA receptors were assigned with a short time constant (less than 5 ms) and the NMDA were assigned with a long time constant (100 ms). This generates predictions to the activity of a decision neural network in the actual brain. For example, the Wang study (2002) suggested that the "memory" of the network is contributed by slow recurrent excitation maintained mainly by NMDA receptors and the Wei and colleagues study (2012) suggested that the limitation in working memory capacity is contributed by the recurrent dynamics of the pyramidal cells of the network. In Chapter 2, I report an extended version of the Wang model that allows the network to process input from three options and make predictions about how the value of a third option could affect choices in value-based decision making.

Dimensional approach: behaviour

The second approach to study how the presence of an irrelevant third option could cause irrational decisions is by investigating the dimensions or decision attributes of the third option. Often people make decisions that involve multiple attributes. For example when

buying a house, people consider not only the price, but also the location, the size, the interior, and the environment. The dimensional approach, although not disregarding the integrated value of all three options that are being decided between, places strong emphasis on interactions that occur between the component attributes of the options. The third option is sometimes known as a *decoy* and its effects on choices between the relevant options are known as *decoy effects*.

Experiments studying decoy effects often present subjects with two options with equal integrated values but where those equal integrated values are actually the consequences of different combinations of component attribute magnitudes. For example, one could have equal preference for buying a 24" computer monitor A for £150 and a 21" monitor B for £120. Here the two component attributers of the options are their size and their cost. An investigation into the decoy effect could then be conducted by offering a third choice, a decoy monitor C, that is 19" in size and for a cost of £120 and investigate how people's preference changes. There are a number of different decoy effects depending on the configurations of decision attributes. The case of the computer monitor choice given just now is an example of an *asymmetric dominance* decoy effect that shifts people's preference for choosing B over A. This is thought to happen because it is obvious that B has a larger size than C but an equal price to C; it is therefore clear that B is better than at least one other option, C, and that advantage is generalized to the comparison between B and A too (Simonson, 1989; Tversky and Kahneman, 1991; Tversky and Simonson, 1993; Roe et al., 2001; Usher and McClelland, 2004; Pettibone and Wedell, 2007). Another example is the *compromise effect* in which the decoy has the same value as the other two options but it is associated with extreme magnitudes on both of its dimensions, for example a 27" monitor D for £180 (Tversky, 1972; Simonson, 1989; Roe et al., 2001; Usher and McClelland, 2004). People are more likely to avoid the extreme D choice (largest but most expensive) and B choice (smallest but cheapest) and select the compromise A choice (medium size and price).

Tversky provided mathematical explanations for how choice preference is shifted in the presence of different types of decoy. The relative advantage model (Tversky and Simonson, 1993) calculates in each dimension the total advantage and the total disadvantage separately of each option compared to all other options. Taking again the example of monitor A (24" and £150), B (21" and £120) and D (27" and £180), rational choice theory predicts equal preference between A and B. Monitor A has an advantage of £30 and a disadvantage of 3" to D, whereas monitor B has an advantage of £60 and a disadvantage of 6" to D (comparison between A and B is left out for simplicity). Since loss aversion has to be taken into account (Kahneman and Tversky, 1979; Tversky and Kahneman, 1991), the model predicts that people are more likely to choose monitor A to avoid the larger disadvantage of choosing monitor B.

Dimensional approach: connectionist model

Connectionist models have been developed to simulate attribute dynamics to predict decoy effects. In the three option version of the leaky, competing accumulator model (LCA model) (Usher and McClelland, 2001, 2004), there are three neuronal populations representing the value of three options and they compete with each other via lateral inhibition. There are two main differences between the LCA model and the Wang model. First, the LCA model lacks the specification of receptor properties, which is central to the Wang model to make more specific predictions to activity related to different receptors, such as the prediction that reward memories are maintained mainly by NMDA receptors but not AMPA receptors (Wang, 2002). Second, the LCA model has an extra layer of populations that receives input from individual attributes of an option (Usher and McClelland, 2004). For example, in decisions where each option has two attributes, reward magnitude and probabilities, the network would contain individual "attribute populations" that receive input from either

reward magnitude or probability of an option. In this kind of design, "value populations" in the LCA model receive excitatory inputs from the "attribute populations" to accumulate evidence to represent an integrated value, instead of receiving inputs from the stimuli directly in the Wang model (Wang, 2002). This design of having attribute populations make it possible to model decoy effects, in which the dynamics between attributes of options is critical.

It is important to incorporate the idea of loss aversion in order to explain decoy effects (Kahneman and Tversky, 1979; Tversky and Kahneman, 1991; Tversky and Simonson, 1993). In one variant of the LCA model, Usher and McClelland (2004) implemented a non-linear input function so that the firing rates with negative values (which is biophysically impossible) are converted to zero. This model exhibits loss aversion properties as a result. This version of the LCA model is able to account for the asymmetric dominance and compromise decoy effects mentioned above.

The potential for an integrated approach

It should be noted that the integrated value approach and the dimension-specific approach for studying the effect of a third irrelevant option on decision making should not be considered necessarily mutually exclusive. It is often for reasons of simplicity rather than principle that studies using an integrated value approach do not often investigate decision dimensions and studies using a dimension-specific approach involve relevant options with equal values. It is possible that a unified model marrying the strength of biological plausibility that the Wang model (Wang, 2002) possesses with the strength of attribute dynamics that the LCA model (Usher and McClelland, 2004) possesses could provide a more comprehensive account for the effect of a third option on decision making.

1.2 Reinforcement Learning

When people make decisions, they need to estimate the value of each option in order to select a choice. However, often choice information is not explicitly provided and the valuation process relies heavily on expectations established via previous reinforcement learning. For example while buying readily-made pasta in supermarkets, the price for each style of pasta is clearly stated and you can trust that it is the price you will eventually need to pay. However, it is difficult to explicitly describe the exact taste of the different types of pasta and you need to rely on your previous experiences eating those pastas in order to know how they taste.

The close relationship between decision making and reinforcement learning was described by Edward Thorndike. In his law of effect (Thorndike, 1933), responses that caused a rewarding outcome are more likely to be repeated in the same context. If the pasta you bought from the supermarket is tasty, then you are more likely to go to the same place and buy the same pasta again next time. This example highlights that during learning, associations are drawn between stimulus-action (the pasta and buying the pasta), stimulus-outcome (the pasta and a good flavour) and action-outcome (buying the pasta and the flavour) so that future decisions could be made based on the acquisition of these associations.

The win-stay/lose-shift strategy

Most of the time, readily-made pasta from supermarkets are mass produced from factories and the taste of the same kind of pasta from the same supermarket is almost always the same. After trying any of the types of pasta for the first time, you can establish a good estimate of how much you like it and decide whether to buy the same pasta again or

not next time. In other words, the outcome expectation of choosing an option in the future can be fully dependent on a single sampling experience of the same option in these situations. This kind of strategy is sometimes known as win-stay/lose-shift (WSLS) (Robbins, 1952) – you purchase the same chicken pasta again if it tastes good or try another one, such as salmon pasta, if it is bad.

In laboratories, learning and decision making that involves WSLS strategy can be investigated using reversal tasks with deterministic schedules. One example of these tasks is object discrimination reversal (ODR) task (Butter, 1969; Jones and Mishkin, 1972; Murray et al., 2007; Rudebeck and Murray, 2008; Kazama and Bachevalier, 2009; Ghahremani et al., 2010; Hampshire et al., 2012). It is frequently designed in a way that requires experimental subjects to choose between two actual objects or two virtual objects presented on a computer screen. One object is associated with a reward whereas choosing the other object does not lead to a reward. The reward contingencies reverse either in the middle of the task or on the next testing day, such that the previously rewarded option becomes non-rewarded and vice versa for the previously non-rewarded option. To continue to gain reward, the experimental subjects need to inhibit choosing the previously rewarded option and switch to choose the previously unappealing option. In this kind of task knowing the outcome of an object on one single trial is sufficient to predict the outcome of the same object for an extended number of trials. Hence, employing the WSLS strategy – repeat any rewarded decisions (win-stay) and forgo any non-rewarded decisions (lose-switch) – allows a quick response to contingency change and maximizes gain in this task.

Numerous lesion studies in monkeys suggest that the orbitofrontal cortex (OFC), especially the central OFC (cOFC) region at area 11/13, is important for performing in ODR tasks (Butter, 1969; Jones and Mishkin, 1972; Dias et al., 1996a; Murray and Izquierdo, 2007; Rudebeck and Murray, 2008). Monkeys with cOFC lesion are typically characterized by

having comparable performance to controls in their initial discrimination performance when they are first identifying the rewarded option in a pair of novel objects. However after reversals, the cOFC lesion monkeys show more choice perseverance. They repeatedly choose the previously rewarded option which is no longer rewarded after the reversal. This suggests that one critical function of the OFC is to allow agents to adapt behaviour in response to environmental changes (Kringelbach and Rolls, 2004; Roberts, 2006).

Intriguingly, this view has recently been challenged by a study that compared ODR performance when a similar cOFC region is ablated by a different lesion method (Rudebeck et al., 2013b). In this study, impairment in reversals was not observed when the lesion was made by using neurotoxins that damaged cOFC cell bodies and spared the fibres of passage in and near the cortex. The impairment in reversal behaviour that had been found in many previous studies was often observed in monkeys in which the cOFC lesion was made by the aspiration technique. The authors argued that the behavioural deficit could be due to damage in the subjacent fibres rather than damage in the cOFC itself. Supporting this view, the authors made a narrow aspiration lesion at a very posterior OFC region in another group of monkeys, aiming at damaging the underlying uncinate fascicle. These monkeys made perseverative errors after reversals that were similar to those monkeys with aspiration lesion in the cOFC. The conclusion that was drawn, therefore, was that lesions to cOFC itself had not been responsible for causing reversal impairments but that the critical determinant of impairment had been the damage sustained by the uncinate fascicle with projects to other areas of the frontal cortex and not just cOFC. The identity of the critical frontal region the deafferentation of which might really have caused the reversal impairment, however, remained unclear.

One weakness of this argument is that there is a substantial body of work that *has* reported reversal-related impairments after lesions that spare fibres of passage because the

lesions were made by excitotoxin injection or because the lesions were neurotransmitter depleting lesions (Dias et al., 1996a, 1997; Saddoris et al., 2005; Schoenbaum et al., 2007; Stalnaker et al., 2007; Walker et al., 2009; Man et al., 2010; Rygula et al., 2010; Clarke et al., 2014; Rygula et al., 2014). Rudebeck and colleagues (2013b), however, pointed that all of these studies had involved rodents or new world monkey such as marmosets rather than old primates such as macaques. They argued that the OFC region identified in these studies might not be homologous with the cOFC region in old world primates. In addition to points of continuity there are clearly important differences in frontal cortical organization in rodents and primates (Preuss, 1995; Passingham and Wise, 2012). The degree of similarity/dissimilarity between OFC anatomy in old and new world primates has received less attention. While there are differences in cytoarchitecture, marmoset OFC connexions resemble those of macaque OFC (Roberts et al., 2007).

One method for identifying brain regions that are important for behavioural adaptation is by performing functional magnetic resonance imaging (fMRI) to record activity throughout the whole brain. Several studies that involve human subjects performing in reversal learning tasks have reported an increased blood oxygen level dependent (BOLD) signal in a lateral OFC region (lOFC; the orbital surface of area 12/47) during reversals (Hampshire and Owen, 2006; Ghahremani et al., 2010; Hampshire et al., 2012). Perhaps the reversal impairment observed in monkeys with OFC lesions using the aspiration technique was caused by deafferentation of this lOFC region if not the cOFC. An early study by Butters (Butter, 1969) compared ODR performance in monkeys with subtotal or total OFC lesions. Here, impairment in reversal performance was found to be most robust in monkeys with lOFC lesions that extended to the lateral surface and in monkeys with total OFC lesions, whereas monkeys with lesions restricted to the anterior OFC or the posteriomedial OFC had reversal performances that were comparable to controls. Kringelbach and Rolls (2004) also conducted a meta-analysis to review the function of the OFC and found that the lOFC is most

commonly reported for its role in behavioural change. In Chapter 5, I report an experiment with macaques performing in an ODR task while undergoing whole brain fMRI and identified an lOFC region that is activated specifically during behavioural adaptation.

Beyond WSLS: reinforcement learning models

WSLS strategy is a simple heuristic to maximize gain in a *deterministic* environment, such as situation that occurs when the same kind of pasta sold in the same supermarket always tastes the same. However it would be problematic if the same strategy were applied to *probabilistic* settings, such as home-made pastas in restaurants where the taste of a given type of pasta might vary between chefs – sometimes there might be a better chef and at other times there might be a worse chef. Employing a WSLS strategy in these situations will result in inappropriately frequent choice switching. A single probabilistic poor outcome that occurs even when choosing an option that has been good on many other occasions will be sufficient for the agent to forgo that option on the next occasion it is offered. Hence, in probabilistic environments, it is important for one to build up an expectation of *how often* an outcome is observed over time and to use the expectation to guide decisions. Such an expectation corresponds to an estimated *value* of the option that lies somewhere between zero (not good at all) and one (always good) and that estimate should not be fully overwritten on the basis of a single outcome occurrence. Instead it should be only partially updated by any single observation. For example, the chicken pasta in the restaurant has long been your favourite dish, although occasionally it does not taste as good (e.g. the good chef is on holiday once a year), your expectation of the chicken pasta should not be totally updated by those infrequent experiences.

Many computational models of reinforcement learning (or "machine learning") have been devised to account for value updating in uncertain environments. The classical Rescorla-Wagner model (Rescorla and Wagner, 1972) states that

$$V_{t+1} = V_t + \alpha(r_t - V_t).$$

The most critical term (r_t-V_t), often known as the reward prediction error (RPE) term, describes the discrepancy between the outcome r_t (the tasting experience of the pasta) and the expected value V_t (the expectation of taste of the pasta that the diner had had prior to eating) at trial *t*. The RPE term is fundamental to many computational models because it tells the model when to learn (whether or not the observation is different from the expectation) and how much to learn so that the new expectation that is formed, V_{t+1}, is appropriate. As mentioned, the expectation should not be fully determined by a single outcome. In the model, the learning rate α is a parameter that lies between 0 and 1 that describes how much the new value V_{t+1} is updated by the RPE. For example, if the good chef has a fixed number of days-off in each week throughout a year, then one should hold a smaller learning rate (any single probabilistic outcome is not very informative for the change in the expected quality of the food) and the expected value for the pasta is proportional to the number of days-off. However, if the number of weekly days-off is random, it is more important to have a more recency-weighted learning and a larger learning rate will ensure quick behavioural change can be made in response to the fast changing average quality of the pasta in every week.

RPE and the dopamine system

The dopamine system plays a crucial role in encoding RPE. Numerous studies have shown that the firing rate of midbrain dopamine neurons is increased above baseline when an

unpredicted reward is received, decreases below baseline when an expected reward is omitted but, critically, is unmodulated when a fully expected reward is received (Mirenowicz and Schultz, 1996; Schultz, 1997; Hollerman and Schultz, 1998; Cohen et al., 2012; Schultz, 2012). Recent advances in optogenetics in rats have provided further evidence for the causal relationship between the RPE code in dopamine neurons and learning. Steinberg and colleagues (2013) used optogenetic tools to specifically activate dopamine neurons in the midbrain ventral tegmental area (VTA) during various learning tasks. The rats were trained to make associations between a cue and a reward and subsequently performed an extinction test in which the reward was consistently omitted to investigate the strength of the association that had been learned. Normally dopamine neuron activity will be inhibited at the time of reward omission during extinction, consistent with the negative RPEs that occur at that timepoint. However, when the inhibition of dopamine neuronal firing due to negative RPEs during extinction was partly counteracted by optogenetic stimulation of the dopamine neurons then the speed with which the cue-reward association was lost was also slowed down. This provides evidence for a causal role of midbrain dopamine neurons in value updating.

Axiomatic approach in studying RPE

One problem for investigating dopamine RPE signals is that RPEs could be easily confounded by other related variables. For example, RPE is often correlated with valance or salience of the outcome. Caplin and Dean suggested a dopamine RPE model that contains axioms to allow systematic falsification of RPE signals in the brain (Caplin and Dean, 2008b, a). The foundation of the model is based on the idea that RPE is a linear subtraction between reward magnitude r and expected value V ($RPE = r - V$). In other words, RPE signals are influenced independently by the reward size and the prediction of the reward.

There are three axioms in the axiomatic RPE model. Axiom 1 is related to the reward magnitude *r* term and states that the rank order of RPE signals for different magnitudes of outcomes should be independent of the expectation. If the signals for outcome A and outcome B have a positive difference in size, the positive signal size difference should hold for all levels of expectation. Axiom 2, by contrast, is related to the expected value *V* and states that the rank order of RPE signals for different expectations should be independent of the reward received. If the signals for expectation X and expectation Y have a positive difference in size, the positive signal size difference should hold for all levels of outcome. Lastly, when reward magnitude is equal to the expectation, the RPE term becomes zero. Hence, axiom 3 states that, for any levels of reward magnitudes and expected values, all fully predicted outcomes should generate the same size of signal.

Rutledge and colleagues investigated BOLD activity in brain regions that receive projections from midbrain dopaminergic neurons while human subjects were choosing between two options with probabilities and magnitudes of a monetary gain or loss that were explicitly shown (Rutledge et al., 2010). By testing against the dopamine RPE model, they found that the striatal regions (nucleus accumbens, caudate and putamen), medial PFC (areas 10 and 32), amygdala and posterior cingulate meet the requirements of the three axioms. However, one limitation of this study is that activity was measured in the form of the BOLD signal, which is not a direct measure of the actual dopamine activity. Recently, Hart and colleagues (2014) measured dopamine levels at the nucleus accumbens core region of the striatum by using fast-scan cyclic voltammetry while rats were choosing between two levers associated with a high probability and a low probability of receiving a large (as opposed to a small) reward. As with the human BOLD data, the changes in dopamine levels that occur in response to different expectations and outcomes are consistent with the dopamine RPE model. For both levels of probabilities, larger changes in dopamine were

related to the reception of a larger reward as well as an expectation of a smaller reward. There was no change in dopamine level in cases where the small and large rewards were fully expected. This provides direct evidence for the involvement of dopamine in encoding reward prediction error.

Alternative reinforcement learning models

The classical Rescorla-Wagner model is a simple model that involves value updating by the use of RPE. One weakness of the Rescorla-Wagner model is that the accounts of learning it provides are couched in terms of changes that occur on a "trial-by-trial" basis. However, there are no clear boundaries between "trials" in many real life situations. By contrast the temporal difference model has an emphasis on the expectations of reward that are updated on a time point by time point basis across the course of actions that eventually lead to a reward (Sutton and Barto, 1998). According to this view, for example, the value update is happening once the pasta is ordered until the time point when the food is consumed. The SARSA (state-action-reward-state-action) learner takes another step further to specify the importance for the conjunction between a state or context and an action that mediates the recipient of a reward (Sutton and Barto, 1998). For example the taste of the pasta (reward) is assigned to the action of ordering the pasta under the state of being in that particular restaurant.

As mentioned previously, the learning rate has to be flexibly adjusted according to the demands of the current context. One factor related to learning rate adjustment is the volatility or instability of the environment. In an environment in which the reward contingencies of options are stable, it is more important to have a low learning rate and to integrate a longer span of reward history in order to make the best estimate of current

reward values. In contrast, in a volatile environment it is more important to have a higher learning rate so that value estimate is more strongly influenced by recent outcomes. A class of reinforcement learning models implements the Bayesian approach to estimate the likelihood of changes in volatility in the environment (Behrens et al., 2007; Nassar et al., 2010), so that learners can adjust their learning rates according to their beliefs about the environment's current volatility. It is possible that environmental stability can be tracked in the brain. Volatility-related modulations of activity can be found in the anterior cingulate cortex (ACC) when subjects are performing a reward-guided decision making task (Behrens et al., 2007). Pupil diameter is often used as an index of autonomic arousal and also as a proxy for noradrenergic activity in the locus coeruleus (though note that this link is indirect and debated). It also shows brief changes when instability is detected in a task in which subjects are required to make predictions about a series of numerical outcomes (Nassar et al., 2012).

Non-contingent learning

Most studies have focused on the type of learning that occurs when an unexpected outcome or RPE is attributed to the stimulus or action that caused it (instrumental learning) or was paired with it (Pavlovian learning). In other words, the credit for an outcome is assigned in to the prior event on which it was contingent. However, humans and animals also learn in a different and more imprecise way in which outcomes are linked to choices in a non-contingent manner. In an early study in which he reported evidence for the law of effect Thorndike asked subjects to learn arbitrary matches between numbers and words (Thorndike, 1933). Intriguingly, he reported that the reward not only strengthened the word-number association that subjects employed on a given trial, but the impact of the reward also "spread" to the word-number pairs from the preceding and following trials that

occurred in close temporal proximity. Similarly, Walton and colleagues (Walton et al., 2010) also found non-contingent learning effects in monkeys when they were learning associations between visual stimuli and probabilistic rewards. They found that when a reward occurs it is linked in a contingent fashion to the choice that led to its occurrence. At the same time, however, the reward is also linked to preceding and subsequence choices. In other words, the credit for the outcome is misassigned to choices on which it was not really contingent. Although the neural correlates for non-contingent learning remain largely unknown, it is known that the OFC is necessary for contingent learning. Walton and colleagues (2010) reported that OFC-lesion monkeys had a specific impairment in learning precise stimulus-reward associations. However, their ability to link stimuli with reward through non-contingent learning mechanisms was comparable to the controls' ability.

In the brain, neuronal populations that encode reward have multiple time constants for holding the reward memory (Seo and Lee, 2008; Bernacchia et al., 2011; Wittmann et al., submitted). Such populations hold information about not just whether reward was received just now but whether it was received recently over a variety of different time scales. Perhaps such neural activity patterns contribute to non-contingent reward learning; not only do they contain information about the reward that has been received just now but also about rewards received on previous trials but at the same time they may lack information about the specific stimulus-reward association that led to the reward. Neurons with short time constants could be important for contingent learning, in which reward memories (but not the stimulus-reward association memory) degrade quickly after the credit is assigned to the relevant stimuli. Neurons with long time constants have more sustained reward memories so that the credit for the rewards they represent, which occurred in the past, could be misassigned to stimuli that are chosen subsequently. In fact, various brain regions contain neurons that exhibit different speeds of decay in their reward memory traces, including the dorsal ACC, dorsolateral prefrontal cortex (dlPFC) and LIP (Bernacchia et al., 2011). In

Chapter 5, I report an analysis that investigated the interaction between the lOFC and amygdala and suggest a mechanism that influences non-contingent learning.

1.3 Ventral Prefrontal Cortex

Anatomy

The vmPFC and OFC are adjacent structures in the frontal lobe (**Figure 1**). As there is no clear boundary between these two regions, the names are sometimes used interchangeably in different studies; it often needs care to review literature related to these regions. To make it even more challenging, it should be noted that when describing the subregions of the vmPFC and OFC, the terminologies are again not always consistent between studies – the vmPFC is sometimes known as the medial OFC, the central OFC (cOFC) is sometimes known as the lOFC, the lOFC is sometimes known as the ventrolateral PFC. On the orbital surface of both the human and macaque brain, the medial orbital sulcus (MOS) and the lateral orbital sulcus (LOS) are useful landmarks that separate the OFC into subdivisions (Carmichael and Price, 1994, 1995b; Ongur and Price, 2000). In this thesis, the vmPFC refers to the ventromedial region that lies on the medial wall and extends to areas 14 and 11m that is on the gyrus rectus medial to the MOS on the orbital surface of the frontal lobe. The cOFC refers to areas 11 and 13 that lie between the MOS and LOS. The lOFC refers to the area lateral to the LOS that includes the orbital region of area 12/47.

The vmPFC and OFC have extensive connections with many other brain regions. Price and colleagues (Carmichael and Price, 1994, 1995b, a, 1996; Ongur and Price, 2000; Kondo et al., 2005) summarized the connectivity in terms of two networks. The "orbital network" that mainly consists of the cOFC and lOFC has connections with sensory regions. The "medial network" that mainly consists of the vmPFC has connections with visceromotor

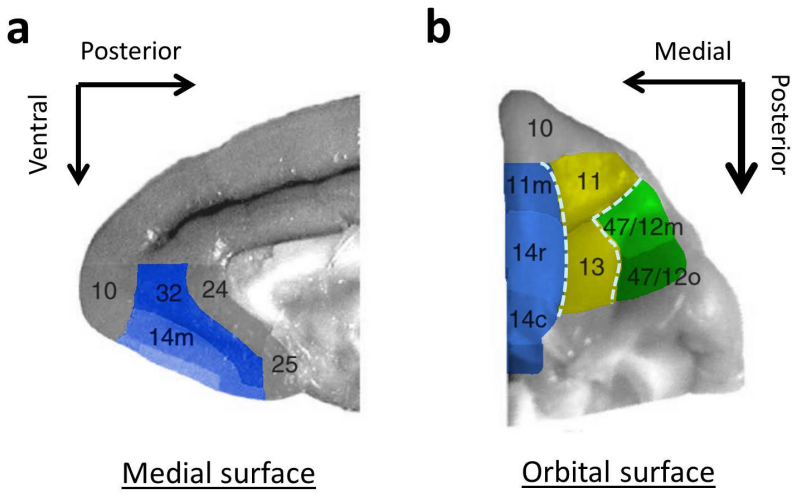


Figure 1. *(a) Medial and (b) orbital views of the macaque frontal cortex. The ventromedial prefrontal cortex (blue) extends from the medial surface to areas 14 and 11m. The central orbitofrontal cortex (yellow) consists of areas 11 and 13. The boundaries are marked by the medial orbital sulcus (left dotted line in* ***b****) and the lateral orbital sulcus (right dotted line in* ***b****). The lateral orbitofrontal cortex (green) consists of areas 47/12m and 47/12o. Adapted from (Mackey and Petrides, 2010).*

and limbic regions such as parts of the hypothalamus and hippocampus. Learning and decision making often require integration of information from multiple sources, often with a high degree of abstraction, as well as coordination with effectors to implement decisions in the form of actions. Arguably, it is the unique connectivity profile of the vmPFC and OFC that makes them ideal regions for decision making and reinforcement learning.

Somatic marker hypothesis

Perhaps one of the earliest records about the functional role of vmPFC and OFC is the case of Phineas Gage, a railway construction worker who suffered damage in these regions

caused by an iron bar driven through his head during an accident in Vermont, United States in 1848 (Damasio, 1994; Damasio et al., 1994). The case is famous because Gage survived the accident but, according to some accounts, he showed dramatic personality changes. Before the accident, he was described as a responsible, efficient, and instrumental person. The iron bar that took away most of his vmPFC and OFC turned him into a disorganised, impatient, and stubborn person.

An early theory that describes the function of vmPFC in decision making is the somatic marker hypothesis (SMH) proposed by Damasio and colleagues (Damasio, 1994; Damasio, 1996; Bechara et al., 2000). In the SMH, the vmPFC (and perhaps most of the OFC in this theory) serves as a substrate to hold the associations between an experience and an emotional state. The associations are known as somatic markers. When an agent encounters a situation that contains elements that are similar to previous experience, the somatic markers in the vmPFC are activated in order to reconstruct the set of emotional states previously linked to that experience. For example, seeing a glass of orange juice in a restaurant could trigger the association between the orange juice and the pleasurable feeling of drinking orange juice in many previous scenarios. The emotional state reconstructed serves as an incentive signal or a bias signal and sets the basis for an agent to decide whether to select or avoid an option.

The SMH was developed based on observations in patients that often, if not always, had vmPFC lesion that extended to the OFC (Damasio et al., 1990; Bechara et al., 1994; Tranel et al., 1995; Bechara et al., 1996). It is not clear in the SMH whether there is a functional distinction between the vmPFC and OFC. However, the SMH highlights two ideas that are critical in this thesis. First, the vmPFC or/and OFC are crucial for assessing the value of options and second, either or both regions have an important role in learning associations between stimuli and rewards. In the following, I will review how value is encoded and

compared in the vmPFC, how stimulus-reward association is learnt in the cOFC and how behavioural adaptation is made in the lOFC.

The subjective nature of the vmPFC value signal

The vmPFC encodes a signal that is related to the value of options during decision making (Knutson et al., 2005; Hampton et al., 2006; Kable and Glimcher, 2007; Plassmann et al., 2007; Boorman et al., 2009; FitzGerald et al., 2009). The results from various lines of studies suggest that such a value signal is subjective in nature.

A particularly clear and direct demonstration of the role that vmPFC has in value representation was provided by Plassmann et al. (2007). They took care to estimate the subjective value of a series of food items by asking each of their participants to engage in a Becker-DeGroot-Marschak auction (this procedure is thought to elicit subjects' true valuations of a choice option) and then used fMRI to scan subjects while they made choices about the same items. Activity on the ventromedial surface of the vmPFC covaried with the estimates of each participant's valuation of the items. The same activity pattern was absent when the participants made similar choices but as a result of instruction by the experimenter. Value signals might not, however, be just a function of how much a person likes the taste of a food. Hare and colleagues (Hare et al., 2009; Hare et al., 2011) investigated how self-control in dietary choices could have an impact on the vmPFC value signal. Human subjects were presented with different food items and were asked how much they would like to consume the food. They also needed to rate how healthy and how tasty the food items were. The fMRI results show that the vmPFC activity was modulated as a function of both the health and taste ratings, suggesting that the value signal in the vmPFC integrates both the primary reward value and the prospective health value that is associated

with the food item. Subjects with stronger impacts of health on food choices also showed stronger impacts of health on the vmPFC signal. Interestingly, the authors found that the dlPFC that is important for cognitive control was coupled with vmPFC as a function of decisions based on food healthiness. It could be the case that the proportion of value contributed by cognitive control to choose healthier food is passed on from the dlPFC to the vmPFC for computation of an integrated value.

The value signal in the vmPFC also reflects subjective preferences when people are making intertemporal decisions. When people are offered a future monetary gain, the same reward is more discounted by the time delay that will elapse before the money is actually received (Loewenstein, 1988; Green and Myerson, 2004). In other words, people subjectively value an immediate reward more than a future reward even though the reward size is the same for both cases. When vmPFC activity is examined, it is best explained by the time-discounted value of the monetary reward, instead of the actual amount of the reward (Kable and Glimcher, 2007). One important note is that the temporal discounting factor varies subject-by-subject. Some people are more impatient of delayed reward while other people have more tolerance of delays. The time-discounted value signal in vmPFC is subjective because it is dependent on the discounting factor that is specific to the subject.

Reinforcer devaluation is another way of manipulating the subjective value of a food reward. Often it involves experimental subjects being fed one kind of food item to satiety. The subjective value they assign to the food item should be reduced when they are asked to consume the food immediately after the satiety procedure (Murray and Izquierdo, 2007; Schoenbaum et al., 2009). The vmPFC value signal is also reduced when the reward that subjects are considering has previously been devalued (Valentin et al., 2007; de Wit et al., 2009).

Taken together, these studies suggest that the vmPFC encodes a subjective value signal that emerges after integrating both external features (such as reward magnitudes and probabilities) and internal states (such as preference for food healthiness, impatience to delayed rewards and satiety) that are relevant for making a decision.

Value comparison in the vmPFC

One important question is how a decision emerges in the vmPFC or whether the value signals are passed on to some other distinct brain area that actually brings about a decision. Race models and competitive models are two potential classes of models that could describe the neural mechanism for making a binary decision. Race models normally include two neuronal populations that each receives input independently from one of the two options (Logan, 2002; Reddi et al., 2003; Gold and Shadlen, 2007). A decision is made once the firing rate of one of the populations reaches the threshold that sets the decision bound. Competitive models mainly differ from races models by having populations that mutually inhibit each other directly or indirectly in order to effect a value comparison process. The cortical attractor network model (Wang, 2002) and the LCA model (Usher and McClelland, 2004) described in Section 1.1 are two examples of these models.

Hunt and colleagues (2012) simulated two-option decisions by using the cortical attractor network model (Wang, 2002) and made predictions about the signatures of a neural network that works under mutual inhibition. They also had human subjects performing in a two-option decision making task while undergoing magnetoencephalographic (MEG) recording. Interestingly, they found that the activity in the vmPFC matched with the competitive model prediction.

In most competitive models, an excitatory population that receives input from one option should exhibit activity that correlates with the value of that option and, importantly, its activity should also inversely correlate with the value of an alternative option due to mutual inhibition during the competition process. A recent study recorded neuronal activity in the vmPFC while monkeys made decisions between two sequentially presented options (Strait et al., 2014). A population of vmPFC neurons showed anticorrelation between the value of the first and the second options – the firing of the neuron that was positively correlated to the value of the first option was negatively correlated to the value of the second option. This provides neurophysiological evidence that the vmPFC works through a mutual inhibitory mechanism for value comparison.

Similar evidence for a value comparison process in the vmPFC was also found in many human fMRI experiments (Boorman et al., 2009; FitzGerald et al., 2009; Lim et al., 2011; Boorman et al., 2013; De Martino et al., 2013). Many of the value comparison fMRI signals that have been reported involve BOLD activity that varies positively as a function of the value of one option and which varies negatively as a function of the value of another option. In other words, the brain region is encoding the value difference between two options. Although exactly how the positive and negative components are contextualized in the value comparison signal is debateable, several forms of vmPFC value difference signal have been reported – value difference between chosen and unchosen options (Boorman et al., 2009; Kolling et al., 2012; Boorman et al., 2013), higher value and lower value options (FitzGerald et al., 2009), attended and unattended options (Lim et al., 2011).

Nonetheless, it is unclear what the exact role of vmPFC is during decision making. It is without doubt the case that vmPFC is critical for many decision making scenarios but the vmPFC is not essential for all kinds of value-based decisions; despite personality change Gage was still mostly a functional person after losing a huge portion of the vmPFC and OFC.

Monkeys with vmPFC/mOFC lesion can perform normally in easy decisions when there is an obvious best option but are only impaired at choosing the best option in difficult decisions when the alternatives are nearly as good as the best option (Noonan et al., 2010).

When people encounter binary decisions that involve two decision attributes, the vmPFC is particularly active in cases when it is critical to integrate the two attributes for making a decision, but it is relatively inactive when one option is clearly better than the other by having an advantage in both attributes (Hunt et al., 2012). When patients with ventromedial frontal lobe lesions are choosing between multiple apartments that are each defined in terms of multiple component attributes, they tend to use an integrated choice-based evaluation strategy (they evaluate the multiple attributes of one apartment and then switch to evaluating another apartment albeit again in terms of all of its component parts) as opposed to healthy controls that tend to use an attribute-based strategy (they compare the same attributes across all apartments and then move to consider another attribute, again across all apartment options) (Fellows, 2006). It is possible that there are multiple decision making mechanisms operating in parallel in the brain so that it is not necessary for the vmPFC to be engaged in all kinds of decisions (Kolling et al., 2012; Rushworth et al., 2012). The vmPFC could have a specific role in making more precise, analytical and goal-directed decisions.

Reward expectation in the cOFC

The cOFC has a connectivity profile that makes it an important region for reward guided learning and decision making. On the one hand, it is connected to the sensory regions associated with the different modalities, such as gustation, olfaction and somatosensation (Carmichael and Price, 1995b, 1996). On the other hand, it is also connected to the ventral

striatum that is important for signalling reward and prediction error (Haber et al., 1995; Haber et al., 2006). At such an intersection between sensation and reward, the cOFC has a critical role in drawing associations between stimulus and outcome that is important for value updating (Rolls et al., 1996; Roberts, 2006; Murray and Izquierdo, 2007; Noonan et al., 2010; Walton et al., 2010; Noonan et al., 2011; Noonan et al., 2012).

The cOFC is crucial for signalling the specific features of a reward associated with a stimulus. When monkeys are taught associations between objects and the reward outcomes that follow, some cOFC neurons respond to signal the value related to a specific type of outcome (Thorpe et al., 1983; Tremblay and Schultz, 1999; Padoa-Schioppa and Assad, 2006, 2008). For example, if object A and B are associated with orange and apple juice respectively and the size of each object presented on any given trial is manipulated so as to indicate the amount of juice that might be obtained by choosing the object, some cOFC neurons fire as a function of the value (the amount of juice associated) with object A while others fire as a function of the value of object B (Padoa-Schioppa and Assad, 2006, 2008). It is noteworthy that the neuronal response is independent of visuospatial features of the stimuli – the same response can be elicited by the same object presented at a different position or a different object that leads to an identical outcome. The neuronal response is also independent of the motor response that is made to select the object.

The notion that cOFC is important for learning stimulus-reward association fits well with the findings in studies that use reinforcement devaluation tasks. This kind of task usually involves subjects choosing between two objects that are associated with two kinds of rewards, such as a grape or a peanut. Subjects are tested before and after a devaluation session in which they are fed with one of the rewards to satiety and normally they demonstrate a preference change to avoid the sated reward, known as a devaluation effect. Typically, such a devaluation effect is weaker in monkeys with bilateral cOFC lesions

(Izquierdo et al., 2004; Murray and Izquierdo, 2007; Rudebeck and Murray, 2011b, a) and monkeys with smaller cOFCs (Burke et al., 2014). As mentioned above, devalued reinforcers are also associated with reduced BOLD signals compared to non-devalued reinforcers in the cOFC in normal humans (Valentin et al., 2007). Finally, the reinforcer devaluation effect also requires interaction between the cOFC and amygdala. Monkeys that had unilateral lesion of cOFC and amygdala in the same hemisphere, meaning that these two structures were still able to communicate in the intact hemisphere, showed a devaluation effect similar to controls (Baxter et al., 2000; Murray and Izquierdo, 2007). However, the devaluation effects were weaker in monkeys if the unilateral lesions in the cOFC and amygdala are made in different hemispheres, such that the cOFC-amygdala connection is bilaterally disrupted. Taken together, the cOFC and its interaction with the amygdala are important for updating the value of a previously established stimulus-reward association so that objects related to a sated reinforcer can be devalued.

It has been argued that not only does the cOFC update the value of a stimulus-reward association when the reward itself is devalued but that the same region may also update the association if the outcome associated with a stimulus is changed. For example, in a reversal task with a Pavlovian design, cOFC neurons respond rapidly to the presence of a stimulus that leads to delivery of blackcurrant juice, however after a reversal, the firing of the same neurons is less rapid in the presence of the same stimulus that now leads to an aversive solution (Thorpe et al., 1983). Here, an important question is whether this signal that reflects some aspect of the outcome expectation (such as reward value and/or identity) associated with a stimulus could be used to guide behavioural change in the same region.

A key test of whether cOFC signals guide behavioural change is whether lesions here disrupt discrimination reversal learning. Demonstrations that even excitotoxic lesions in the OFC of the marmoset disrupt discrimination reversal learning (Dias et al., 1996b, 1997; Clark

et al., 2008; Rygula et al., 2010) suggest a critical role for some part of the OFC in bringing about a change in choice strategy in the face of changing environmental contingencies. However, as discussed above (Section 1.2 Reinforcement Learning), whether it is the cOFC region of the OFC in old world monkeys that is the critical region has recently become unclear because the view that the cOFC is important for object reversal learning has been challenged (Kazama and Bachevalier, 2009; Rudebeck et al., 2013b).

Careful consideration of other aspects of the neural activity in cOFC raise doubts about whether it is concerned with every aspect of reward-guided choice. For example, in the study by Padoa-Schoppia and Assad (2006), the authors reported a population of cOFC neurons that responded as a function of the value of any option that is chosen, in contrast to another population in the same region that respond to the individual options. However, it is particularly unlikely that the cOFC is involved in the kind of value comparison process performed in the vmPFC. Models that involve mutual inhibition for value comparison predict that neurons that receive input from one option not only fire as a function of the value of that option, but also fire as a function of the *inverse* value of the alternative option due to mutual inhibition (Busemeyer and Townsend, 1993; Usher and McClelland, 2001; Wang, 2002; Ratcliff and McKoon, 2008) (See Section 1.1 Decision making for details). This is the pattern that was recently demonstrated in the study that made recordings in vmPFC neurons (Strait et al., 2014). By contrast, the firing of cOFC neurons in response to an option is not modulated by the presence of another option that is presented at the same time (Padoa-Schioppa and Assad, 2008). In other words, the firing of these neurons exhibits a menu invariant property suggesting that they contain information about the value of a particular kind of reward, rather than participating directly in the value comparison process.

Moreover, unlike monkeys with medial OFC/vmPFC lesions, monkeys with lesions that included lOFC and parts of cOFC lesion were not impaired in making reward-guided

decisions as a function of decision difficulty once they had learned the relevant stimulus-reward associations (Noonan et al., 2010). They do, however, experience difficulties in the initial learning of stimulus-reward contingencies (Noonan et al., 2010; Walton et al., 2010). Such results also argue that the reward features coded in the cOFC may be passed on to a different region, such as vmPFC, to guide the actual decision-making process.

In summary, reward expectation signals are prominent in OFC and it is clear that the cOFC region is important when animals make choices about which rewards to pursue given their current state of motivation and satiety. The most medial parts of the OFC, mOFC, and the vmPFC may play a special role in comparing the values of potential choices. Despite a long history of investigations focusing on the OFC and behavioural reversal there is currently controversy about the role played by the cOFC region in old world monkeys, on which most research focuses, in bringing about changes in behaviour in response to changes in the reward environment. Despite this controversy, evidence from new world monkeys suggests that some part of the OFC may be important for some aspect of reward-guided behavioural change.

Cognitive inhibition and adaptation in the lOFC

In the lOFC, area 12o has a special connectivity profile. It has abundant connections with both the orbital network, which has access to multiple sensory inputs in regions such as the anterior temporal lobe and perirhinal cortex (Carmichael and Price, 1995b; Ongur and Price, 2000; Kondo et al., 2005), and the medial prefrontal networks, which connects to visceromotor and limbic regions such as the hypothalamus and hippocampus and which is composed mainly of the vmPFC (Carmichael and Price, 1995a, 1996; Ongur et al., 1998; Ongur and Price, 2000). This perhaps puts area 12o in the lOFC an ideal position to serve as a region for behavioural adaptation, a process that requires integration of updated values in response to environmental change and implementation of decisions into actions.

There is a handful of data from old world monkeys, such as macaques, that investigate the role of lOFC in behavioural adaptation. Typically these studies have focused on the general region of area 12/47 because it is not possible to identify the precise location of the 12o subdivision without a full cytoarchitectonic analysis. While lesion and neurophysiological studies that purport to investigate OFC in macaques are often focused on area 11/13 in the cOFC, studies that investigate the function of the ventrolateral PFC or inferior prefrontal convexity usually involve lesions or neuronal recordings in regions that are dorsal to the lOFC and ventral to the principle sulcus, such as the lateral surface of area 12/47, area 45a and area 9/46v (Rushworth et al., 1997b; Romo et al., 1999; Rushworth et al., 2005; Kennerley and Wallis, 2009). Thus, both types of investigation tend to overlook lOFC regions such as 12o. Iversen and Mishkin (1970) conducted one of the few studies that investigated behaviour of monkeys with lOFC lesions. While performing in an auditory discrimination task that involved a response to a reward related tone and no response to a non-reward related tone, these monkeys made more errors and also more task-inappropriate responses during the inter-trial intervals. In an object reversal task, the same lOFC monkeys as well as monkeys with lesions in the cOFC (area 11/13) that extended into the orbital surface of vmPFC (area 14) initially made more mistakes than the controls. In particular, the lOFC monkeys' deficits in reversal learning were characterized by more perseverative errors, whereas the mistakes of the 11/13/14 monkeys were mainly a consequence of making stochastic choices (possibly contributed by the lesion in area 14 given what we now known about this region from research published in the last few years).

It is possible that one critical function of the lOFC is to inhibit any context irrelevant behaviour, such as responses during the inter-trial interval, or to forgo a previously rewarded option, and to select context-relevant behaviour, such as choosing an option that although previously unrewarded is now rewarded (Roberts and Wallis, 2000; Man et al., 2009). However, there is uncertainty about the conditions under which response

suppression impairments are manifest (Chudasama et al., 2013), whether they can or should be linked to any pattern of regional variation in the responsiveness of OFC neurons to negative outcomes, including mild aversive events, such as airpuffs, and loss of positively reinforcing events (such patterns have proved difficult to identify) (Morrison and Salzman, 2009; Rich and Wallis, 2014) or to the role of ventrolateral prefrontal regions in attentional selection, switching, and/or response suppression (Dias et al., 1996a; Monchi et al., 2001; Arana et al., 2003; Hampshire and Owen, 2006; Ghahremani et al., 2010; Rygula et al., 2010; Hampshire et al., 2012).

To address some of these issues, it would therefore be useful to examine the whole extent of the OFC in an old world monkey in order to identify the critical region for behavioural adaptation in response to environmental change and even to ascertain whether such a region exists. This might not be done easily by conventional methods used in monkey experiments, such as lesion and neurophysiological recording. In Chapter 5, I report an fMRI experiment with awake macaques performing in an object discrimination reversal task. I identified a lOFC region in the orbital surface of area 12/47 that is active when the monkeys used the outcome of a choice to generate subsequent adaptive behaviour.

1.4 Posterior Parietal Cortex

Anatomy

The posterior parietal cortex (PPC) is an association region that contains rich sensory and motor information (Grefkes and Fink, 2005; Sereno and Huang, 2014). It contains the intraparietal sulcus (IPS) as an important landmark that divides the PPC into the superior parietal lobule (SPL) and inferior parietal lobule (IPL). This thesis has a strong focus on two divisions, the lateral intraparietal area (LIP) and medial intraparietal area (MIP).

The LIP in macaque is located in the lateral bank of the IPS. The human LIP, in contrast, is located more medially in the IPS perhaps due to the expansion of the IPL and dorsal visual stream during evolution (Grefkes and Fink, 2005; Mars et al., 2011). The LIP receives projection from visual sensory areas in the occipital and parietal lobes and is interconnected to eye movement areas such as the frontal eye field and the superior colliculus (Lynch et al., 1985; Blatt et al., 1990; Felleman and Van Essen, 1991; Grefkes and Fink, 2005). Due to its unique connectivity profile, the LIP has been the locus of many neurophysiological studies in which macaques are tested using tasks that require detection of visual stimuli and responses that are made by eye movements.

The MIP is located in the medial bank of the IPS in both monkeys and humans. Unlike the LIP, the MIP does not have rich interconnections with visual or eye movement regions. It has projection to the dorsal premotor area and is crucial for generating hand reaching movements (Felleman and Van Essen, 1991; Johnson et al., 1996; Cohen and Andersen, 2002).

An important feature of the PPC is that it contains retinotopic and somatotopic maps for the contralateral space (Goldberg et al., 1990; Sereno et al., 2001; Silver and Kastner, 2009; Sereno and Huang, 2014). For example, a given neuron in the left LIP might be sensitive to objects presented within its response field which is likely to be in the right visual field or a given left MIP neuron might be active when a hand movement is executed so as to reach a similar location (Rushworth et al., 1997a; Sereno et al., 2001; Sereno and Huang, 2014). Numerous neurophysiological studies in reward-guided decision making make use of this property by presenting objects that are associated with different reward values inside the receptive fields of macaque PPC neurons to investigate how signals related to value and to the process of evidence accumulation about where to look next emerge in

this area (Shadlen and Newsome, 1996; Platt and Glimcher, 1999; Shadlen and Newsome, 2001; Dorris and Glimcher, 2004; Kable and Glimcher, 2009; Seo et al., 2009).

Decision making

The PPC is important for decision making, in particular the LIP region has been the focus of many neurophysiological studies in macaques. The early study of Shadlen and Newsome involved monkeys performing in a random dot motion (RDM) task (see Section 1.1 Decision Making for details of the task), while the activity of LIP neurons was being recorded (Shadlen and Newsome, 1996). The LIP was selected as a region of interest for oculomotor tasks widely used in macaques because it connects directly to areas MT and MST that are sensitive to motion and to the FEF and superior colliculus that are related to eye movements (Maunsell and Van Essen, 1983; Lynch et al., 1985; Blatt et al., 1990; Felleman and Van Essen, 1991; Grefkes and Fink, 2005; Lee and Maunsell, 2010; Kolster et al., 2014). Shadlen and Newsome showed that the firing of LIP neurons ramps up gradually when an animal is about to make a saccade towards the target located within the response field of these neurons (Shadlen and Newsome, 1996). Critically, this pattern of firing is a truly decision-related signal, not just a motion perception signal, because the same choice-related activity can be seen even when animals are making decisions from fully random dot movements in the absence of coherent dots.

Not only do LIP neurons participate in perceptual decisions such as those based on dot movements, they are also involved in value-guided decision making. Platt and Glimcher trained monkeys to choose between visual cues that led to juice reward (Platt and Glimcher, 1999). The amount of juice and probability of juice delivery associated with the visual cues were varied across blocks so that the value of choices was systematically manipulated. They

found that the firing of LIP neurons is modulated as a function of the value of the choices presented in their response field.

Neurophysiology and neuroimaging studies in monkeys and humans show that the PPC value signal is reflective of the subjective preference of an agent (Sugrue et al., 2004; Kable and Glimcher, 2009). In making intertemporal choices, some animals have more tolerance of a delay of reward delivery, while other more impatient animals discount the value of a reward that is delayed more steeply. LIP neurons in monkeys fire more rapidly as a function of the subjective discounted values of the stimuli within their response field. In other words, for any given amount of reward, the firing of LIP neurons is more attenuated in monkeys with less tolerance of delays (Louie and Glimcher, 2010). In humans, fMRI signals in the PPC are also related to delay discounted values (Kable and Glimcher, 2007; Ballard and Knutson, 2009; Peters and Buchel, 2009) and disruption of the right PPC is also associated with stronger preferences for immediate rewards as opposed to delayed rewards (Essex et al., 2012).

Multiple alternative decisions

One useful way of studying the properties of decision signals in the PPC is by having subjects participate in multiple alternative decisions. Churchland and colleagues trained monkeys to perform in a RDM task that involved two or four possible directions of coherent dot movement (Churchland et al., 2008). In both trials with two and four options, the activity of LIP neurons ramps up more rapidly before a decision when there is a stronger coherence of dot movements favouring the subsequently chosen direction that lies within the response field of the neurons. Interestingly, by comparing LIP activity between the two trial types, it was shown that LIP neurons exhibit a lower firing rate before and during the onset of

random dot presentation in four choice trials than in two choice trials. However, the LIP neurons reach a common maximum firing rate of around 90 Hz for both trial types right before a saccadic decision is made. This shows that in contexts when decisions are expected to be more complex, such as making a decision on four choice trials, the LIP can adjust the baseline firing but keep the same decision threshold so that more evidence is necessary before a decision is made.

Given their limited resources, neural systems often operate in a limited range of sensitivity. It is important for the brain to flexibly adjust the sensitive range in different contexts. As mentioned in Section 1.1 Decision Making, one mechanism that serves for this purpose is *divisive normalization*. Divisive normalization occurs when neurons fire more rapidly for stronger inputs from stimuli *inside* their response field but less rapidly for stronger total inputs from stimuli both *inside and outside* their response fields (Heeger, 1992; Reynolds and Heeger, 2009; Carandini and Heeger, 2012). In other words, the sensitivity of a neuron to stimuli within its response field is stronger in the context of lower total inputs from all stimuli. This means that even small changes in the strength of a stimulus in the neuron's response field will be more likely to be detected when that stimulus occurs in the context of other weak as opposed to strong stimuli. Evidence for such a mechanism is found in the neuronal activity of various visual areas, such as during contrast detection in the lateral geniculate nucleus (Bonin et al., 2005), orientation detection in the primary visual cortex (Heeger, 1992), and motion detection in area MT (Heeger et al., 1996; Shadlen et al., 1996). Since LIP is closely linked with various visual regions, it is possible that LIP also exhibits activity changes predicted by divisive normalization.

In order to study divisive normalization, a decision making task that allows manipulation of context is crucial. Most forms of the RDM task only involve manipulation of evidence favouring one perceptual decision – there are coherent dots moving towards one

direction at a time. The context is often stable because evidence for alternative options originates purely from noise. It is not possible to investigate whether the context has an impact on PPC signals in the manner that divisive normalization would have predicted. Louie and colleagues (2011), however, recorded LIP neuronal activity while presenting monkeys with stimuli associated with different amounts of juice reward inside and outside their response field. The context was varied by systematically manipulating the reward size associated with a stimulus and the number of stimuli outside the response field of the LIP neurons. As predicted by divisive normalization, the impact on LIP firing of a stimulus inside the response field is stronger when the total reward size of all stimuli outside the response field is smaller.

Decision making or attention?

Despite recent interest in studying its role of decision making, the PPC has been the focus of studies related to visual attention. Damage in the PPC is often associated with hemispatial neglect, in which patients have deficits in attending or exploring the contralesional side of space (Vallar, 1998; Husain and Rorden, 2003). For example, patients with lesions in the right PPC are significantly slower when they are asked to make a leftward hand movement to make a button press as opposed to a rightward hand movement (Mattingley et al., 1998). Malhotra and colleagues (2009) tested right PPC patients with different variants of a task in which they needed to make responses to stimuli appearing on five potential locations aligned vertically on the screen. While showing normal performance when they were asked to make a response when a target letter appeared (non-spatial task), the patients made more errors when the instruction was to only make a response when the stimuli appeared on two of the five potential locations (spatial task). This suggests that the PPC is critical for sustaining attention when spatial information is important.

The PPC is important for multiple forms of attention. Bays and colleagues examined the contribution of PPC to goal-related and stimulus-related oculomotor responses (Bays et al., 2010). They presented PPC lesion patients with left visual neglect with an 8×5 matrix of green squares, while their eye movements were recorded. On some trials, a letter that was isoluminant to the squares was presented on top of one of the squares; subjects had to make a goal-driven eye gaze to search for the letter and report whether the letter was the target letter X by making a button press in order to obtain a reward. On the other trials, the salience level of one of the squares was altered by a change in orientation, brightness or onset time. Although subjects did not need to make any manual response, they were expected to make a stimulus-driven eye gaze to detect the odd square. By examining the eye movements, it was established that the PPC patients with left visual neglect exhibited a gradual reduction in fixation when the target letter or the odd square appeared on the left rather than on the right position. This suggests that the PPC is important for both goal-directed and stimulus-directed attention.

It is argued that value and attention are two qualities that are easily correlated in decision making experiments. It could be possible that some value signals reported in the PPC are actually contributed by attention (Leathers and Olson, 2012). In decision making tasks, stimuli associated with higher values are often stronger in salience as well. One method to orthogonalize the value and salience associated with the options is to introduce stimuli associated with aversive outcomes, such as monetary loss or mild air puff to subjects' eyes. By doing so, stimuli with strong salience could either be highly rewarding (positive values) or aversive (negative values). In the human PPC, there is a distinction between a salience signal encoded in the inferior parietal lobule and a value signal encoded in the superior parietal lobule (Kahnt et al., 2014). The location of the salience signal matches with the clinical observation that hemispatial neglect is observed most often in patients with more inferior lesions in the PPC (Vallar, 1998; Husain and Rorden, 2003). It might seem

surprising that the value signal is found in the superior parietal lobule, whereas the LIP region that is important for decision making is located in the lateral bank of the intraparietal sulcal wall adjacent to the inferior parietal lobule. However, this is LIP's anatomical location in macaques. By contrast, as briefly mentioned above, comparative anatomical studies show that the human equivalent of the LIP region is actually found in the superior parietal lobule (Grefkes and Fink, 2005; Mars et al., 2011; Sereno and Huang, 2014). Together these results suggest the PPC contains divisions primarily subserving either attention or decision making.

1.5 The current thesis

Rational choice theory makes an axiomatic assumption that decisions are independent of irrelevant alternatives (Luce, 1959). The first question addressed by this thesis concerns how the presence of a third irrelevant option could cause suboptimal decision making (Chapters 2-4). In a changing environment, it is important for an agent to use reward information to guide adaptive decisions. The second question addressed by this thesis concerns how a win-stay/lose-shift (WSLS) strategy is encoded in a neural mechanism for generating flexible behaviour (Chapter 5).

In Chapter 2, I will test rational choice in multi-alternative decisions by extending the cortical attractor model for binary decisions (Wang, 2002) to accommodate a third irrelevant option. This adaptation enables the model to make predictions to how the presence of a third irrelevant option could have an impact on making suboptimal decisions between two other options.

In Chapter 3, I will examine whether human behaviour show any violations to the assumption of independence of irrelevant alternatives. A series of experiments will be reported that involve testing subjects with different variants of a decision making task

containing three alternatives. I will provide evidence that suboptimal decisions are made in the presence of a third irrelevant option. The results are compared with the predictions made by the biophysical model in Chapter 2 and by the divisive normalization model (Louie et al., 2011; Louie et al., 2013).

In Chapter 4, I will investigate the neural mechanism underlying suboptimal decisions with multiple alternatives. I will examine how the neural signatures related to decision making in the ventromedial prefrontal cortex (vmPFC) and posterior parietal cortex (PPC) are affected by the presence of the irrelevant option. I will also examine how the vmPFC and PPC signals are related to different aspects of behaviour.

In Chapter 5, I will examine the neural mechanism related to adaptive behavioural change with focus on the implementation of WSLS strategy. I will report an experiment that involves macaques performing in an object discrimination reversal task while undergoing functional magnetic resonance imaging. The different contributions of the orbitofrontal cortex, anterior cingulate cortex and amygdala to guide flexible behaviour will be investigated. The results will be discussed by comparing controversial findings in these regions in the literature.

Chapter 2. Suboptimal Decisions with Multiple Alternatives: Biophysical Model

Rational choice theory makes predictions that decisions between two options should not be influenced by the presence of a third irrelevant alternative. However, evidence has shown that this theory could be violated in many situations. In this study, we extended a biophysically plausible model that receives input from three options. By manipulating the value of a third irrelevant option, the model predicted that more suboptimal choices were made when the value of the third option was smaller; this contradicts the assumption of rational choice. The phenomenon is caused by overall weaker inhibitory activity in a decision-making neural network when the third option value was small, which leads to an increase in stochasticity in the network and fewer optimal choices. Comparable predictions were made when the third option was only presented transiently or for a time period equal in duration to the other two options.

2.1 Introduction

Traditional economics theories made assumptions that human makes rational decisions, known as the rational choice theory (von Neumann and Morgenstern, 1944). Under this theory, one important assumption is that people make decisions that are independent of irrelevant alternatives. In other words, when choices were made between two options, the decisions should not be affected by the presence of other distractors, such as a very bad option or an unchooseable option. However, the assumption of rational decision making has received criticisms. For example, it is violated by the phenomenon of the "decoy effect".

In the decoy effect the way in which people choose between two multi-attribute options is affected by the presence of an additional third option being present which is more similar to one or other of the two options on one dimension (Pettibone and Wedell, 2007). For example the way that people choose a lobster or a steak defined by two attributes such as price and taste is sometimes affected by the presence of a third alternative apartment that is similar to one of the original options in terms of one dimension (for example price). In the phenomenon of choice intransitivity, if a person has a preference for option A over B and also for option B over C they may still not exhibit a preference for option A over C (Tversky, 1969).

With advances in computational technologies over recent decades, a series of computational models has been developed to simulate neuronal activity and decision making. Most studies have focused on perceptual decisions for making binary choices, such as the biophysical cortical attractor model proposed by Wang (Wang, 2002), the model put forward by Shadlen and Newsome that simulates motion detection at visual temporal area MT (Shadlen et al., 1996) and drift diffusion models for evidence accumulations (Ratcliff and McKoon, 2008). Although models have been built to simulate multiple option (rather than just two option) perceptual decisions (Roe et al., 2001; Usher and McClelland, 2001), an investigation using a *biophysically plausible* model to make predictions on the effect of irrelevant options on optimal *value-based* decision making is lacking.

The current study adapts a biophysical cortical attractor model (Wang, 2002), which was originally used to simulate binary perceptual decisions. Properties of the neurons, such as membrane potentials, synaptic connectivity strength and time constants for different receptor currents, were parameterized using findings from previous neurophysiological studies. In this binary choice model, there were two excitatory pyramidal neuronal populations in which each population received input from sensory evidence that favours one

of the two perceptual decisions. The evidence is accumulated and maintained by a recurrent excitation mechanism within each pyramidal population. Both pyramidal populations are connected to a pool of inhibitory interneurons which has a critical role of mediating the competition between the pyramidal populations. The competition is resolved and a decision is made once a pyramidal population remains in a high attractor state and suppresses the alternative pyramidal population to a low firing state via the interneurons. Previous studies have shown that this model predicts the activity of decision making regions in the brain, namely the ventromedial prefrontal cortex (vmPFC) and intraparietal sulcus (IPS), when people were making decisions between two options (Hunt et al., 2012).

To predict how the value of an irrelevant distractor affects decisions between two relevant options, we adapted the original biophysical cortical attractor model (Wang, 2002) by adding a third pool of excitatory pyramidal neurons that received input from the distractor. In other words, we used the network to examine how choices would be made between two options when a third option was present and we examined how the influence of the third option varied as a function of its value. We examined a situation in which the model was given two options that it could choose between and we examined what happened when a third option, that could not be chosen, was transiently presented to the network. This simulates the situation that might, for example, occur when a person is making a choice between two items in a supermarket after they have just seen an advert for a third item that is not available to be purchased in the supermarket. Similarly it simulates the situation that might occur when an animal chooses between two food sources after having seen a third type of food that is currently out of its reach. It might be thought that if the third item were very salient then it might disrupt the decision making process. In general, in investigations of attention, it is the salience of a distractor that determines its power to disrupt other cognitive processes. However, to briefly anticipate the pattern of results we found, it was the opposite type of distractor, a very poor third alternative that

had the most disruptive effect on the way in which choices were made between the two available options.

In another simulation we examined the impact of a third option that was no longer a distractor but which was presented to the network in the same way as the other options and which was available for choice. In this simulation we were constrained to examine the effect of a third option that was lower in value than the other two options (if it were higher in value then it was likely to actually itself be chosen).

We found that the firing difference between excitatory populations, as an index of optimal model's choice, decreased when the value of the distractor was worse. In this class of biophysical models, strong inhibitory activity is important in efficiently resolving the competition between excitatory populations for making optimal and non-stochastic choices. By analysing the activity of the inhibitory interneurons, we were able to put forward a mechanistic explanation for the phenomenon – smaller distractor values were associated with weaker overall inhibitory activity. Similar results were found when two complementary versions of the model were used for the simulation.

2.2 Methods

We extended a biophysical cortical attractor network model in a full spiking form (Wang, 2002) for making decisions between two options which has previously predicted value difference signals in vmPFC and IPS when there are two choices (Hunt et al., 2012; Jocham et al., 2012). Our model was similar but included three rather than two populations of excitatory pyramidal neurons that represented potential choices (**Figure 2**a). Each population's neurons received an input proportional to the value of one option. Just as in the original model, there was strong recurrent excitation between neurons *within*

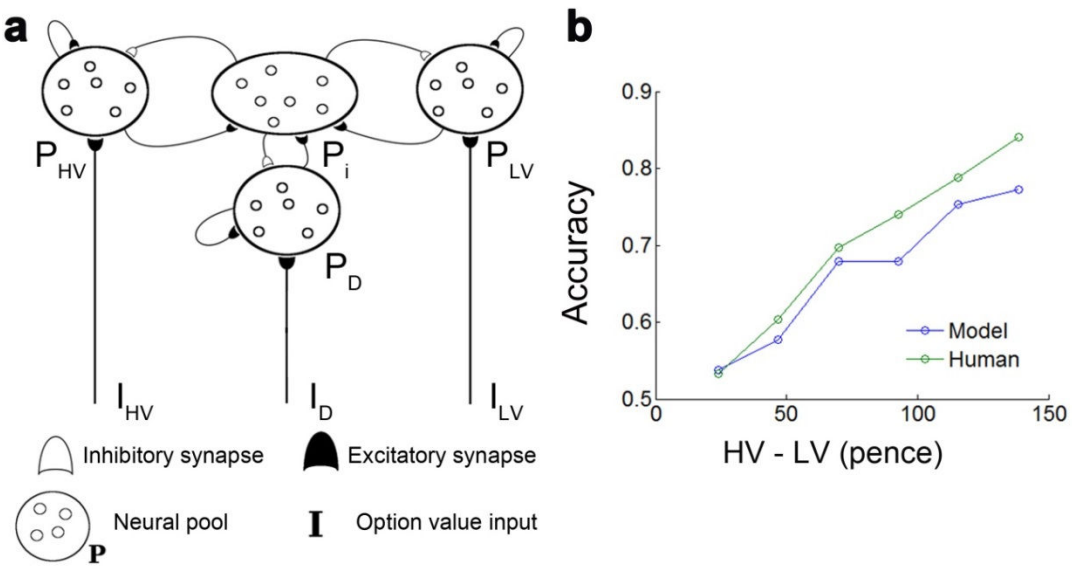


Figure 2. *Biophysical model predictions.* ***(a)*** *The biophysical model is similar to models used elsewhere (Wang, 2002; Hunt et al., 2012) but extended to contain three pools of excitatory pyramidal neurons (P_{HV}, P_{LV}, P_D) corresponding to choices of either the high value or low value available option or the distractor D. There is recurrent excitation between neurons within pools but inhibitory interneurons (P_i) mediate competition between pools. Activity in each pool is initially affected by an input proportional (I_{Hv}, I_{LV}, I_D) to the value of each option (I_{HV} and I_{LV} was on between 200-1000 ms and I_D was on between 200 and 300 ms) but inhibition between pools leaves only a single pool in a high firing attractor state. The corresponding option is then chosen.* ***(b)*** *The scaling of the value related input was selected at 17.5 arbitrary units with a variance of 25 arbitrary units so as to match the average relationship found between value difference and accuracy seen in the human subjects tested in the subsequent behavioural experiment (**Figure 8**). The model's "choices" were defined once P_{HV} and P_{LV} had a firing difference of 6 Hz.*

populations but interneuron-mediated inhibition *between* populations. The inhibition between populations instantiates a competition that leads to an attractor state in which one single population has a high firing rate, and the others low firing rates; the corresponding option is thus 'chosen' by the network. High value options are more likely to be chosen by the network because their representative population receives stronger external excitatory input.

Excitatory neurons parameterization (Distractor model)

The design of the excitatory neuronal populations was mostly the same as in the implementation of the model reported in Wang et al. (Wang, 2002). A modification made to account for differences in our task was that we split the overall number of option specific excitatory pyramidal neuronal pools into three equal populations, each containing 240 neurons. The excitatory neuronal populations (P_{HV}, P_{LV}, P_D) were assigned to receive input from a higher value option (HV), lower value option (LV) and distractor (D).

We examined how the comparison between two options was affected by the presence of a distractor in a manner that would generate predictions of behaviour and brain activity. Obviously, if the distractor has the highest value, then it might itself be selected. In those cases behaviour will be silent as to how the first two options were compared. If the distractor is always lowest in value then it will have little parametric range and its value is easily confounded with those of the first two options. We therefore examined the impact of a distractor that is only available transitorily – an option with a potential value but which could not itself be chosen. To ensure that it could briefly be considered a potential alternative, the identity of the distractor was only revealed shortly after (100ms) all three options were presented. Therefore, initially, all three populations of neurons, P_{HV}, P_{LV} and P_D (**Figure 2**a) become active and inhibit one another indirectly through an inhibitory population (P_I; **Figure 2**a). The distractor option input was removed 100 ms after presentation and the inputs for the other two choosable options were switched off 800ms after presentation. Timings of stimulus onset and distractor removal can be seen in Figure 3.

The model simulation contained 149 trials of different combinations of option value input. We repeated the model 36 times (i.e. 5364 trials in total) and pooled results for our statistical analyses. We sampled the activity of P_{HV} and P_{LV} in the period 700-1200 ms after

presentation to estimate which option would be most active, and therefore likely to be chosen, in a speeded reaction time (RT) decision. The scaling of the value-related input was selected at 17.5 arbitrary units with a variance of 25 arbitrary units to match the frequency of choices between the two available options reflected in human subjects' performance in the task described in Chapter 3 (Figure 2b).

Excitatory neurons parameterization (Three-option model)

In an alternative model, instead of having the distractor option with a shorter input to its corresponding population, we had all three options sending inputs to their corresponding population for an equal amount of time and ***the option with lowest value was considered as a distractor***. In particular, all three excitatory populations received input from their corresponding options between 200 ms and 1000 ms by an option value to firing rate scaling factor of 35 arbitrary units with a variance of 10 arbitrary units. This scaling factor meant that the model made non-random decisions with an accuracy level approximately similar to the subjects on similarly difficult decisions. All the simulated trials had HV and LV values fixed (335 and 279 pence respectively). We used fixed values of HV and LV so that any effect found in the subsequent analyses could be attributed to variance in D (rather than to changes in other factors such as the sum of the values of choosable options). The fixed values chosen approximately correspond to a moderately difficult decision for human subjects.

Statistical analysis

We focus on examining activity and inhibition in neuron pools corresponding to the values of available options and distractor (Figures 3, 4, 5). The difference in value between the

available options and the distractor can be indexed in various ways, HV-D, LV-D, or (average of (HV and LV)-D), all of which are correlated (a consequence of keeping the correlations between HV-LV, HV-D, and D low), and essentially similar results follow in all cases. Here, we report results using HV-D as this might be thought the strongest test of our hypothesis; HV-D is clearly a value difference that is similar in formulation to HV-LV but it should be irrelevant for decision-making.

We examined the impact of available options and distractor on neuronal firing by performing time course analyses. In particular, the original firing time courses were averaged from 10 samples per millisecond to one sample every 10 milliseconds. The time courses were time-locked to the onset of the initial phase and were analyzed by applying a general linear model (GLM) to every time point. The effect sizes (β weights) and standard error of regressors were calculated and plotted in form of effect size time courses (Figures 3, 4, 6). In addition, we also generated model "choices" once P_{HV} and P_{LV} had a firing difference of 6 Hz (Figures 5, 7).

2.3 Results

Biophysical model of multiple option decision-making

We focused on the difference in P_{HV} and P_{LV} population activity levels (Figure 3) because it: 1) is the key decision variable determining likelihood of HV being chosen – an accurate choice of HV is more likely when the P_{HV}-P_{LV} firing rate difference is large; 2) may correspond to the HV-LV value difference signal recorded from vmPFC when decisions are made correctly (Boorman et al., 2009; FitzGerald et al., 2009; Philiastides et al., 2010; Rushworth et al., 2011; Hunt et al., 2012; Kolling et al., 2012; Wunderlich et al., 2012; De Martino et al., 2013). We look for this signal and its influence by the irrelevant value

difference HV-D in the last part of our investigation and so we ensured the input parameters HV-LV and HV-D shared less than 25% of their variance in the set of options we gave the network. Signals with similar levels of shared variance are separable in fMRI data (Rushworth et al., 2011).

We carried out our test by regressing the P_{HV}-P_{LV} population activity difference onto HV-D – the irrelevant relative difference in value between an available option and distractor – and found a *negative* relationship even when the regression model included other factors (HV-LV and HV+LV, Figure 3a); P_{HV}-P_{LV} population difference, the signal in favour of optimal choice, decreased as distractor value decreased (i.e. HV-D increased). The effect was even more prominent in trials when HV-LV was small (**Figure 3**b). We examined HV-D because, like HV-LV, it is a value difference and so it is feasible to look for a corresponding value difference signal in brain activity in later parts of the investigation. It is, of course, possible to carry out the analysis by regressing P_{HV}-P_{LV} onto the absolute distractor value – D (**Figure 3**c) – but the inference remains the same: the P_{HV}-P_{LV} signal difference decreased as distractor value decreased. To see how the effect emerges, the average P_{HV}-P_{LV} firing activities at two different levels of D are also shown (**Figure 3**d).

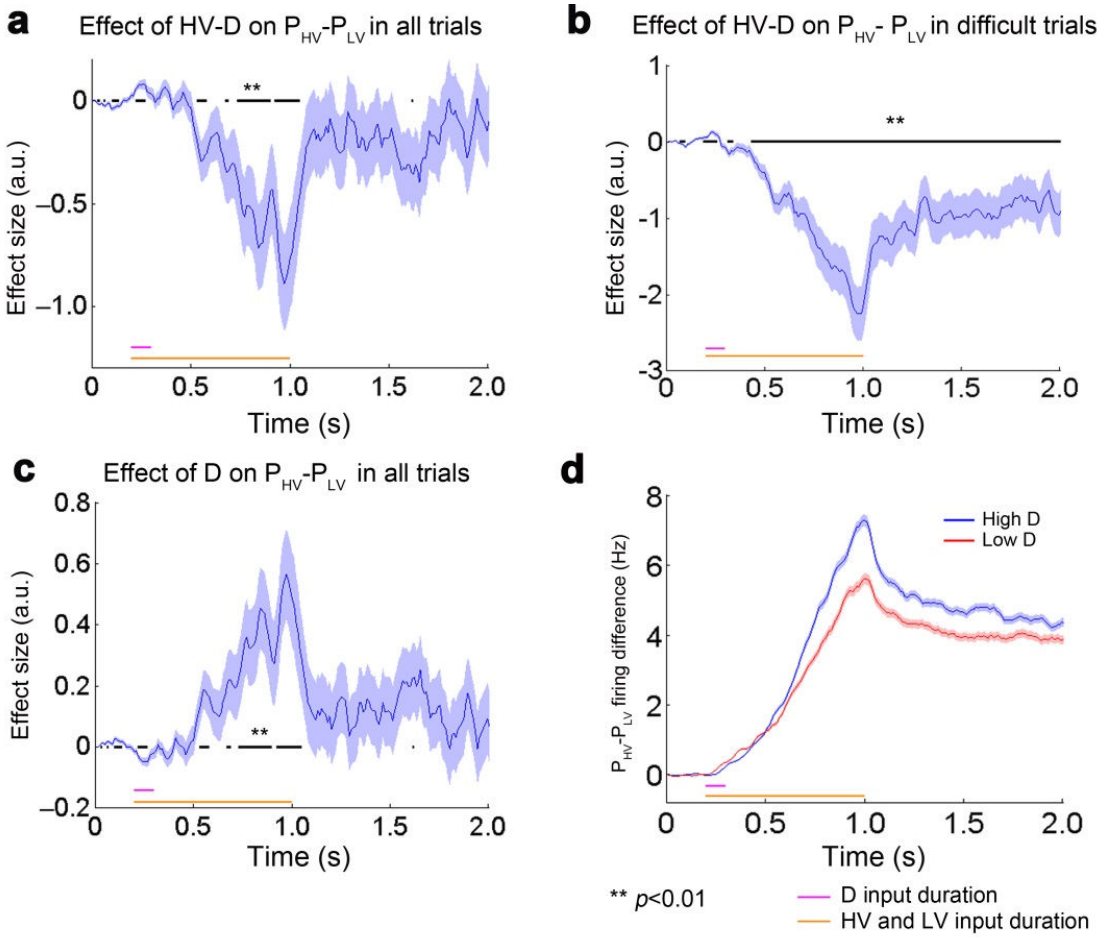


Figure 3. ***(a)*** *The magnitude of difference in the activity of P_{HV} and P_{LV}, which corresponds to the value difference signal (HV-LV) used in many fMRI analyses, decreased as a function of increasing size of the irrelevant difference between the available options and the distractor (HV-D) in a regression analysis.* ***(b)*** *The effect is even clearer on trials when HV-LV was small.* ***(c)*** *A complementary analysis that showed P_{HV}-P_{LV} activity difference increased as a function of distractor value.* ***(d)*** *To see how the effect emerges, P_{HV}-P_{LV} firing differences at two different levels of D are also shown. Shading denotes s.e.m. ** p<0.01.*

The reason for the effect became apparent when inhibitory interneuron activity was regressed onto HV-D (**Figure 4**a); there was less inhibitory interneuron activity when HV-D was higher (the relative value of D was lower). In other words, low value distractors lead to *less* inhibition between populations, which in turn lead to P_{LV} population activity being closer

to, and sometimes exceeding, P_{HV} population activity during the time period when network activity first makes non-random decisions (**Figure 4**b). This will sometimes lead to LV rather than HV being chosen when the relative value of D is lower. When there is increased inhibition, it has a bigger impact on P_{LV} because it is more likely to ensure P_{LV}'s activation level falls to a point where it can no longer exert any competitive effect against P_{HV}. It is also noted that the effect of HV-D on inhibitory activity changed from negative to positive after 800 ms. Since the competition was resolved more quickly with larger initial inhibitory activity in cases when D was large, the overall activity levels of P_{HV} and P_{LV} also dropped earlier in the trial, which in turn resulted in a smaller input to the inhibitory population at the end of the trial.

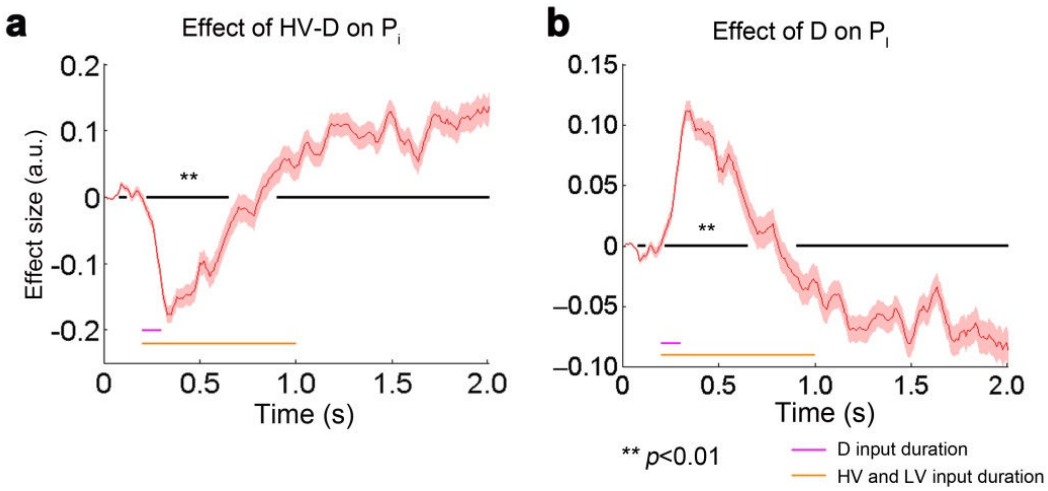


Figure 4. (a) *The amount of inhibitory activity (P_i) increased as HV-D decreased (larger relative distractor value) and P_D inhibited representations of the other two options, P_{HV} and P_{LV} more. In those cases, the competition between P_{HV} and P_{LV} was resolved more quickly, causing a smaller overall input to the inhibitory neurons.* ***(b)*** *A complementary analysis that showed P_i activity increased as a function of distractor value.*

Shading denotes s.e.m.

*** p<0.01.*

Such regression analyses of model activity can be linked to the regression analyses of fMRI data we report in Chapter 4. However, to predict human choice behaviour we looked at the impact of D on which choice the model made – in other words we examined whether D had an impact on whether P_{HV} or P_{LV} entered the higher firing attractor model state. After an initial rise in both populations' activity the model entered stable attractor states (**Figure 5**a). The initial rise in both populations is a consequence of the relatively high input scaling employed to best simulate human choice behaviour (**Figure 2**b). It is consistent with the observation that activity in neural circuits for decision-making initially reflects the sum of choice values prior to reflecting value difference (Hunt et al., 2012). Analysis of attractor states revealed the model was more likely to make an erroneous LV choice when D was much lower in value than the choosable options (**Figure 5**b).

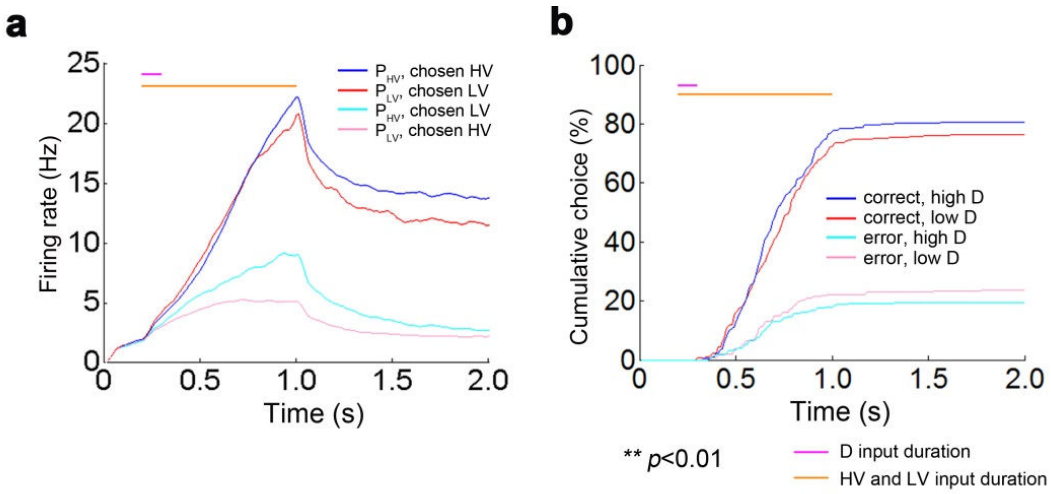


Figure 5. (a) *The excitatory population reached a high attractor state when the corresponding option was chosen by the model.* ***(b)*** *When D was much lower in value than the choosable options the model was more likely to make the wrong LV choice;* P_{HV} *became less likely to enter the higher firing attractor model state.*

Population activity predictions for a three-option biophysical model

We note that by examining the impact of transient value information that does not relate to a choosable option, we can devise the most challenging, and arguably most ecologically valid, test of distractor value impact on decision-making; if even transitorily available value information can disrupt the way in which a decision is made then the phenomenon we are investigating may be a prevalent one. However, we report an additional model simulation demonstrating that the distractor effects remain robust even when it is an equally available option compared to the other two. Obviously in this situation D must be lower in value than both HV and LV or else D itself might be chosen.

In this alternative model, instead of having one option (the distractor option) with a shorter input to its corresponding population, we had all three options sending inputs to their corresponding population for an equal amount of time and ***the option with lowest value was considered as a distractor***. In particular, all three excitatory populations received input from their corresponding options between 200 ms and 1000 ms by an option value to firing rate scaling factor of 35 arbitrary units with a variance of 10 arbitrary units. This scaling factor meant that the model made non-random decisions with an accuracy level approximately similar to the subjects on similarly difficult decisions. All the simulated trials had HV and LV values fixed (335 and 279 pence respectively). We used fixed values of HV and LV so that any effect found in the subsequent analyses could be attributed to variance in D (rather than to changes in other factors such as the sum of the values of choosable options). The fixed values chosen approximately correspond to a moderately difficult decision for subjects.

Similar to **Figure 3**, we analyzed the parametric effect of D value on the firing difference between P_{HV}-P_{LV} (***Figure 6***a, b) and also on the firing of P_i by regression analyses

(***Figure 6***c). We found that increasing values of D led to larger P_{HV}-P_{LV} differences and also higher firing rates of P_i.

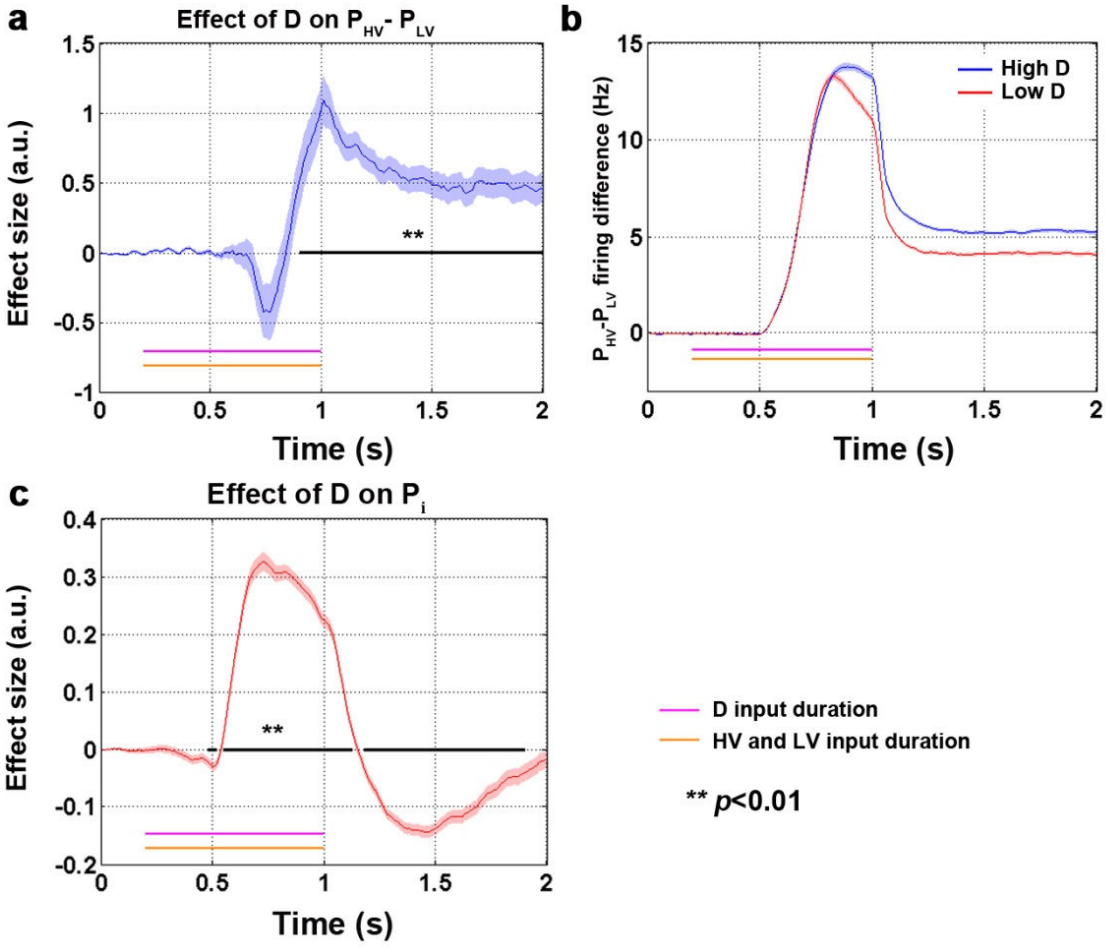


Figure 6. ***(a)*** *Similar to the model with a transitory distractor, the size in P_{HV}-P_{LV} firing difference decreased as a function of smaller non-transitory distractor input.* ***(b)*** *To see how the effect emerges, P_{HV}-P_{LV} firing differences at two different levels of D are also shown.* ***(c)*** *Initially the amount of inhibitory activity (P_i) increased as D was larger. In those cases, the competition between P_{HV} and P_{LV} was resolved more quickly, causing a smaller overall input to the inhibitory neurons after about 1200 ms. Shading denotes s.e.m. ** p<0.01.*

These results were comparable to those in the earlier model with a transitory distractor, suggesting that larger third option values affected the firing difference between the populations for the other two options and the firing of the inhibitory neuron population in a similar manner even when the third option was an available choice.

Next, we examined the effect of D on the model's "choice" when an attractor state was achieved by either the P_{HV} or P_{LV} population 1100 ms after the onset of a trial. We carried out an analysis of population activity as a function of choice that was comparable to the one illustrated in **Figure 5**a for the distractor task. As before, activity in both P_{HV} and P_{LV} populations rose initially but once again the model subsequently entered stable attractor states **Figure 7**a. As in the distractor task model the initial rise in population activity was a consequence of the relatively high input scaling which ensured that empirically measured human performance in the same task was captured accurately. As we have noted, it is also consistent with the observation that activity in neural circuits for decision-making initially reflects the sum of choice values prior to reflecting value difference (Hunt et al., 2012).

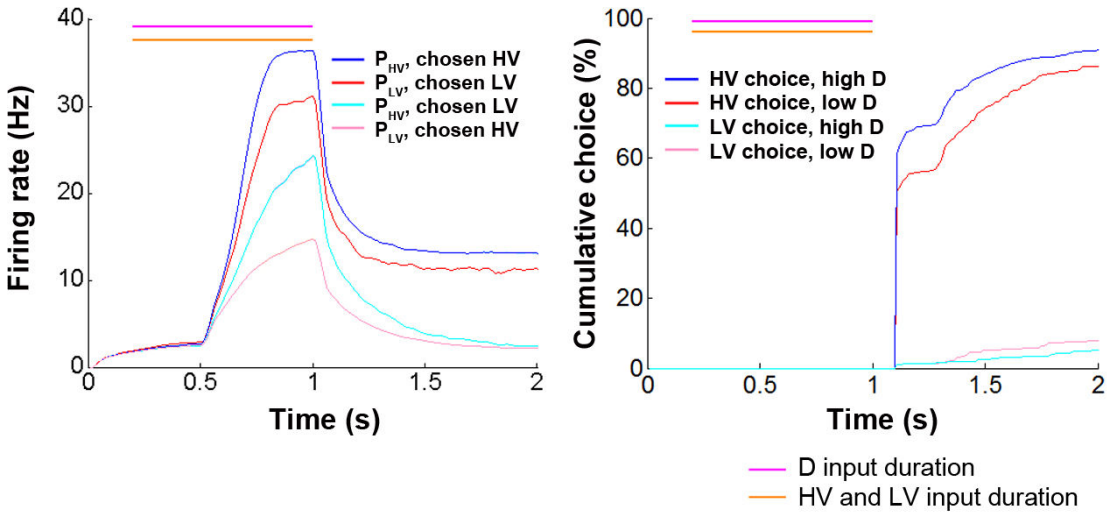


Figure 7. (a) *The excitatory population reached a high attractor state when the corresponding option was chosen by the three-option model.* ***(b)*** *When the value of D was small the model was more likely to make the wrong LV choice; P_{HV} became less likely to enter the higher firing attractor model state.*

Attractor states were defined when either the P_{HV} or the P_{LV} population had a firing rate higher than the other population by at least 6 Hz and the variance of the firing stayed less than 30% of the firing for at least 300 ms. As in **Figure 5**b we separated the trials by using a median split of D value and then investigated the cumulative percentage of HV and LV choices that the model made over time. However, in order to show that the distractor effect on choice was driven by the recurrent dynamics of the model's populations, choices were only assigned after 1100ms when the inputs were switched off. Consistent with our value difference analysis, when D had a higher value, more HV choices (**Figure 7**b, blue and cyan lines) and less LV choices (**Figure 7**b, red and magenta lines) were made.

In summary, taking all our diverse analyses of the biophysical model together, we were able to show consistently that no matter whether the distractor input is transient or not, higher third option values caused larger firing difference between P_{HV}-P_{LV} and more optimal choices by the model.

2.4 Discussion

We made simulations by using two versions of a biophysical model of decision-making that employed mutual inhibition mechanism and showed consistent predictions that contradicts the rational choice theory. Both models predict that the difference in activity between the populations encoding each of two options was *reduced* in the presence of a low value third alternative. In addition, the frequency with which the best option was chosen was *reduced* in the presence of a low value third alternative (Figures 3, 6). The smaller third alternative exerts weaker input to the inhibitory populations, causing more stochastic competitions between the other two options (Figures 4, 6).

Comparison to the original model

The original biophysical model for binary choice includes two excitatory populations, each receives input that is related to evidence favouring one decision (Wang, 2002). The activity levels are sustained with a recurrent excitation mechanism within each population and the two populations are interconnected via a pool of inhibitory interneurons that mediated their competition. Decisions are made once an excitatory population wins the competition and enters a high firing attractor state while the other population loses and enters a low firing attractor state. On this basis, we adapted the model by adding a third excitatory population (**Figure 2**). We extended the model in two different ways. In one new version of the model the third distractor option was only transiently available and so it only exerted a transient effect on the network. This version of the model simulates the situation in which a person might only be able to choose between two options but the brief prior presentation of an additional option that cannot actually be chosen might affect decision making. For example a person considering purchasing one of two different flights on an airline's website might be influenced by the price of a third flight that they also see on the website but which they cannot take because it is fully booked. In the second model the third option was no longer a "distractor"; it was presented to the network for the same amount of time as the other options and could actually be chosen just as the other options could be. This meant that the third option was now always lower in value than the other two options (otherwise it itself was likely to be chosen).

Both the two variants of the biophysical model exhibited three features similar to the original model (Figures 5, 7). First, the excitatory neuronal activity levels ramped up after receiving stimulus input and these activity levels were maintained via recurrent excitation. Second, the competition between excitatory neurons was mediated by a

population of inhibitory interneurons in which activity ramped up as a consequence of the high input from the excitatory neurons. Finally, the competition was resolved by having one excitatory population remaining in a high firing attractor state such that the corresponding option was selected by the model.

Divisive normalization model

Divisive normalization is one influential model that attempts to describe cortical connectivity. Under this account, the firing of the neurons increased as a function of the input strength of their corresponding stimuli and, critically, these neurons' firing is also weakened as a function of the sum of inputs from all stimuli in the environment (Carandini and Heeger, 2012). This kind of model has been used to describe neuronal activity in various perceptual modalities, including visual (Heeger, 1992) and auditory (Depireux et al., 2001) pathways. Recently, attempts have been made to use divisive normalization models to describe neuronal codes for option values during perception-guided and value-guided decision making. Louie and colleagues (2011) recorded individual neuron firing rates in the lateral intraparietal cortex (LIP) in monkeys while they were performing in a matching task with one to three options presented at a time. LIP neurons exhibited higher activity when an option associated with a larger reward value was presented within the corresponding visual response field. Consistent with the divisive normalization prediction, when the total value of the options presented outside the response field of a given LIP neuron was larger (either as a result of manipulating the number of options or the reward magnitudes), the firing rate of that neuron was also reduced. This suggests that divisive normalization does not only describe the computations of sensory neurons, but potentially could also serve as a model to describe high cognitive systems (Carandini and Heeger, 2012).

Comparisons between biophysical cortical attractor and divisive normalization models

When decisions were made between two options A and B, similar predictions are provided by both divisive normalization and the biophysical cortical attractor models – the value-dependent firing of a given neuronal population that represents option A *decreases* as a function of *larger* value inputs associated with the alternative option B. An explanation given by divisive normalization would be that a stronger divisive effect was exerted by a larger total value (A+B) in the environment as a consequence of larger values of option B. The biophysical model provides a mechanistic explanation that the population for option A receives more inhibition from interneurons and exhibits lower firing as a consequence of stronger activity in the option B population.

Intriguingly, the predictions of the models start to diverge when they are applied in three-option decisions (options A, B and C). Our biophysical model predicts that the firing difference between populations responding to options A and B (as an index of how likely the model would choose the best option A) *decreased* as a function of *smaller* values of option C (**Figure 3**). In a different study, Louie and colleagues (2013) simulated the activity of a neural network with a divisive normalization mechanism that makes decisions between three options. By using a similar approach of investigating firing difference between populations for options A and B, their model made an opposite prediction to our findings – the firing difference *decreased* as a function of *larger* values of option C.

It should be noted that the biophysical cortical attractor network model and the divisive normalization model should not be treated as two opposing models. In fact, a variant of the biophysical attractor model could demonstrate normalized activity levels when simulated with a random dot motion task that involved multiple possible directions for

coherent dot motions (Furman and Wang, 2008; Wei et al., 2012). In this kind of model, each pool of excitatory populations exhibits directionally tuned activity as a continuous network, as opposed to the discrete option input in the original model (Wang, 2002) and in our model. The activity tuning was contributed by two features that are absent in our model – the tuned input of excitatory neurons in response to stimuli with differing directions of motion and the inclusion of an additional level of interactions – lateral glutamatergic connectivity between excitatory populations where the strength of connectivity was tuned as a function of the similarities in directional preference between populations. Wei and colleagues (2012) demonstrated that such a kind of continuous attractor network exhibited normalization as a function of the number of stimuli. This is similar to the normalization effect in LIP neurons reported in the Louie and colleagues study (Louie et al., 2011).

It is possible that "discrete" attractor network models, like our three-option model, and continuous attractor network models, like Wei and colleagues' model (Wei et al., 2012), could be better fits to neuronal activity in brain regions with and without a simple spatial topography respectively. It is noted that there is no evidence for a simple topographic organization in the vmPFC, unlike the parietal cortex that contains a visuo-spatial gradient for representing objects in the contralateral visual field (Silver and Kastner, 2009). While it has been well-documented that both regions participate in encoding option values and in making decisions (Shadlen and Newsome, 1996; Boorman et al., 2009; Kable and Glimcher, 2009; Hunt et al., 2012), magnetoencephalography (MEG) data have shown that vmPFC is particularly involved in decisions when it is necessary to integrate more than one decision variable to make optimal choices (Hunt et al., 2012). We argue that our biophysical model could make the best prediction for the effect of a distractor in these kinds of decisions. For instance, when decisions were made between an optimal choice A and a suboptimal choice B in the presence of a distractor and where each option's value is a function of two different aspects – its associated reward magnitude and reward probability. When option A has a

larger magnitude and a smaller probability than option B then accurate decisions must be based on integrating both magnitude and probability information or else by deciding that, on a given trial, one modality or the other is the more relevant one to focus on. In such situations, more suboptimal choices (option B choices) were made in the presence of smaller value distractors.

To summarize, we performed simulations using a biophysically plausible model and made predictions that there were more suboptimal decisions and smaller firing differences between the excitatory neurons responding to the relevant options in the presence of a small value third option. This phenomenon was caused by the existence of weaker inhibitory activity which in turn led to less accurate resolution of competitions between the excitatory populations. Consistent results were found no matter whether the third option exerted a transient or sustained input to the network.

Chapter 3. Suboptimal Decisions with Multiple Alternatives: Human Behaviour

The biophysical cortical attractor model made predictions that rational choice is violated in the presence of a third option. We empirically tested this prediction by designing a behavioural task in which decisions were made between (a) two options, (b) two options in the presence of an unchooseable distractor or (c) three options. We ran different versions of the task on human subjects in five different experiments. We found an effect of the third option that is consistent with our biophysical model prediction – more suboptimal choices were made when the third option value was small. The effect was more robust on difficult trials and on trials when integration of different decision variables was needed; the effect was most robust when consideration of both reward probability and reward magnitude was needed. Finally, we found that the third option influenced decisions in a manner consistent with divisive normalization in situations when there was no demand for integrating decision variables in order to make decisions.

3.1 Introduction

In the previous Chapter, we ran simulations by using a biophysically plausible model and made predictions that when the value of a third alternative is smaller, more suboptimal decisions are made. We performed a series of human behavioural experiments in order to provide empirical evidence for the predictions.

In multiple alternative decision making, there are a number of different theories making very different predictions on how the presence of a third alternative could influence people's choice. One traditional view in economics is the rational choice theory (von Neumann and Morgenstern, 1944), in which one important axiom is that human decision making is independent of irrelevant alternatives (Luce, 1959). In other words, when choices are made between two options (for example a lobster on dinner menu A and a steak on dinner menu B) decisions should not be influenced by the presence of a poor third option (for example a tiny sandwich on dinner menu C) or an unavailable third option (for example a nice seafood platter on dinner menu D that is sold out). If a person slightly prefers the lobster to the steak, then the tiny sandwich or the sold out seafood platter on the menu should not affect the person's decision to choose the lobster.

However, rational choice theory has been violated in a number of ways. One of those is the asymmetric dominance decoy effect that describes how similarities in option attributes could have an impact on decisions (Pettibone and Wedell, 2007). Taking again the example of dinner menu, for instance, a lobster (£30) is slightly more appetitive but more expensive than a steak (£27) and much more appealing but costs the same price as a tiny sandwich (£30). In such case the lobster and the steak are equal in value (appetitive level discounted by price), while the sandwich is the worst option. However, the decoy effect predicts that people are more likely to choose the lobster because it has the same price but is more appetitive than the sandwich (asymmetric dominance). The sandwich would prime people to feel that the lobster has a better value than it and so the lobster is more likely to be chosen than when just the lobster and the steak are presented together. The effect is thought to happen because the value of each of the items on the menu is determined by the two components features they possess: appetitive level and cost. It is difficult to compare the value of such multi-attribute items but when two of them, in this example the sandwich and the lobster, are equal on one dimension then becomes easy to see which item is better

than the other; when price is equal then only appetitive level is important and it is clear that appetitive level of the lobster is higher than the sandwich. The lobster's relative advantage in comparison to the sandwich seems to carry over in to the way that the lobster is compared with the steak so that it ends up being chosen.

Divisive normalization is another theory that describes how the presence of a third alternative could affect binary choice (Louie et al., 2011; Carandini and Heeger, 2012). The normalized option value that guides decisions in this model is estimated by the option's own value divided by (normalized by) the total value of all options presented in the environment. In the dinner menu example, if there is a preference for the lobster over the steak, then this preference will become weaker if the value of the tiny sandwich becomes better, for example by enlarging the sandwich's size or by reducing the sandwich's price. This happens because the total value of all the option value is larger when the sandwich option becomes better and the normalized value difference between the lobster and the steak becomes smaller.

In Chapter 2, we also reported a quite distinct effect – how the presence of a third alternative could influence decision making by using biophysical model simulation. Interestingly the prediction is opposite to that of divisive normalization model – more suboptimal decisions (weaker preference of lobster to steak) are made when the third option is poor (a tinier sandwich). We ran a series of behavioural experiments to test empirically whether human decision making follows the predictions of our biophysical model (positive relationship between optimal choice and third option value), the divisive normalization model (negative relationship) or rational choice theory (no relationship). We found strong evidence supporting the biophysical model prediction. However in the last analysis in this chapter, we also found evidence for a behavioural pattern that is consistent to the divisive normalization prediction. Overall, the value of the third option had a positive

effect on optimal choice when there was a demand for integrating decision variables for making a choice (£30 for a tasty lobster versus £27 for a less appealing steak), which is consistent with the biophysical model prediction. However, the third option value had a *negative* effect on optimal choice in decisions where *any* decision variable would be sufficient for guiding decisions (£30 for a tasty lobster versus £33 for a less appealing steak), which is consistent with the divisive normalization prediction.

3.2 Methods

Three pilot distractor experiments (Studies 1, 2 and 3), one three-option experiment, and one fMRI experiment were performed using various versions of the same decision making task. Different task parameters were used in the different versions and are explained in the *Experimental designs* section below.

Subjects

Twelve subjects participated in Study 1, seven subjects participated in Study 2, three subjects participated in Study 3, twelve subjects participated in the three-option experiment (Study 4), and twenty one subjects participated in Study 5 which also involved fMRI data collection. I report the behavioural results of this fifth experiment while the fMRI results are reported in the next chapter. All subjects were healthy, right-handed, aged 19-34 years and had no psychiatric or neurological history. Informed consent was obtained from every subject. The study was approved by the Oxfordshire National Health Service Research Ethics Committee.

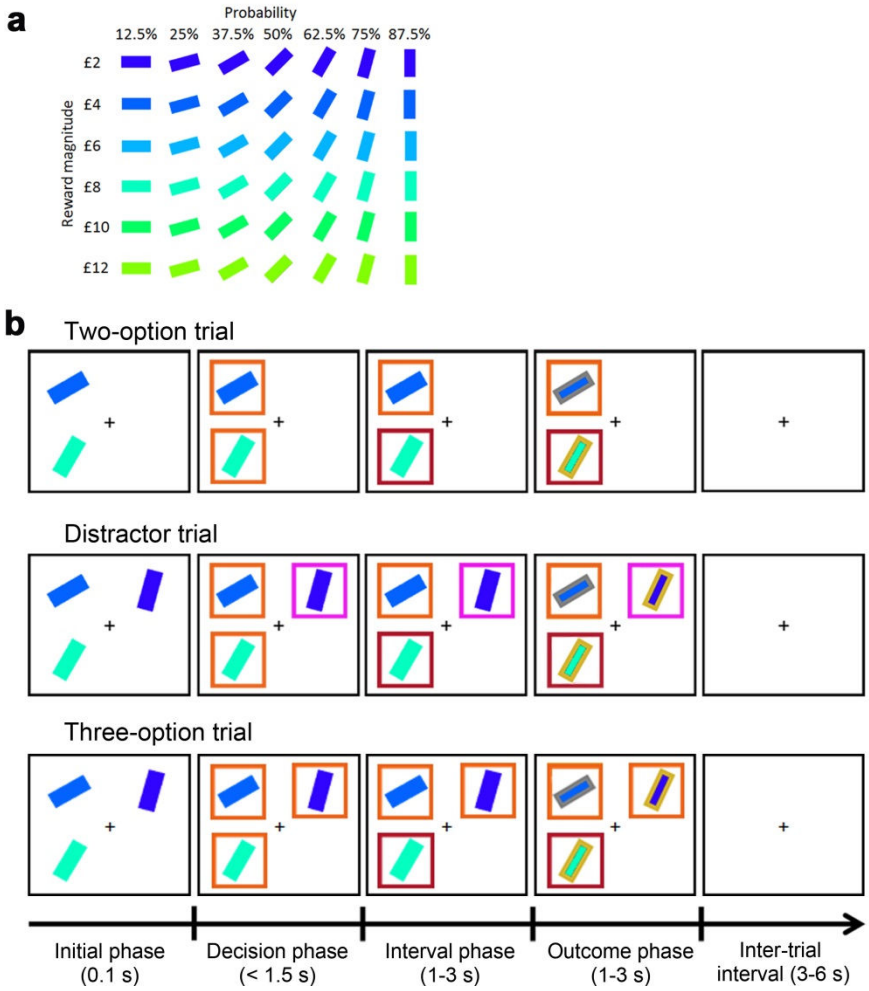


Figure 8. *Behavioural task.* ***(a)*** *An example decision variable-to-stimuli mapping in the fMRI experiment. Prior to the experiments subjects learned that stimulus orientation and colour indicated the probability and magnitude of rewards if the stimulus was chosen.* ***(b)*** *In the initial phase of two-option trials subjects saw two visual stimuli indicating two possible choices. These were immediately surrounded by orange squares, indicating that either option might be chosen, in the decision phase. A further colour change during the interval phase of one box indicated which choice the subject made. In the outcome phase of the trial the outline colour of the chosen stimulus indicated to the subject whether the reward had been won. The final reward allocated to the subject on leaving the experiment was calculated by averaging the outcome of all trials. Distractor trials unfolded in a similar way but, in the decision phase, one stimulus, the distractor, was surrounded by a purple square to indicate that it could not be chosen while the presentation of orange squares around the other options indicated that they were available to choose.*

Experimental task

In our decision making task, subjects chose repeatedly between stimuli associated with different reward magnitudes and probabilities, represented by colours (green to blue or red to blue) and orientations (0° to 90°) of rectangular bars. Associations between visual features and decision variables were counterbalanced across subjects. An example decision variable-to-stimuli map is shown in **Figure 8**a. There were three types of intermingled trials across different versions of the experiment – (i) two-option trials, (ii) "distractor" trials, where three options were presented, but one was signalled as unavailable to select, and (iii) three-option trials, where any of the three presented options could be selected (**Figure 8**b). Each trial began with a central fixation cross indicating an inter-trial interval followed by an *initial phase* in which two (two-option trials) or three (distractor trials or three-option trials) stimuli were presented in randomly selected screen quadrants. Then, in the *decision phase*, orange boxes were presented around two stimuli (two-option trials or distractor trials) or three stimuli (three-option trials) indicating those options were available for choice. In distractor trials, a purple box was also presented around the third stimulus to indicate a distractor. Subjects were instructed to choose an option within a deadline (varies between experiments, see *Experimental designs*). Subjects were warned that they were "too slow" if no response was made within a deadline (1.5 s) or that they made an "error" if a blank quadrant was mistakenly chosen and the trial was aborted afterwards. After an option was chosen, the box surrounding it turned red in the *interval phase*. The edge of each stimulus turned yellow or grey in the *outcome phase* to indicate whether the choice had been rewarded or not respectively. The final reward allocated to the subjects on leaving the experiment was calculated by averaging the outcome of all trials.

Subjects learned the task and visual feature associations and experienced Distractor trials in a practice session before the experiments. At the end of practice, all subjects chose

option with the highest value on >70% of two-option trials when it was associated with both higher reward magnitude and probability.

Experimental designs

In the different behavioural experiments, the designs varied by changing the number of trials, the reward magnitude and probability sets for the stimuli, and the event timings. These are summarized in Table 1.

Table 1. *Parameters for different experiments.*

	Study 1	Study 2	Study 3	Study 4 Three-option	Study 5 fMRI experiment
Number of subjects	12	7	3	12	21
Number of trials	100 two-option 100 distractor	200 two-option 200 distractor	200 distractor	100 three-option	150 two-option 150 distractor
Stimuli					
Reward magnitudes (£)	2, 4, 6, 8, 10, 12	1, 2, 3, 4, 5, 6	2, 4, 6, 8, 10, 12	2, 4, 6, 8, 10, 12	2, 4, 6, 8, 10, 12
Reward probabilities (%)	14.3, 28.6, 42.9, 57.1, 71.4, 85.7	12.5, 25, 37.5, 50, 62.5, 75, 87.5	12.5, 25, 37.5, 50, 62.5, 75, 87.5	14.3, 28.6, 42.9, 57.1, 71.4, 85.7	12.5, 25, 37.5, 50, 62.5, 75, 87.5
Event durations					
Inter-trial interval (s)	3	3	3	3	3-6
Initial phase (s)	1	1	none	1	0.1
Decision phase deadline (s)	0.8	1	1.5	4	1.5
Interval phase (s)	2	2	2	2	1-3
Outcome phase (s)	1.5	1.5	1.5	1.5	1-3

In Study 1, we adopted a schedule of 100 trials for each trial type. In order to avoid confounds in regression analyses caused by our parameters correlating with each other, we ensured that the shared variance between the options with high value (HV) and low value (LV), and the distractor (D), the difference between HV and LV (HV-LV) and between the HV and D (HV-D) was always under 25%. In Study 2, we doubled the number of trials to check for consistencies with the results of Study 1. We also conducted (Study 4) a three-option experiment using a similar design to Study 1 to check whether a similar third option effect could be found when all three options were available to choose. A longer decision deadline was used to allow for greater task difficulty. Note that in such cases the distractor was defined as an available option with the lowest value and was unlikely to be chosen by the participants. Since in Study 1 and 2, there was a long initial phase (1 s) before the distractor identity was revealed in a distractor trial, it was possible for subjects to be involved in a dual decision process – to decide a first priority choice and a second priority choice and then to choose the second priority if the first priority was a distractor. To avoid this, we conducted Study 3 that did not have an initial phase for choice prioritization.

One disadvantage of a paradigm without an initial phase was that subjects could totally ignore the stimulus inside the purple box that indicates distractor identity and that may cause a weaker distractor effect. In the main fMRI experiment, we adapted the design in two ways that could potentially give the most robust distractor effects. First, there was a brief initial phase (0.1s) such that subjects were more likely to process the information of an option that would turn into a distractor and the duration was short enough to avoid dual decisions mentioned above. Second, we provided incentives for subjects to attend to the visual features of every stimulus by interleaving "catch" trials between decision phase and interval phase in 15% of all the trials (**Figure 9**). In this way we ensured that it was even more unlikely that subjects would ignore the distractor. In a "catch" trial, the word "MATCH" was presented once subjects selected an option in the decision phase (1 s). Then, an

exemplar stimulus was presented at the centre of the screen and subjects had to indicate, within 2 s, the position of the same stimulus presented before and during the decision phase. Feedback was then given to indicate whether the response was correct or not (1.5 s). The trial then continued with the resumption of the interval phase, followed by the outcome phase. Each correct response in the "catch" trial added an extra 10 pence to the final reward and each error response leads to no extra reward on that trial.

For all the experiments with both two-option and distractor trials (i.e. the fMRI experiment and Studies 1, 2), each distractor trial was matched with a two-option trial which had the identical pair of *available* options in the distractor trial. The trials were randomly interleaved. To allow comparison between the behavioural results in the fMRI experiment and the biophysical model predictions in Chapter 2, we used an identical set of option value configurations in both cases.

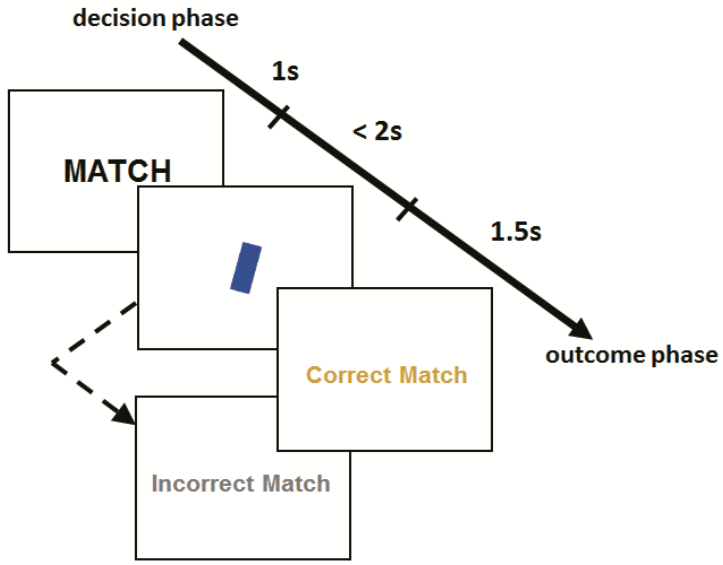


Figure 9. *Catch trials were randomly inserted between the decision and outcome phase for 15% of all trials in the fMRI experiment.*

Behavioural analysis

The value of each option was calculated by multiplying the associated reward magnitude and probability. The difference in value between the available options (HV, LV) and the distractor (D) can be indexed in various ways: HV-D; LV-D; or [(average of HV and LV)-D], all of which are correlated (a consequence of keeping the correlations between HV-LV, HV-D, and D low), and essentially similar results follow in all cases. In most cases, we report results using HV-D as this might be thought as the strongest test of our hypothesis; HV-D is clearly a value difference that is similar in formulation to HV-LV but it should be irrelevant for decision-making. It would therefore be especially striking if HV-D influenced the impact HV-LV had on behaviour.

Logistic regression analyses were performed for every subject to predict accuracy (HV choice) using different general linear models (GLMs). The β weight of each regressor in the GLM of every subject was calculated and then a one-sample t-test against zero was computed. Such a test identified regressor effects that replicated throughout the group of participants. Two-dimensional plots were generated to illustrate behavioural effects by binning each distractor trial (**Figure 13**) or three-option trial (**Figure 11**) according to HV-LV and HV-D and calculating the average accuracy within each bin.

3.3 Results

We designed a task to test in human subjects the impact of an additional distracting alternative on decision-making. Subjects made decisions between two options in the presence of an unchooseable distractor in Studies 1-3 and in Study 5 (the main fMRI experiment), and between three available options in Study 4, the three-option experiment.

Human behaviour in Studies 1 to 3

We ran three studies outside the MRI scanner that all involved a similar task paradigm with two choosable options in the presence of a distractor. We investigated the effects of the relevant comparison HV-LV and the irrelevant comparison HV-D on decision making. **Figure 10** shows that the effects on accurate HV choices are consistent across the pilot studies. As might be expected, fewer HV decisions were made when there was a smaller HV-LV value difference (Study 1, β=0.711, t_6=4.979, p=0.003; Study 2, β=0.464, t_6=5.400, p=0.002; Study 3, β=0.426, t_2=5.854, p=0.028). More importantly, fewer HV choices were also made when there was a *larger* HV-D value difference (i.e. relatively smaller distractor value; Study 1, β=-0.514, t_6=-3.321, p=.005; Study 2, β=-0.375, t_6=-3.202, p=0.019; Study 3 difficult trials, β=-0.243, t_2=-4.791, p=0.041). Note that in Study 3 the negative effect of HV-D was significant only when the analysis was focused on half of the trials when HV-LV was smaller than the mean, so the decisions were more difficult (**Figure 10**, green bars). Log transformation was also performed for the HV-LV and HV-D effects in Study 3 because of the small sample size (n=3) and relatively large variance.

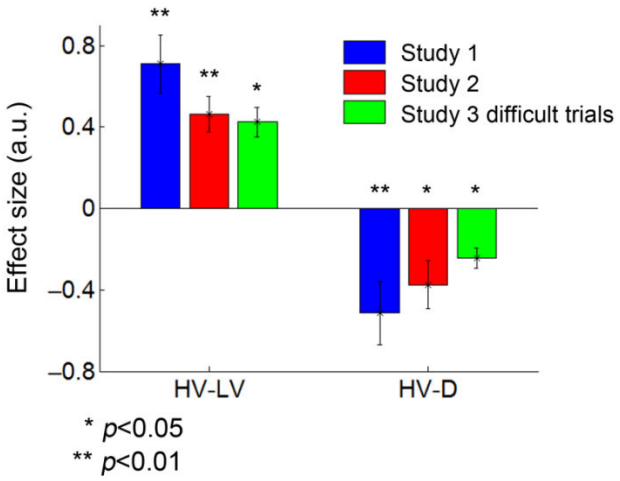


Figure 10. *Pilot behavioural experiments. We ran three studies outside the scanner using a similar task paradigm which mainly differed across experiment by having different durations for the task events (see* **Table 1***). We found consistent results across the three studies; the HV-LV value difference had positive effects and the HV-D value difference had negative effects on choosing the optimal option HV. In other words, more suboptimal choices were made for smaller relative distractor values.*

Study 4: The distractor effect was independent of the availability of the distractor

In order to test whether the "distractor effects" on human behaviour were dependent on whether the third option could be chosen or not, we also performed Study 4 in which all three options were available choices. As in the three-option biophysical model, here an option with the lowest value in a trial was considered as a distractor. In the analysis, we excluded trials when D was chosen because it would not be surprising to see D itself being chosen more often when its value increased, and we were more interested in investigating whether the accuracy of choosing HV *relative* to choosing LV would change when D's value decreased; this analysis is analogous to those conducted for Studies 1-3 that examined accuracy of HV choice relative to LV choice when D could not be chosen. The regressors in

this analysis were value differences between HV-LV and between HV-D, the total HV+LV+D value and an interaction regressor (HV-LV)(HV-D).

This analysis revealed results comparable to those in Studies 1-3. Subjects were more accurate in choosing HV as HV-LV increased (β=1.256, t_{11}=2.747, p=0.019, from bottom to top in **Figure 11**). Again, more LV choices were made when HV-D was large (β=-0.296, t_{11}=-2.154, p=0.054; from left to right in **Figure 11**). Taken together with our results in Studies 1-3, this showed that in cases when the value of a third option was very small, whether it was an available or unavailable distractor, it became more difficult to choose the best option meaning that more second best choices were made. Finally, the (HV-LV)(HV-D) interaction had a marginal positive effect (β=1.242, t_{11}=1.972, p=0.074), suggesting that the negative HV-D effect was stronger when HV-LV was smaller which is similar to our results in Study 3.

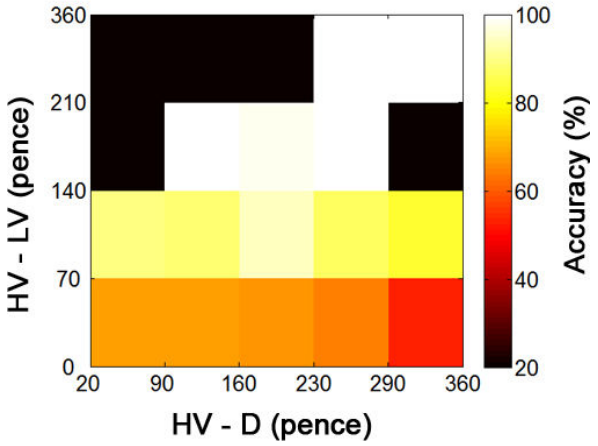


Figure 11. *Human behaviour in a three-option task. Subjects were more accurate in choosing HV as HV-LV increased (from bottom to top in the figure; similar to the positive HV-LV effect in Studies 1-3). More mistakes were made by choosing LV when HV-D was large (from left to right in the figure; similar to the negative HV-D effect in the distractor trials Studies 1-3). Note that black boxes are bins with no trials.*

To further confirm that the distractor effect was driven by the variance of D, we repeated the same analysis but substituted the HV-D and (HV-LV)(HV-D) terms by D and (HV-LV)D respectively, again trials with D choice were excluded in this analysis. The results were consistent with those in the "HV-D" analysis: there was a positive effect of HV-LV (β=2.669, t_{11}=10.696, p<0.001), a marginal positive effect of D (β=0.321, t_{11}=2.102, p=0.0594) and a negative effect of (HV-LV)D (β=-0.788, t_{11}=-3.326, p=0.007).

The distractor's relative and absolute values were related to different kinds of suboptimal choices

We confirmed the effect of the third option on optimal decision making in a complementary logistic regression analysis in which we investigated the effect of each individual option value, instead of looking at the value differences, on choosing HV versus not choosing HV (trials with D choices were included), to confirm that a third option affected subjects' decision by both its *individual value* and its *relative value* to other options. Unlike the "value difference analysis" above, we did not exclude those trials when D was chosen. We included the following regressors to predict HV choice: HV, LV, D for examining individual value effects and the HV×D, LV×D interaction terms for examining the relative value effect. We found a positive impact of HV (β=2.503, t_{11}=10.302, p<0.001) and a negative impact of LV (β=-2.924, t_{11}=-11.232, p<0.001) on choosing the best option, which was similar to the positive HV-LV effects in the previous analysis above and in the "distractor task" in our main manuscript. We also found a negative impact of D (β=-0.899, t_{11}=-2.716, p=0.02), confirming the first kind of third option effect *(individual value effect)*: *increasing* D value makes it harder to choose the best option due to more D choices. Finally, the HV×D and LV×D interaction terms had a negative effect (β=-1.809, t_{11}=-2.813, p=0.017) and a positive effect (β=2.725, t_{11}=4.530, p<0.001) on choosing the best option respectively. This

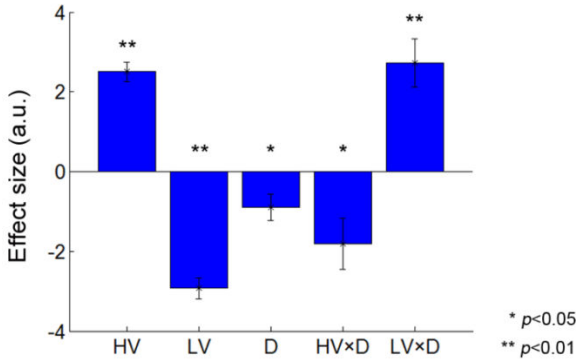


Figure 12. *Two kinds of suboptimal choice effects caused by the distractor in the three-option task. Larger absolute values of D were related to more suboptimal choices, which was likely to be caused by choosing the distractor itself. On the other hand, a relative value effect of distractor was borne out by having a combination of a negative HV×D and positive LV×D effects, which was similar to the negative HV-D effect in difficult trials that was related to suboptimal choices of the second best option reported in all of our behavioural experiments.*

shows that by *decreasing* D value, it was *harder* to choose the best option in situations when HV had a *low* value and when LV had a *high* value. In other words, it was *harder* to choose the best option for increasing HV-D when HV-LV was *small*, due to more LV choices as suggested by the previous analysis using value difference regressors, this confirms the second kind of third option effect *(relative value effect)*. This second phenomenon was analogous to the positive (HV-LV)(HV-D) effect reported in our main manuscript (i.e. it was harder to choose the better option when HV-LV was small and distractor value was small).

To summarize, here we have provided evidence from a three-option experiment for 1) an intuitive individual value effect of the third option D such that higher D values causes suboptimal decisions due to D choices and 2) a counterintuitive relative value effect of a third option such that lower D value causes suboptimal decisions due to LV choices. Together with the results from Studies 1-3, we showed the relative value effect of the

distractor on suboptimal decision making was consistent regardless of whether or not the third option could be chosen.

Study 5: Human behaviour in the fMRI experiment

In the fMRI experiment, we included distractor trials and two-option trials to investigate the "distractor effect" so that we were able to focus on the more surprising relative value effect on suboptimal decisions by masking the less surprising individual value effect. Also we were able to sample the full parametric range of third options (including cases when the third option had a value higher than the HV) to investigate the counterintuitive relative value effect. The set of option values in this experiment corresponded to those used when testing the model in Chapter 2. Each distractor trial had a matched two-option trial with identical reward probabilities and magnitudes for the two available options. Any behavioural effects found only on distractor trials must therefore be due to distractor presence.

Similar to the experiments above, we investigated the effect of distractor value on decision-making by performing a logistic regression analysis. We included value differences between the available options (HV-LV), their sum (HV+LV) and value difference between the higher value option and the distractor (HV-D) as regressors, plus an interaction regressor (HV-LV)(HV-D). Another regressor indexed two-option or distractor trial type.

The results were consistent to those in Studies 1-3 and the three-option experiment. Subjects made more mistakes as HV-LV decreased (β=0.554, t_{20}=13.737, p<0.001) but also the same was true when HV-D was large (β=-0.179, t_{20}=-2.292, p=0.033), suggesting it was more difficult to choose HV when the distractor had a much lower value. Finally, the interaction term (HV-LV)(HV-D) was also significant (β=0.279, t_{20}=3.466, p=0.002), showing that the effect of HV-D causing inaccurate decisions diminished when HV-LV was larger

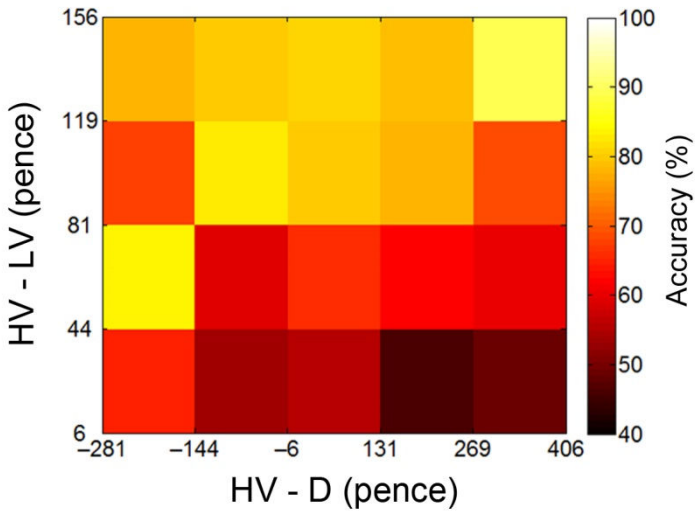


Figure 13. *More suboptimal choices were made in the presence of a low value distractor. Average choice accuracy (% HV choice) of all subjects is indicated by colour as a function of both the difference in value between the two available options (HV-LV) and the difference in value between available options and the distractor (HV-D). Choices are less accurate when the HV-LV difference is small (bottom of figure contains more red and dark colours than the top) but also, importantly, when HV-D was large (i.e. small distractor; more red and dark colours at the right hand side than the left hand side of the figure) especially when HV-LV was small.*

(**Figure 13**). Subjects were less accurate in distractor than two-option trials (β=-0.081, t_{20}=-2.455, p=0.023).

Next, we analyzed the distractor trials and two-option trials separately but using the same regressors. In the two-option trials, the distractor value in each trial of the GLM was taken from the matched distractor trial. In **Figure 14** we show that the HV-D effect was only significant on distractor and not on two-option trials.

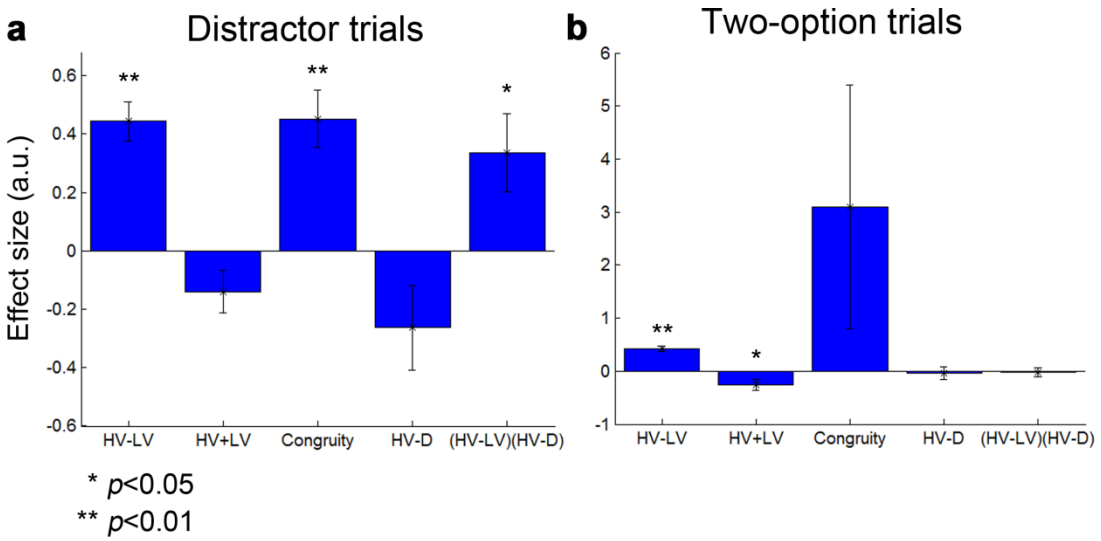


Figure 14. *The effect of HV-D on human behaviour was specific to Distractor trials. Here we present a more extensive analysis of performance in the distractor and two choice tasks in all subjects (n=21). When we analyzed data from both the two-option and distractor trials we could see that subjects made more mistakes (chose HV less) as HV-LV decreased and when HV-D was large (suggesting it was more difficult to choose HV when the distractor had a much lower value). In this analysis, we added an additional binary regressor, Congruity, that describes whether both the HV option's reward magnitude component and the HV option's reward probability component were both larger than the LV option's reward magnitude and reward probability components. In other words, it tested whether both components of the HV option were congruent with one another in suggesting HV was the better option. HV choices were more common when there was congruency. Similar effects were also present when (a) we focused on analysing the data from just the distractor trials. HV-D had a marginally negative impact on accuracy but the clearly significantly positive (HV-LV)(HV-D) interaction term indicated the HV-D effect was stronger when HV-LV was smaller. (b) When we analyzed the data from just the two-option trials using notional distractor values obtained from the matched distractor trials the results showed that the subjects' choices were not affected by the value of the notional distractor.*

The distractor effect is more robust on difficult trials

To test the insights derived from our GLM analysis of behaviour which had revealed an HV-D effect, we calculated an alternative GLM using regressors of HV-LV, HV+LV, and D.

By using this approach we were able to check that any effect that we had linked with the HV-D term in the first analysis was not simply due to the impact of the values of HV rather than the value of the distractor (D). This is possible in this second analysis because the impact of the values of the available options and the distractor are indexed by separate regressors – by HV+LV and by D respectively. Because the first analysis had suggested that the impact of the distractor was greatest when HV-LV was small (when decisions were difficult) and not when HV-LV was large, we now carried out two separate analyses with our new set of regressors after dividing the trials into those where the HV-LV difference was small or large.

In the first case, on trials where HV-LV was small and decisions were difficult (**Figure 15**, blue), we observed a significant impact not just of HV-LV on choice accuracy (β=0.710, t_{20}=3.627, p=0.002) but also an impact of D (β=0.157, t_{20}=2.010, p=0.049) which confirmed that when distractors were low in value (corresponding to the situation where HV-D is large) then subjects made less accurate choices. However, as the initial regression models had suggested the effect disappeared when decisions were easy (the positive interaction effect of (HV-LV)(HV-D)) and we looked at accuracy on trials where the HV-LV difference was large (**Figure 15**, red). Once again a main effect of HV-LV was found: (β=0.560, t_{20}=3.602, p=0.002); but the effect of D now in the opposite direction and no longer reached significance: (β=-0. 147, t_{20}=-1.751, p=0.095).

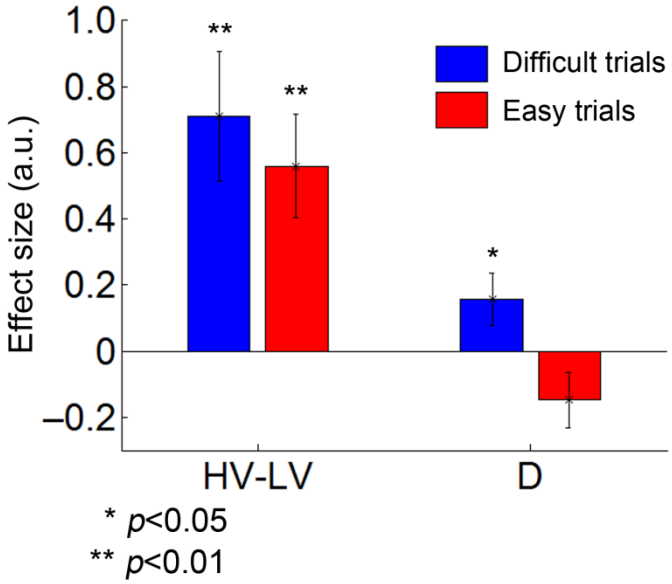


Figure 15. *The effect of HV-D on human behaviour was caused by the value of the distractor. On difficult trials (blue), we observed a significant impact not just of HV-LV on choice accuracy but also an impact of D which confirmed that when distractors were low in value (corresponding to the situation where HV-D is large) then subjects made less accurate choices. On easy trials (red), a main effect of HV-LV was found but there was no effect of D.*

The linearity of the distractor effect was not influenced by the value rank of the distractor

In another alternative analysis, we considered that the negative effect of HV-D could be confounded by the value rank of the distractor. It is possible that when D had the highest value, subjects had to engage in more inhibition to avoid choosing D and made more errors. To test for this, we split the HV-D regressor by whether the distractor was highest or lowest in value. The results of this general linear model were comparable to those reported in above. We found a positive effect of HV-LV (β=0.558, t_{20}=8.283, p<0.001), a negative effect of HV-D when distractor was highest in value (β=-0.281, t_{20}=-3.796, p=0.004), a negative effect of HV-D when distractor was lowest in value (β=-0.174, t_{20}=-2.454, p=0.023) and a positive effect of (HV-LV)(HV-D) (β=0.384, t_{20}=3.458, p=0.003; **Figure 16**a). No significant difference was found between the two HV-D regressors (t_{20}=1.030, p=0.315). We also

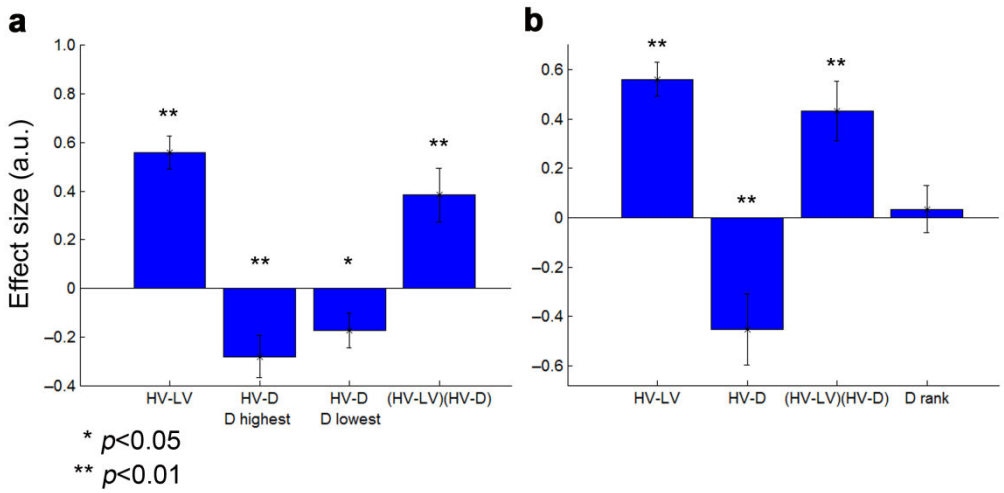


Figure 16. *The effect of HV-D was not dependent on the value rank of the distractor.* ***(a)*** *We split the HV-D regressor by whether the distractor was highest or lowest in value. Both HV-D regressors had a negative impact on accuracy.* ***(b)*** *In another analysis we used an additional regressor that described the rank order of the value of D, instead of splitting HV-D into two regressors. A similar HV-D effect was found in this model.*

performed another analysis in a similar spirit that, instead of splitting HV-D into two regressors, used an additional regressor that described the rank order of the value of D (i.e. whether the value of D was the highest, middle or lowest among the three options). Similar results were found in this model: a positive effect of HV-LV (β=0.566, t_{20}=8.211, p<0.001), a negative effect of HV-D (β=-0.435, t_{20}=-2.926, p=.008) and a positive effect of (HV-LV)(HV-D) (β=0.441, t_{20}=3.525, p=.002; **Figure 16**b).

The distractor had opposite effects on decision making depending on task demands

We have shown that in human decision making, the presence of a smaller third option value has an effect of causing more suboptimal choices, which is consistent with the results of our biophysical model simulation in Chapter 2. We argued that the biophysical model could potentially be a good model to describe vmPFC activity (see Discussion of

Chapter 2). Magnetoencephalographic (MEG) data also show that both the vmPFC and parietal cortex are involved in decision making, however the vmPFC is particularly disengaged in decisions when there is no need for integration of multiple decision variables (i.e. congruent trials in our experiments when HV had both reward probability and magnitude larger than that of LV). In such cases, we hypothesize that the "distractor effect" could be particularly weak.

To test our hypothesis, we ran a similar GLM to predict when the optimal HV choice would be taken that involved the regressors HV-LV, HV-D and (HV-LV)(HV-D). However, this time we split the HV-D and (HV-LV)(HV-D) regressors as a function of whether it is an incongruent trial that required integration of decision variables or a congruent trial that did not require any integration. We also inserted a binary Congruity regressor as a controlling factor to partial out the non-parametric difference between the two pairs of HV-D and (HV-LV)(HV-D) regressors. On incongruent trials, we found a negative HV-D effect (β=-0.025, t_{20}=-2.567, p=0.018; **Figure 17**a) similar to those we have reported before, while interestingly the (HV-LV)(HV-D) interaction term that had a positive effect in other analyses was not significant (β=-0.003, t_{20}=-0.359, p=0.726). This suggests that subjects made *more* suboptimal choices when the distractor value was *small* and the effect was *uniform* for all levels of difficulty (indexed by HV-LV) on incongruent trials. Surprisingly, on congruent trials, HV-D had a *positive* effect on accurate decisions (β=0.014, t_{20}=2.498, p=0.021) and (HV-LV)(HV-D) had also a positive effect (β=0.020, t_{20}=3.360, p=0.003). This suggested that the predictions for congruent trials were the opposite – *less* optimal decisions were made in the presence of *larger* distractors (i.e. smaller HV-D) and the effect was *stronger* when there was a larger HV-LV value difference.

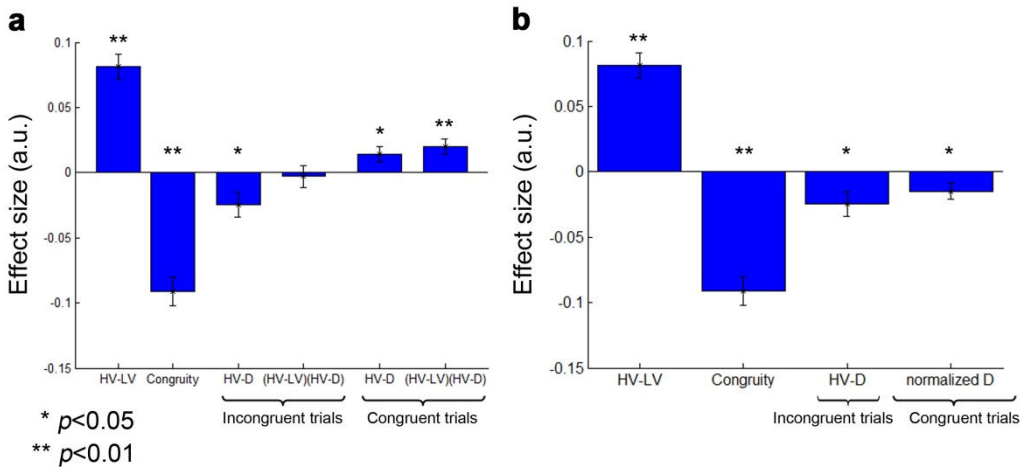


Figure 17. *The distractor effects were dependent on task demand.* ***(a)*** *On incongruent trials when subjects needed to integrate both reward magnitudes and probabilities of HV and LV for decision making, smaller distractor values (i.e. larger HV-D) were related to less accurate decisions. However on congruent trials when both reward magnitude and probability of HV were larger than that of LV, larger distractor values were related to less accurate decisions. The distractor effect on these trials was even more robust when HV-LV value difference was large.* ***(b)*** *In an alternative model, the results show that the positive HV-D and (HV-LV)(HV-D) effects on congruent trials are contributed by a divisive normalization mechanism effect reflecting the value of the distractor.*

The distractor effect followed divisive normalization predictions on congruent trials

The combination of the positive HV-D main effect and the positive (HV-LV)(HV-D) interaction effect on congruent trials seems consistent with the divisive normalization account for how the distractor value could be predicted to affect decision making (Carandini and Heeger, 2012). In this model, the normalized value of the distractor follows:

$$normalized\ D = \frac{D}{HV+LV+D}$$

(Louie et al., 2011; Louie et al., 2013).

In one way, normalized D is *increased* as a function of the objective distractor value

(the numerator of the equation) and this is analogous to the positive HV-D effect that larger D was related to less HV choice. In another way, normalized D is decreased as a function of the summed objective value of all options presented in the environment (the denominator of the equation). This latter influence might be particularly prominent in cases where the LV value is small and the variance of D has a *larger impact* on the overall HV+LV+D denominator of normalized D. This phenomenon is analogous to the positive (HV-LV)(HV-D) effect, in which the parametric changes in distractor value have a bigger influence on HV choice when HV-LV is large (i.e. LV is relatively small).

To confirm that the effects of (HV-D) and (HV-LV)(HV-D) on congruent trials are contributed by the normalized distractor value, we substituted these two terms by normalized D in the GLM. The results showed that there is a negative normalized D effect on HV choice (β=-0.015, t_{20}=-2.359, p=0.029; **Figure 17**b), suggesting that more suboptimal decisions were made when the normalized distractor value was larger on congruent trials. This result is consistent with the divisive normalization prediction. Nonetheless, here we found a negative non-normalized HV-D effect (β=-0.025, t_{20}=-2.567, p=0.018) on incongruent trials, suggesting that more suboptimal decisions were made when the relative distractor value was smaller on incongruent trials and that is consistent with our biophysical model prediction.

3.4 Discussion

It is debateable how the presence of a third option could affect decision making between two other options – whether the third option has a positive effect, a negative effect or no effect. Our biophysical model (Chapter 2) supported the view of a positive distractor effect (i.e. negative HV-D effect), whereby larger values of the third option are related to more optimal decisions through boosting of the inhibitory activity of the network and bring about

a less stochastic competition between options. Conversely, the possibility of a negative distractor effect (i.e. positive HV-D effect) is supported by the divisive normalization model (Reynolds and Heeger, 2009; Louie et al., 2013), in which the value difference between two options is normalized by the total value of all options and becomes smaller in the presence of a larger distractor. Finally, the "independent of irrelevant alternatives" axiom of rational choice theory advocates that the presence of a low available or unavailable third option should have no effect on decision making (Luce, 1959). In a series of behavioural experiments, we were able to provide consistent evidence to support the idea that more optimal decision making is positively related to the value of the third option, no matter if it is a low value option that is available to be chosen (Study 4; **Figure 11**) or an unchooseable distractor (Studies 1-3 and Study 5 – the fMRI experiment; **Figure 13**). Such a finding is consistent with the prediction of the biophysical model in Chapter 2.

One important note is that on the distractor trials in which people made binary choices, optimal decisions were all contributed by avoiding the available option with a lower value. However, in the three-option trials, a positive distractor effect on optimal decisions could be contributed by avoiding the second best and/or the worst distractor option. It would be highly counterintuitive if larger distractor values were related to more avoidance of the distractor option. In one analysis for the three-option trials, we showed that the kind of positive distractor effect we found in all our behavioural experiments was contributed by choosing the second best option less often when distractor value *relative* to the best option was larger (**Figure 12**), but at the same time the distractor per se was also chosen more often when it had a larger *absolute* value.

It is intriguing that our behavioural and biophysical model results made opposite predictions to the compelling divisive normalization model which has been used to describe firing in neurons that are related to decision value, attention, visual and other perceptual

information (Reynolds and Heeger, 2009; Carandini and Heeger, 2012). In this model, the negative distractor effect predicts not only that larger distractor is related to more distractor choices, but critically it also predicts that larger distractor are related to weaker preferences for the best option over the second best option (Louie et al., 2013). Although the absolute value effect of distractor in our three-option trials is also a kind of negative distractor effect (**Figure 12**), it is insufficient to support the divisive normalization account because it does not describe specifically how the preference between the best and second best options is affected. However, such a kind of effect was found in a study by Louie and colleagues (Louie et al., 2013). They tested human and monkey subjects with three-option decision making tasks, in which human subjects chose between food items of different subjective values and monkey subjects chose between stimuli that were associated with different amount of fluid reward. They found negative distractors effect in both humans and monkeys that conform to the divisive normalization account – more suboptimal decisions were made in the presence of larger distractor values.

Neuroscientific considerations, particular in relation to regional variations in neural coding, could provide some hints to explain the discrepancies between the results of our study and Louie's and colleagues' study (Louie et al., 2013). VmPFC and parietal cortex are two important regions for decision making (Shadlen and Newsome, 1996; Boorman et al., 2009; Kable and Glimcher, 2009; Hunt et al., 2012); however, there are two major differences between these two regions. Anatomically, there is a visuo-spatial topographic map in the parietal cortex. Neurons here only respond to options presented within part of the visual field, although their firing rates are normalized when more options are presented outside the response field (Goldberg et al., 1990; Sereno et al., 2001; Louie et al., 2011; Sereno and Huang, 2014). By contrast there is no evidence for a similar form of spatially topographic mapping in vmPFC or for the operation of normalization mechanism similar to those reported in parietal cortex. MEG data was collected in the whole brain while human

subjects were making binary decisions that sometimes required integration of decision variables on some trials (for example option A had a larger reward magnitude and a smaller probability than option B; similar to the incongruent trials in our study) but not on the other trials (option A had a larger reward magnitude and probability than option B; congruent trials) (Hunt et al., 2012). While both vmPFC and parietal cortex were active on the incongruent trials, the vmPFC was relatively inactive on the congruent trials. This suggests that the parietal cortex, which exhibits divisive normalization, could be more related to decisions in the congruent trials, whereas the vmPFC activity could be more related to incongruent trial decisions.

In a similar manner, we compared distractor effects on congruent and incongruent trials in the last analysis of this chapter (**Figure 17**). We showed that on congruent trials the distractor exerted a divisive normalization effect on decision making accuracy – there was a negative effect by the normalized distractor value on optimal decision between the two available options suggesting that as the distractor increased in value it became harder for subjects to choose accurately between the two other options. This is similar to the behavioural findings in the study by Louie and colleagues (Louie et al., 2013), in which humans and monkeys made decisions based on simple single attribute visual stimuli (these were associated with different foods in the humans and fluid quantities in the monkeys). By contrast on incongruent trials, there was an opposite effect of distractors – the effect predicted by the biophysical model in the previous chapter – larger distractor values had a positive impact on decision accuracy. It is therefore possible that the divisive normalization effects on decision-making are mediated by parietal cortex but when distractors exert the opposite effect that is predicted by the biophysical model then the effect is mediated by the vmPFC. This question will be addressed in Chapter 4 when fMRI data are discussed.

While the positive distractor effect on incongruent trials may cause failure of

optimal choice in some circumstances, it has the advantage of favouring "satisficing" choices when time for deliberation is limited (Simon, 1956) –choices that, while not ideal, are likely to be good given the options currently available. This is because failure of optimal choice only occurs on incongruent trials when HV and LV are often close in value and the identity of the optimal choice is not obvious. If time and resources for taking a decision are at a premium, as in many natural settings, then the positive distractor effect allows fast satisficing decisions to be made. It leads people to pick the second best option when the best and second best options are close in value and when the best and second options are both much better than another alternative.

Apart from the predictions of the biophysical model and the divisive normalization models, much work has been conducted to study decoy effects – a collection of phenomena that relate to the similarities in decision variables between a decoy/distractor option and two available options. Again these effects cause deviation from rational choice when decisions are made between the two choosable options (Usher and McClelland, 2004; Pettibone and Wedell, 2007). We argue that the distractor effects that we reported in this chapter are orthogonal to the decoy effects. One common design for studying decoy effects is that the two choosable options are made to have the same value. However, different combinations of decision variables are used to define the options on different trials and each option should be chosen half of the time. Although this is similar to the incongruent trials in our experiments, one key difference is that we avoided having choosable options with equal value in our experiment in order to have only one optimal choice at a time. It is generally the case that when a decoy option shares an attribute with a choosable option, then the decoy will prime people to select that option more often (Tversky and Simonson, 1993; Pettibone and Wedell, 2007). For example, the asymmetric dominance decoy effect involves two options with the same value (£30 for a nice lobster versus £27 for a steak) and a decoy with a *smaller integrated value* but with one attribute that it shares with one of the

choosable options (£27 for a tiny sandwich – i.e., a less valuable option but with the same price as the steak). In this case people's preferences are shifted in favour of the option (steak) that resembles the decoy in one of its attributes. However, in a study of the "phantom decoy" effect, which the decoy is unavailable like our distractor trials, a decoy with a larger value could also cause the same preference shift to a smaller value decoy. For example, people were more likely to choose the £27 steak than the £30 lobster in the presence of a "sold out" tasty seafood platter that costs £27 (same price but more appetitive than the steak).

In my study, the distractor was not designed in a way to manipulate similarity between the component attributes of choosable options and the distractor and instead we took special care to decorrelate the distractor's value from the values of the choosable options and value differences between the choosable. This also means that the effects of distractor value were orthogonal to predictions that might have been made on the basis of the decoy effect; regardless of any similarities in component options more optimal decisions were made when the *value* of a poor distractor became better and even more optimal decisions were made when the distractor value exceeded the best option (**Figure 16**). In contrast, previous investigations of the decoy effect predict that when one of the attributes of a decoy becomes more and more similar to those of an option A (in the presence of option B which has the same value but different attributes to option A), then this should cause people's preferences to shift to option A regardless of whether or not the value of the decoy is smaller or larger than the other options (Tversky and Simonson, 1993). It is possible that decoy effects had also influenced our subjects' decisions, however we argue that our distractor value effects are orthogonal to and are not confounded by decoy effects.

In summary, we found that the presence of a third option had several effects on decision making. First, in general, a poor third option is related to more suboptimal

decisions. This is a consequence taking more second best option choices (Figures 10, 11, 13), in a manner consistent with the biophysical model predictions in Chapter 2. Second, this kind of distractor effect is more robust on trials when integration of the component attributes that comprise an option in order to estimate its value is needed. In fact, the distractor effect reversed (larger distractors were related to more second best option choices) on trials when there was no demand for integration of decision variables (**Figure 17**). This is consistent with the divisive normalization prediction.

Chapter 4. Suboptimal Decisions with Multiple Alternatives: Neural Mechanism

Despite widespread interest in neural mechanisms of decision making, most investigations focus on decisions between just two options. Studies have shown that signals related to the value difference between options emerge during binary choice. A biophysical cortical network model (Chapter 2) predicts that the presence of third options with poorer values is related to weaker value difference signals. However, opposite predictions are made by the divisive normalization model. We tested human subjects on a decision making task in which choices were made between two options in the presence of an unchooseable third option, while undergoing functional magnetic resonance imaging (fMRI). We examined value-related signals in the ventromedial prefrontal cortex (vmPFC) and the parietal cortex. Consistent with the biophysical model prediction, the value difference signal in the vmPFC attenuated in the presence of a poor third option. In contrast, weaker value signals in the parietal cortex were related to larger third option values as predicted by the divisive normalization model. The size of the third option effect on the vmPFC and parietal signals were related to choice accuracies and reaction times respectively.

4.1 Introduction

Value representations are ubiquitous in the brain (Vickery et al., 2011), but there is special interest in vmPFC and intraparietal sulcus (IPS) because several lines of evidence suggest they use value information to guide a decision process (Boorman et al., 2009; Noonan et al., 2010; Camille et al., 2011; Wang, 2012). Not only do these areas carry information about reward expectation (Platt and Glimcher, 1999; Kable and Glimcher, 2007; Rangel et al., 2008; Kable and Glimcher, 2009; Lebreton et al., 2009; Bouret and Richmond, 2010; Louie and

Glimcher, 2010; Rushworth et al., 2011; Hunt et al., 2012; Monosov and Hikosaka, 2012), but vmPFC, and sometimes IPS, also carry "value difference" signals reflecting the difference in value between chosen and unchosen options during decision-making in a way that suggests value comparison between options is taking place (Boorman et al., 2009; FitzGerald et al., 2009; Philiastides et al., 2010; Rushworth et al., 2011; Kolling et al., 2012; De Martino et al., 2013).

Despite broad interest in the neural mechanisms of decision-making, little is known about the comparison process when there are multiple, and not simply two, alternatives to choose between. One important study of visually-guided decisions (evidence accumulation processes in value-guided and visually-guided decisions share many features (Philiastides et al., 2010; Hunt et al., 2012; Summerfield and Tsetsos, 2012)) explicitly compared multi-alternative and binary visual decisions but the degree of evidence for only one of the two/four possible choices was manipulated on each trial (Churchland et al., 2008). By contrast, in Chapter 2, we performed simulations using a biophysical model and investigated decisions where the value related to each of several potential options is parametrically varied on each decision so that the interactions between neural representations of several potential choices can be investigated. We made predictions that the strength of a value difference signal was attenuated in the presence of a third option with poor value.

Contrary to our biophysical model simulation, the compelling divisive normalization model makes opposite predictions about the effect of a third option on value-related signals (Louie et al., 2011). Under this account, the neuronal code for the value of an option is computed by the option value divided by the total value of all options presented in the environment (Carandini and Heeger, 2012; Louie et al., 2013). In cases of multiple alternative decision making, the model makes predictions that smaller value signals are related to larger third option values, as opposed to smaller third option values predicted by

the biophysical model.

In this chapter, we investigated how value signals were influenced by the third option in vmPFC and the medial IPS regions (MIP). Human MIP corresponds to macaque MIP (Mars et al., 2011) and is concerned with value-guided decisions (Mars et al., 2011; Hunt et al., 2012) and evidence accumulation(Gould et al., 2012) when choices are made with the hands. We found that poorer third options were related to weaker vmPFC signals that were related to the value difference of options, consistent with the biophysical model predictions. On the other hand, we found evidence that MIP signals related to the value of a chosen option in the contralateral visual field were modulated in a way that was consistent with the divisive normalization model (Louie et al., 2011).

4.2 Methods

Experimental subjects and task

The subjects' demographics and the design of the decision making task has been described in Study 5 of Chapter 3. In brief, the experiment involved twenty one subjects performing in the decision making task while undergoing fMRI scanning. The task involved 150 distractor trials and 150 two-option trials. On two-option trials, subjects made decisions between two available options with different combinations of reward magnitudes and probabilities. On distractor trials, there was an additional unchooseable distractor present and subjects had to choose between the other two available options.

Imaging data acquisition and preprocessing

Imaging data were acquired on a Siemens 3T MRI scanner. FMRI data were acquired with 3x3x3 mm^3 voxel-resolution, TR=3 s, TE=30 ms, flip angle=87°, slice angle =15°. Signal dropout in vmPFC was minimized by applying z-shimming at this region (Deichmann et al., 2003). Field maps were acquired using a dual echo 2D gradient echo sequence: 3x3x3 mm^3 voxel-resolution, TR=445 ms, TE 1=5.19 ms, TE 2=7.65 ms. T1-weighted structural images were acquired using an MPRAGE sequence: 1x1x1 mm^3 voxel-resolution, 174x192x192 grid, TR=2200 ms, TE=4.53 ms, TI=900 ms.

FMRI data were analyzed using FMRIB's Software Library (FSL) (Smith et al., 2004). Preprocessing was performed by applying high-pass temporal filtering (3 dB cutoff of 100 s), Gaussian spatial smoothing (full-width half maximum of 5 mm), motion correction (Jenkinson et al., 2002) and field map correction for signal distortion (Jenkinson, 2003). Motion artefact was removed by performing independent component analysis using FSL MELODIC. Each individual's fMRI data set was registered to the same individual's high-resolution structural image and then into the standard Montreal Neurological Institute (MNI) space using affine transformations (Jenkinson and Smith, 2001).

FMRI data analysis

Whole-brain analyses were conducted using a univariate GLM approach. At the group level, analyses were performed using FMRIB's local analysis of mixed effects (FLAME) (Beckmann et al., 2003; Woolrich et al., 2004), with outlier deweighting (Woolrich, 2008) and cluster-based thresholding criteria of $z>2.3$ and $p<0.05$ cluster-corrected(Worsley et al., 1992), unless otherwise specified.

To identify brain regions encoding value difference, we included the following regressors: chosen-unchosen value difference in two-option trials, chosen-unchosen value difference in distractor trials, and HV-D in distractor trials, time-locked to stimulus onset; left hand response, right hand response, time-locked to subjects' response; magnitude of reward feedback and feedback indicating inappropriate responses (mistakenly choosing a blank space or distractor) and slow responses, time-locked to feedback onset; and feedback on catch trials time-locked to its onset.

We extracted MIP and vmPFC signals from three-voxel radius spherical masks at coordinates from our previous studies of MIP (Mars et al., 2011) and vmPFC (Boorman et al., 2009; Kolling et al., 2012) (which corresponded to the peak effect in the current study). Coordinates in MNI space were transformed to subject space by affine transformation. The extracted time courses were 10 times oversampled by cubic spline interpolation. The time courses were then time-locked to onset of the initial phase and were analyzed by applying a GLM to every time point of each subject separately. The mean and standard error of the β weights of regressors were calculated across subjects so that time courses of effect sizes could be plotted. To illustrate the value difference signals in vmPFC and MIP we analyzed the distractor trials by including regressors of HV-LV, HV-D, HV+LV and magnitude of reward feedback in the GLM (Figure 18c, d). In Figure 22a we also included an additional (HV-LV)(HV-D) interaction term regressor. Note that because we focused on correct trials only, the HV-LV signal that we focus on corresponds to the chosen-unchosen value signal previously reported in vmPFC.

In addition, a second analysis approach was used in an attempt to look for divisive normalization of spatially specific value signals in parietal cortex. In this analysis we looked at modulation of the MIP signal corresponding to the value of a contralateral option (COpt signal). We then examined how the COpt signal changed as function of the value of stimuli

on the ipsilateral side of space (ipsilateral value: IV). If there is divisive normalization of a spatially specific value signal then the COpt signal will be reduced when IV is higher. To assess whether this was the case the following regressors were included in the GLM: COpt, IV, and the interaction between these two regressors COpt×IV (which served as our critical measure of modulation), magnitude of reward feedback, and a binary variable describing trial type. Error bars in all the figures represent standard error.

To investigate modulation of the HV-LV value difference signal by distractor (**Figure 19**a, b), we separated the distractor trials into four bins according to HV-D difference (>150, 0 to 150, -150 to 0, <-150). Four corresponding bins were identified in the two-option trials. The matched trials were, by design, identical to distractor trials in terms of available options (HV and LV) but obviously they lacked a distractor. We included in our analysis for a given participant only matched pairs of trials when correct responses were made in both trials. It is important to use only correct trials in the analysis because the vmPFC decision signal reflects the chosen-unchosen option value difference and this would differ between correctly and incorrectly performed versions of the same trial. We applied a GLM to each bin of every subject that included regressors of the HV-LV value difference and the magnitude of reward feedback. The peak HV-LV value difference signal in each bin was extracted in every subject within a 5-11 s time window from onset of the initial phase. Linear regression was performed to predict HV-LV signal size by the HV-D value difference rank associated with the four bins for each trial type. The β weight for HV-D rank was an index of modulation in HV-LV signal size by HV-D. A two-tailed one sample t-test compared the group's β weights with zero to determine consistency and significance of modulation.

A similar analysis was performed to investigate the divisive normalization effect on the COpt signal. We took correct trials where choices were made with the hand contralateral to the ROI and binned them according to IV. Effect sizes in each bin were derived from a

GLM with a regressor of COpt. The maximum COpt signal of each subject was extracted from a time window where group effects peaked (MIP, 12-18 s; vmPFC, 5-11 s).

We performed partial correlations to examine relationships between neural signals and different aspects of behaviour. We extracted the effect size of the impact of HV-D on the HV-LV signal (Figure 22a), as indexed by the maximum (HV-LV)(Op-D) signal (13-19 s), from every subject and correlated it with the subjects' accuracies in distractor trials. The control variables were sizes of HV-LV and HV-D signals and RT in distractor trials. In Figure 22c, we subtracted the maximum (HV-LV)(HV-D) signal in a 13-19 s time window from that in a 2-7 s time window to obtain the size of signal reduction and correlated this with subjects' distractor trial accuracies. A second analogous analysis in the divisive normalization analysis (Figure 25b) examined partial correlation between each subject's maximum effect size for the interaction term COptxIV (12-18 s) and mean RT. Control variables were sizes of COpt and IV signals and accuracy.

Psychophysiological interaction analysis

We performed a psychophysiological interaction (PPI) analysis (Friston et al., 1997) to examine how vmPFC coupled with other brain regions to encode HV-LV and HV-D value differences. Each PPI regressor was generated by multiplying the demeaned physiological regressor (time course of vmPFC BOLD) by the demeaned, convolved psychological regressor (either HV-LV or HV-D value differences in distractor trials). We entered the following regressors into the GLM: time course of vmPFC, HV-LV value difference in distractor trials and its PPI with vmPFC, HV-D value difference in distractor trials and its PPI with vmPFC, HV-LV value difference in two-option trials, left hand response, right hand response, number of stimuli presented, magnitude of reward feedback, feedback on the catch trials and

feedbacks indicating inappropriate and slow responses. We initially used a cluster-based threshold ($z > 2.3$, $p<0.05$; Figure 23a) and then a more stringent cluster-based thresholding criterion of $z > 2.7$, $p<0.05$ (Figure 23b) to reduce the possibility of any false positive results in follow-up analysis.

4.3 Results

Impact of distractor value on neural signals of value

We investigated whether distractor value had an impact on neural signals for decision-making in a way that our biophysical model in Chapter 2 would have predicted. There is a vmPFC blood oxygen level dependent (BOLD) "value difference signal" (HV-LV) when subjects choose correctly (Boorman et al., 2009; FitzGerald et al., 2009; Philiastides et al., 2010; Rushworth et al., 2011; Hunt et al., 2012; Kolling et al., 2012; De Martino et al., 2013). The first analysis therefore looked for regions where BOLD was positively correlated with HV-LV in both correctly performed two-option and distractor trials. It identified a similar vmPFC region to that seen in previous studies (Figure 18a).

The time course of the vmPFC HV-LV value difference signal in distractor trials was extracted at coordinates from previous studies (Boorman et al., 2009; Kolling et al., 2012) (Figure 18c) and used in analyses with orthogonal contrasts. We also extracted the HV-LV signal in the MIP region of interest (ROI) at a coordinate investigated by another study (Mars et al., 2011) (Figure 18b,d). The biophysical model (Chapter 2) predicted the strength of HV-LV value difference signals should be attenuated with low distractor values, so we tested whether the same was true in the brain. We divided the distractor trials into four bins according to HV-D

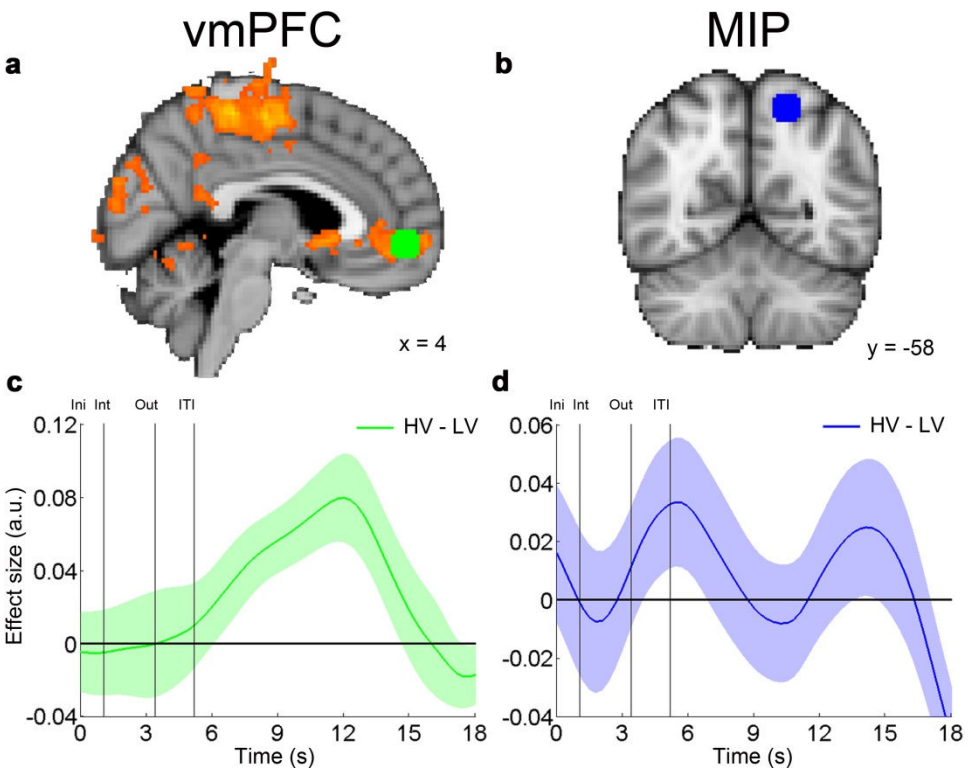


Figure 18. *Value difference signals in vmPFC and MIP.* ***(a)*** *As in previous experiments there was a prominent value difference effect between available options in vmPFC (sagittal section). An ROI in this region (green) was defined for the subsequent time course analyses.* ***(b)*** *An ROI in MIP (blue) with coordinates taken from Mars et al. (Mars et al., 2011) (coronal section).* ***(c)*** *Time course of HV-LV value difference signal in correct trials in vmPFC and* ***(d)*** *MIP.*

Ini, initial phase; Int, interval phase; Out, outcome phase; ITI, inter-trial interval.

Shadings in ***(c, d)*** *denote s.e.m.*

and plotted peak HV-LV effect size in each bin (Figure 19a) and tested whether there was a consistent decrease in peak effect size across subjects as HV-D increased (consistent negative slope for each individual's best fitted line in Figure 19a). Consistent with model prediction, vmPFC HV-LV value difference signals were weaker when distractor values were smaller (i.e. HV-D was larger; average slope: β =-0.065, t_{20}=-3.200, p=0.005; Figure 19a).

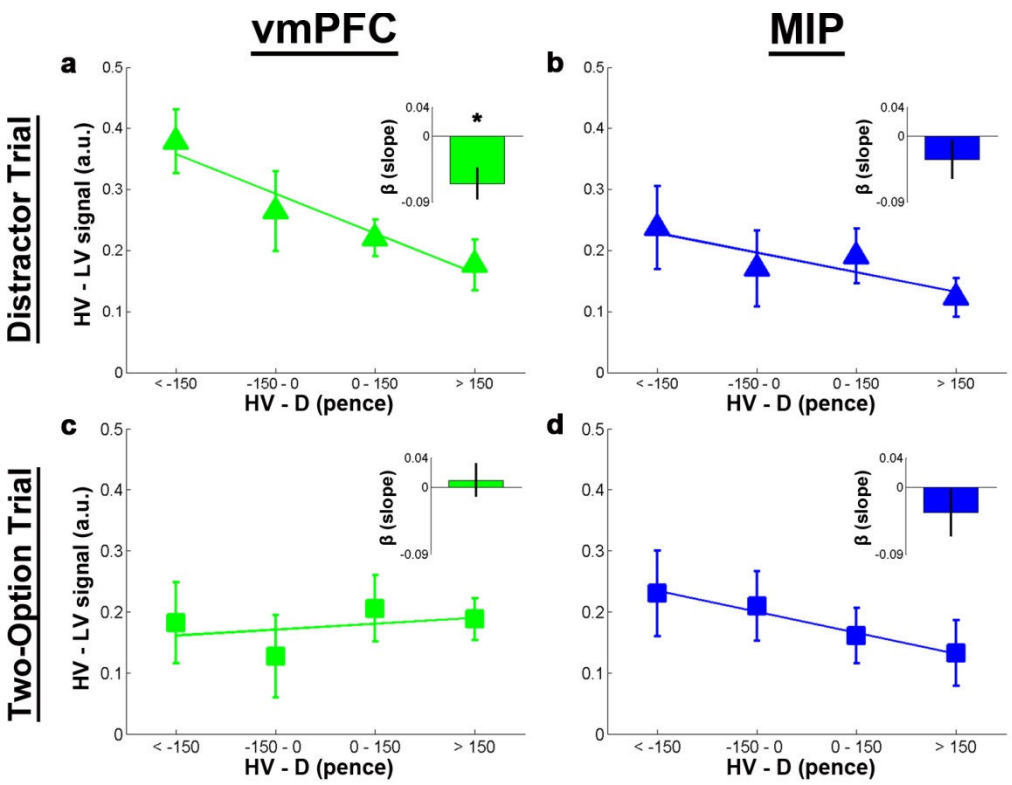


Figure 19. *VmPFC value difference signal was weaker when distractor value was small.* ***(a)*** *When the HV-LV value difference signal was binned as a function of the difference in value between available options and the distractor (HV-D) it was clear that, in vmPFC, it declined significantly as HV-D increased (in other words as D was lower) in distractor trials* ***(c)*** *but not in matched two-option trials.* ***(b, d)*** *No modulation, however, was seen in MIP in either distractor trials or matched Two-option trials.*

Error bars in s.e.m.

** Effect size of the slope, p=0.005*

To demonstrate that change in HV-LV signal across bins was really due to changing distractor values rather than some confound in our design, we also binned the matched two-option trials in an identical manner. It was therefore possible to examine HV-LV signals in two-option trials as a function of the absent distractor value that would have been present on its matched distractor trial (**Figure 19**c). As an additional precaution, we only analyzed trials performed correctly in both two-option and distractor trial conditions. This ensured

that any HV-LV signal modulation differences between conditions could not be attributed to differences in choices. Since we did not restrict our biophysical model analyses in Chapter 2 by models' choice, here we confirmed that the distractor effects in the biophysical model remained present even when analysis was restricted to correctly performed, difficult trials (**Figure 20**). Note that, just as it cannot predict other aspects of behavioural change between two and three choice situations, the model cannot predict the relative heights of the lines in figures 19a and 19c (without knowledge of how neurons are assigned to different options in the brain) but it does predict that only the line in 19a should be sloped and this is what we found. There was no significant effect of the absent "distractor" on two-option trials (β=0.010, t_{20}=0.451, p=0.657; **Figure 19**c); and effects in two-option and distractor conditions differed significantly (t_{20}=2.211, p=0.039). HV-LV signal weakening in vmPFC in distractor trials is a consequence of distractor presence.

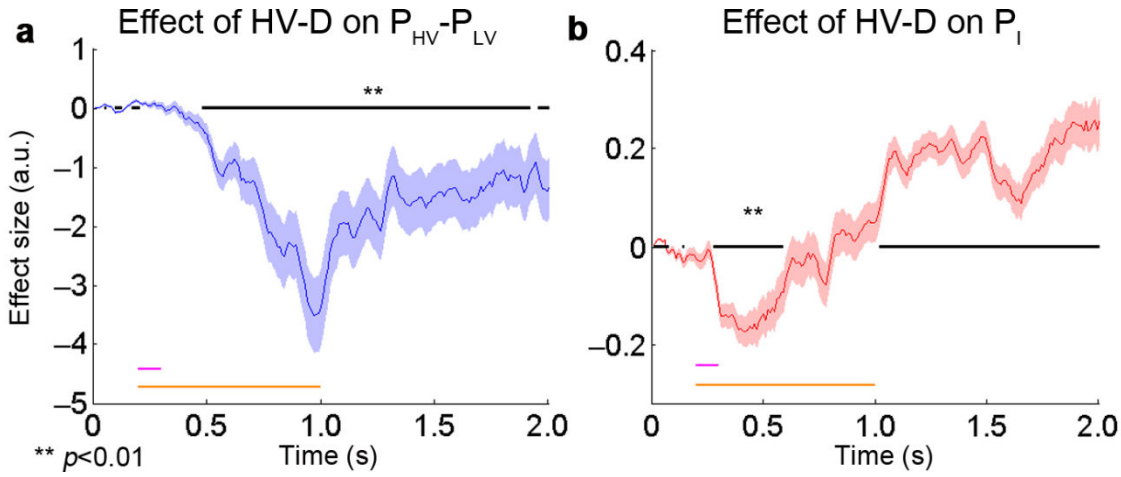


Figure 20. *Biophysical model predictions for correct difficult trials. A similar impact of HV-D* ***(a)*** *on the firing difference between HV and LV populations, P_{HV} and P_{LV}, and* ***(b)*** *on inhibitory firing P_I could also be seen when we analyzed only difficult trials on which the biophysical model "selected" the correct option.*

Shadings denote s.e.m.

Notably, the impact of distractor values on HV-LV signals in MIP did not reach significance on either distractor (β=-0.032, t_{20}=-1.271, p=0.218; **Figure 19**b) or two-option trials (β=-0.034, t_{20}=-1.121, p=0.276; **Figure 19**d), nor was there any difference between conditions (t_{20}=-1.230, p=0.233). An ANOVA comparing change in HV-LV signals in MIP and vmPFC in two-option and distractor conditions confirmed a significant interaction ($F_{1,20}$=6.104, p=0.023). In additional control analyses, we also showed that the effect of HV-D on vmPFC signal was neither driven by variance in HV (**Figure 21**a-d) nor HV+LV (**Figure 21**e-h). Our results suggest that distractor impact on accuracy might be mediated via vmPFC rather than parietal cortex.

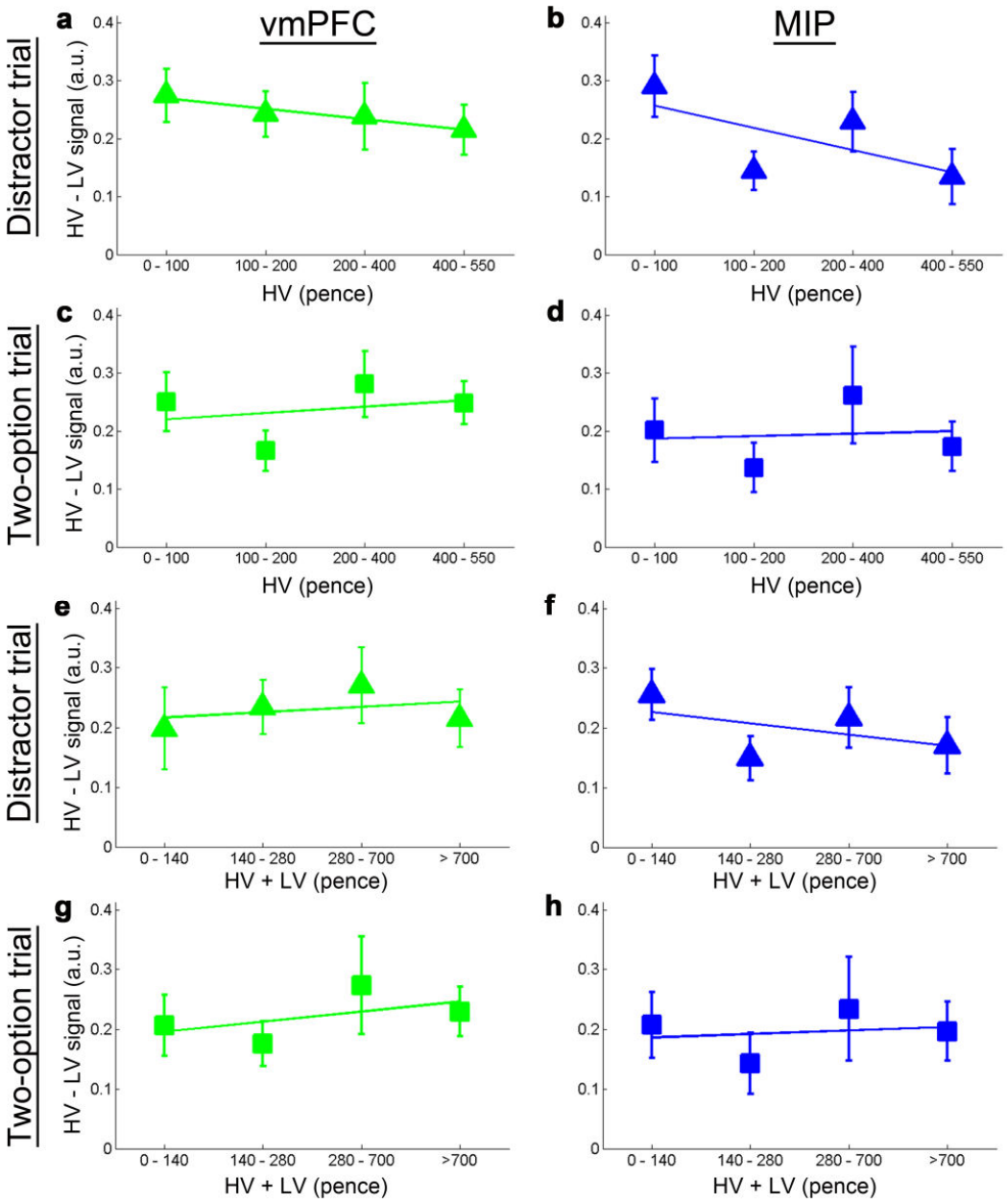


Figure 21. *HV-LV signal was not modulated by values of HV or HV+LV .* ***(a-d)*** *When the HV-LV value difference signals were binned as a function of the value of HV, there was no significant change in effect size in either type of trial in vmPFC or MIP.* ***(e-h)*** *Similarly, when we binned the HV-LV value difference signals as a function of the overall option value (HV+LV), there were no significant changes in effect size in either type of trial in vmPFC or MIP.*

To test the possibility that the HV-LV value signal in vmPFC became smaller when the HV-D value difference was bigger we looked at vmPFC value difference signals in a

complementary way. Instead of binning HV-LV signals by HV-D, we carried out an analysis of vmPFC BOLD looking at the interaction term (HV-LV)(HV-D) (**Figure 22**a). A key strategy to achieve high accuracy in distractor trials is to focus on the most relevant value comparison – that between the available options (HV-LV). Incorporating the irrelevant HV-D information into this comparison (i.e. a signal for (HV-LV)(HV-D) interaction) could result in less accurate representation of the relevant value difference and this may lead to suboptimal decisions. We examined individual differences in average accuracies in distractor trials to see if they were related to individual differences in the vmPFC (HV-LV)(HV-D) signal by a partial correlation analysis, with RT and HV-LV, HV-D signals as controlling factors. Individuals with

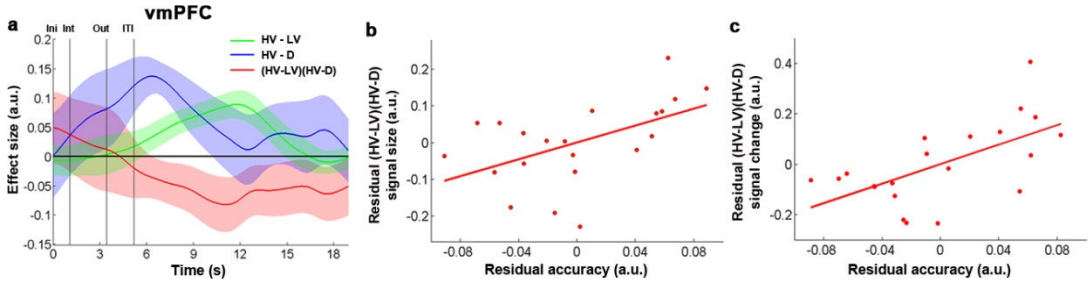


Figure 22. *Distractor impact on accuracy could be mediated via vmPFC.* ***(a)*** *The distractor-induced modulation of vmPFC HV-LV signals can be seen by a complementary analysis that included regressors corresponding to both the value difference between available options (HV-LV) and to the value difference between available options and the distractor (HV-D) and their interaction (HV-LV)(HV-D).* ***(b)*** *Individuals with more negative (HV-LV)(HV-D) interaction signals, especially in a late 13-19 s time window, were associated with less accurate decisions, after controlling for the effects of HV-LV, HV-D signals and RT.* ***(c)*** *It appeared that failure to reduce the size of the impact of the (HV-LV)(HV-D) interaction term on vmPFC BOLD activity was associated with inaccurate choice; less reduction of the negative (HV-LV)(HV-D) signal in this late time window compared to an earlier 3-7 s time window was associated more inaccurate decisions.*

Shadings in ***(a)*** *denotes s.e.m.*

lower accuracies had more negative (HV-LV)(HV-D) signals (r=0.512, p=0.030; **Figure 22**b). The relationship between the (HV-LV)(HV-D) signal and inaccurate behaviour was especially strong at late time periods suggesting prolonged (HV-LV)(HV-D) modulation of vmPFC BOLD led to suboptimal decisions. To further test this hypothesis, we extracted this signal from an earlier period and calculated how much the signal size changed. Individuals that showed stronger *reduction* of the negative (HV-LV)(HV-D) signal had higher accuracies (r=0.607, p=0.008; **Figure 22**c).

Psychophysiological interactions of vmPFC

So far we have referred to vmPFC as a decision-maker but an alternative account emphasizes its role in attentional selection (Noonan et al., 2010; Lim et al., 2011). The finding that variation in accuracy is related to variation in vmPFC (HV-LV)(HV-D) signal is consistent with a variant of the latter hypothesis in which vmPFC does not just compare values of potential choices but selects key choices that might be compared as opposed to irrelevant distracting, albeit value-laden, information in the environment.

We looked for further evidence for whether vmPFC might, respectively, facilitate or suppress representation of information relevant or irrelevant for the value comparison at hand. We carried out a psychophysiological interaction (PPI) analysis of all trial data seeking brain regions in which BOLD changed its coupling with vmPFC as a function of the irrelevant value contrast (HV-D). We found a large swathe of central orbitofrontal cortex (cOFC) extending into ventrolateral prefrontal cortex to be negatively coupled with vmPFC as a function of HV-D (cluster-forming threshold $z>2.3$, $p<0.05$ cluster-corrected, Figure 23a). Unlike vmPFC, cOFC is less important for value comparison but instead lesion, single neuron recording, and fMRI studies (Izquierdo and Murray, 2004; Izquierdo et al., 2004; Bouret and

Richmond, 2010; Noonan et al., 2010; Walton et al., 2010; Rushworth et al., 2011) suggest it represents precise associations between stimuli and specific outcomes. Here, we extend this to suggest that the relevance of the stimulus-reward representation encoded in cOFC may affect its coupling with vmPFC. A second part of the PPI analysis, focusing on the same cOFC region that negatively coupled with HV-D, found an area of positive coupling with vmPFC as a function of the relevant value comparison, HV-LV (cluster-forming threshold $z>2.7$, $p<0.05$ cluster-corrected, Figure 23b, c).

Causality cannot be inferred from PPI analysis, and it is unclear whether the results reflect an increased influence of vmPFC over cOFC, or vice versa. The former interpretation is consistent with an otherwise surprising finding (Noonan et al., 2010) that lesions in this vicinity sometimes cause macaques to become worse at choosing as the difference in value

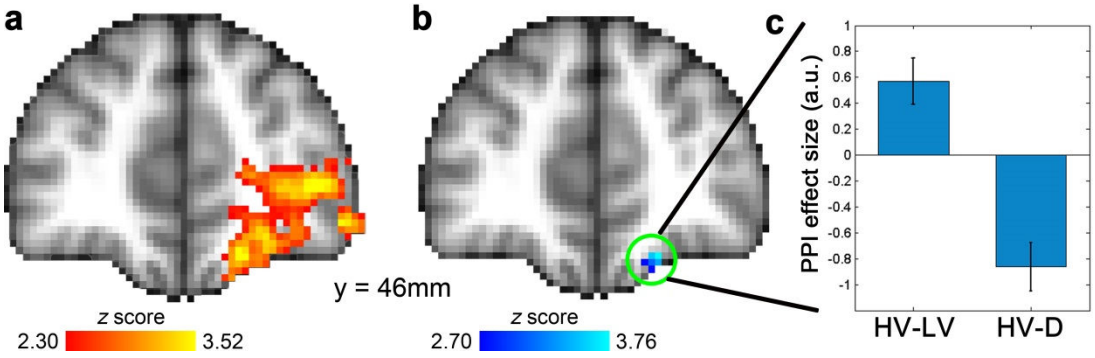


Figure 23. *Psychophysiological Interaction (PPI) results.* ***(a)*** *Coupling with vmPFC as a function of the negative effect of the irrelevant comparison between available options and the distractor (HV-D) increased in central orbitofrontal and ventrolateral frontal cortex. Representations associated with the irrelevant comparison HV-D were reduced in tandem with vmPFC activity in these regions that are important for representation of individual stimuli and their associated outcomes.* ***(b)*** *Coupling with vmPFC as a function of the positive effect of the decision-relevant value difference (HV-LV) increased in part of the same region.* ***(c)*** *Positive effect of coupling as a function of HV-LV and negative effect of coupling as a function of HV-D with vmPFC.*

between two better options and a third, lower value item increases (Noonan et al., 2010) whereas normally such differences make decision-making easier. Such a deficit might be expected if vmPFC lesions prevented focusing on the most relevant comparisons when making decisions. Other lesion findings might be interpreted in a related manner (Fellows, 2006).

Divisive normalization of spatial value signals in MIP

Our demonstration that decision-making accuracy (Chapter 3) and vmPFC value difference signals (**Figure 19**) are reduced in the presence of low value distractors seems at odds with an influential account of multi-alternative decision-making (Louie et al., 2011; Carandini and Heeger, 2012). It argues that divisive normalization, a standard form of gain control operating in sensory systems and which may explain attentional modulation in sensory systems (Lee and Maunsell, 2009, 2010), also impacts on choice valuation. According to this account, value signals are diminished by *high*, rather than low, value distractors. For example, IPS neuron activity is correlated with the value of saccades to targets in their receptive fields but the activity is normalized by values of alternative targets outside the receptive field (Louie et al., 2011); the activity associated with a target is *reduced* by a greater amount if there is a *high* value alternative elsewhere in the visual field than if there is a low value alternative.

It is, however, important to remember that divisive normalization is prominent in sensorimotor areas. Typically neurons exhibiting divisive normalization, like those in IPS, have spatially-specific receptive fields and there is evidence that IPS selects spatial locations for behavioural priority (Sugrue et al., 2005). By contrast, range normalization when reported in prefrontal cortex appears on the timescale of sessions, and to a lesser extent on

individual trials (Padoa-Schioppa, 2009). It is therefore possible that trial-to-trial divisive normalization in parietal cortex might be apparent if we look for *spatially-selective* value signals.

In the next analysis, rather than looking for activity covarying with the chosen option's value irrespective of its spatial position, as we had done up to now, we looked for activity covarying with the value of choices made to a particular side of space (Palminteri et al., 2009). We focus here on MIP and vmPFC signals covarying with values of correct choices made with the *contralateral* hand. We tested whether such contralateral option (COpt) choice signals were diminished when the *total value* of the options associated with the *ipsilateral* response (ipsilateral value: IV) was higher. Consistent with divisive normalization accounts (Louie et al., 2011) the MIP COpt signal was significantly diminished as IV

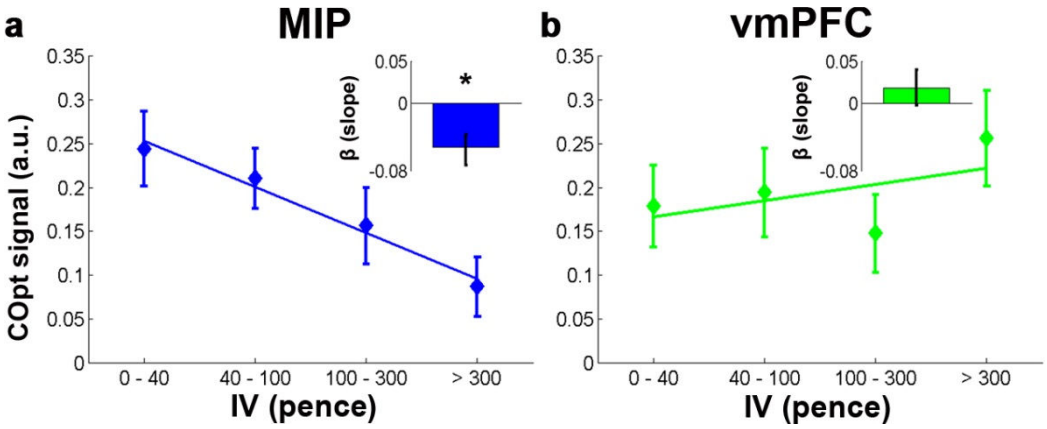


Figure 24. *Divisive normalization of spatially specific contralateral option (COpt) choice signals in MIP. Value signals specific to contralateral responses* ***(a)*** *decreased as the total value of stimuli (ipsilateral value: IV) associated with the other, ipsilateral, response increased in MIP (b) but not vmPFC.*
Ini, initial phase; Int, interval phase; Out, outcome phase; ITI, inter-trial interval.
Error bars in ***(a, b)*** *denote s.e.m.*
** p=0.003*

increased (β=-0.053, t_{20}=-3.490, p=0.002; **Figure 24**a). No such effect, however, was seen in vmPFC (β=0.015, t_{20}=0.684, p=0.502; **Figure 24**b).

We also found that RTs varied in the manner predicted by a divisive normalization account. RTs increased as IV increased (β=0.142, t_{20}=3.240, p=0.004; in other words subjects were slower when the total value of options was larger in the hemifield opposite to the COpt), whereas RTs were unaffected by values of unchosen options or distractor when they were on the same side of the COpt (β=-0.028, t_{20}=-0.450, p=0.658). Next, we therefore investigated whether individual variation in RT was related to individual variation in the divisive normalization of MIP COpt signals. This also provides a way of testing whether MIP activity is actually related to task performance. We first estimated the impact of distractor value on the COpt signal in each subject by looking at the COpt×IV interaction. As expected, the interaction term was associated with a negative modulation of the BOLD signal suggesting that the COpt signals diminished as a function of IV (***Figure 25***a). Across subjects variation in the COpt×IV interaction term was negatively correlated with RT (r=-0.485, n=21, p=0.042; ***Figure 25***b) after controlling for accuracy and other neural signals.

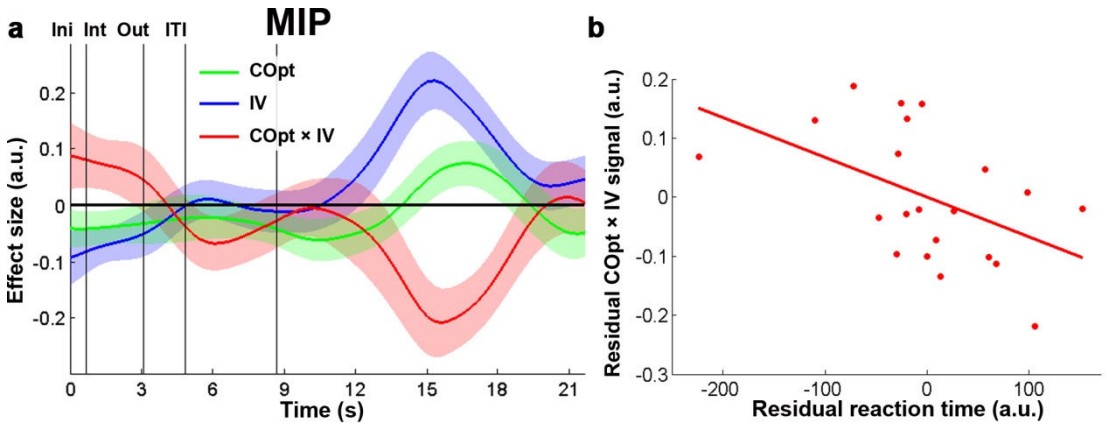


Figure 25. *An analysis quantifying in each subject the size of the modulation of the spatially specific value signal (COpt) in MIP by including a regressor indexing the total value of stimuli associated with the ipsilateral response (IV).* ***(a)*** *There was a negative effect of the interaction term COpt×IV suggesting that MIP spatial value signals were diminished in the presence of value associated with the ipsilateral response.* ***(b)*** *The size of the COpt×IV effect was negatively correlated with RT suggesting that greater divisive normalization of spatial value signals in MIP was associated with slower RTs, after controlling for COpt, IV signals and accuracy.*

Ini, initial phase; Int, interval phase; Out, outcome phase; ITI, inter-trial interval.

Shading in ***(a)*** *denotes s.e.m.*

4.4 Discussion

In a multialternative decision making task, we found signals in the vmPFC and MIP that made value comparisons between task relevant options. In Chapter 2, the biophysical model simulations predicted that value comparison signals diminished in the presence of an irrelevant third option. Interestingly, we found that the value difference signals in the vmPFC, but not the MIP, interacted with the distractor in a way that was consistent with the biophysical model prediction. In other words, the presence of irrelevant third options with lower values was related to weaker value comparison signals for the relevant and choosable options in the vmPFC. In Chapter 3, we showed that smaller third option values were associated with more suboptimal choices in human behaviour. The pattern of vmPFC

activity that is reported here could mediate the disrupted behavioural pattern. In support of this contention we also found that individuals with stronger distractor effects in their vmPFC value comparison signals were also the ones that made more suboptimal choices (**Figure 22**b). This provides evidence for the relevance of the vmPFC signals, at a neural level, to the actual choices that were made by subjects at a behavioural level.

It has been argued that the vmPFC compares value during decision making via a mutual inhibitory mechanism. In Chapter 2, we have explained how a biophysically plausible model of a neural network receives inputs from options and generates decisions through a process of competition, mediated by inhibition, between two pools of neurons representing each option. Various studies have suggested that this could be the working mechanism for the vmPFC. Perhaps the most direct evidence comes from a recent single neuron recording study. When monkeys were presented with two options to choose between, there was a substantial number of value-sensitive neurons in the vmPFC that responded to both of the options but in an opposite manner (e.g. if the firing of a neuron was correlated positively with the value of option A it was also negatively correlated with value of option B), suggesting mutual inhibitory competition between neurons (Strait et al., 2014).

A binary choice version of the biophysical model also predicts neural activity that is similar to the vmPFC activity obtained from human MEG recording (Hunt et al., 2012). One feature of this class of biophysical model is that optimal, non-stochastic choices are dependent on the efficiency of the mutual inhibition mechanism. In humans, it is not possible to take time-resolved measures of neurotransmitters to test whether lower levels of inhibition occur when the option have low values as the model predicts. Average vmPFC GABA concentrations, however, can be measured and lower concentrations are associated with poorer performances in two-option decision-making (Jocham et al., 2012). Furthermore, the biophysical model in Chapter 2 made predictions that smaller third option

values were related to weaker inhibitory activity that reduced value comparison signals. Although inhibitory activity could not be measured directly in our fMRI data, our results showing that weaker vmPFC value comparison signals were related to smaller third option values (**Figure 19**a) provide additional evidence for a mutual inhibitory mechanism for "decision making" in the vmPFC.

One way of understanding the interaction signal in the vmPFC (**Figure 22**a) is that the distractor is incorporated in the value comparison signal. In other words, the vmPFC was making 1) task relevant comparisons between available, choosable options and 2) task irrelevant comparisons between available and unavailable options and 3) the two types of comparisons interfere with each other (interaction). We found that subjects with stronger interaction signals made more suboptimal choices. However, this correlation was more robust when we investigated the signal size in a late time window rather than an early time window (**Figure 22**b), suggesting that even for high performers there was a transient distractor interference in their value comparisons. They were more capable of getting rid of the interference from the irrelevant value comparison and focus on the relevant value comparison in their vmPFC (**Figure 22**c). This would require a selection mechanism for identification of task relevant comparisons. In our PPI analysis, we found that the coupling between vmPFC and cOFC/ventrolateral PFC (vlPFC) was stronger as a function of HV-LV and weaker as a function of HV-D – these two regions were synchronized for the coding of the relevant comparison and desynchronized for the coding of an irrelevant comparison. It is possible that the cOFC/vlPFC has a role in selecting task relevant signals and inhibiting task irrelevant signals. A similar region is important for making flexible changes to generate adaptive behaviour when the environmental contingencies that determine reward change (Butter, 1969; Hampshire and Owen, 2006; Roberts, 2006). Here for the first time we report that the cOFC/vlPFC could be more strongly coupled with the vmPFC for task relevant value comparison during decision making.

It is not possible to describe the impact of the distractor on vmPFC value difference signals in terms of divisive normalization accounts such as those proposed by Louie and colleagues (2011) to describe distractor effects in parietal cortex, even if vmPFC neurons with activity that is positively and negatively related to value are intermingled as has been reported in other frontal lobe regions (Wallis and Kennerley, 2010) . Critically, however, when the fMRI data were re-analyzed, a *spatially selective* value signal – a contralateral chosen value signal specific to one response side – was found in MIP and its size was reduced in proportion to the value of alternatives on the other response side (**Figure 24**a) just as predicted by divisive normalization accounts (Louie et al., 2011). Behavioural tasks requiring more spatially resolved responses (reaching movements to four different locations corresponding to the different potential stimulus quadrants) and employing more spatially resolved recording of brain signals might be necessary to identify a normalizing effect of stimuli presented in a different quadrant in the same hemifield as the Copt (Louie et al., 2011).

While our version of the biophysical model predicts both subjects' choices and vmPFC activity patterns, it is possible that biophysical models of other networks might produce the activity patterns both we and Louie and colleagues (Louie et al., 2011) observed in parietal cortex. One candidate is the biophysical model that simulates neural network activities in a random dot motion task (Furman and Wang, 2008; Wei et al., 2012). The model differs from the early Wang model (Wang, 2002) by having 1) multiple excitatory populations that exhibit different tuning preference to motion direction and 2) excitatory-excitatory connections in which the strengths are dependent on the similarity in tuning preference of the neurons being connected. This kind of design in the network mimics the topographic representation of parietal neurons and allows multiple attractor states as "bump attractors" in the network. When such a network is applied in value-based decision making, as oppose to the original applications simulating working memory or perceptual

decision making, it seems likely that as the value of one item is increased it would, via a shared inhibitory interneuron pool, have a divisive normalization effect on the other bump attractors representing other responses.

That the current biophysical model provided a better description of vmPFC than IPS activity it is unlikely to reflect some vagary of task design. Care was taken to index values with conjunctions of simple visual features known to rely on parietal cortex (Ashbridge et al., 1997; Sereno and Maunsell, 1998) . Similar speeded response requirements were used in monkey experiments that reported parietal value signals (Platt and Glimcher, 1999; Louie and Glimcher, 2010). Moreover, training of the sort subjects undertook before scanning leads to greater IPS activity during action selection (Grol et al., 2006). That we were able to find value signals in MIP attests to the appropriateness of our protocol. Instead, the results suggest fundamental differences between vmPFC and IPS decision-making mechanisms. VmPFC neurons may employ more abstract and flexible coding principles than do spatially selective IPS neurons. Little is known of vmPFC neuron properties but what is known suggests differences from cOFC (Bouret and Richmond, 2010; Noonan et al., 2010; Kaping et al., 2011; Noonan et al., 2011; Rudebeck and Murray, 2011b; Rushworth et al., 2011; Monosov and Hikosaka, 2012; Noonan et al., 2012; Wallis, 2012).

We note that previous studies have presented evidence for encoding of the difference in value between options (Seo et al., 2009) and for value normalization (Louie et al., 2011) in the intraparietal sulcus. Here we provide some evidence that the human intraparietal region encodes the value difference between options (**Figure 18**d) but that spatial value signals are normalized in the same brain region even by the values of irrelevant stimuli (**Figure 24**).

Our results resonate with the notion that IPS and vmPFC constitute two components of a system for value-based choice (Kable and Glimcher, 2009) but suggest both components

might simultaneously attempt to select the course of behaviour. That individual variation in both vmPFC (**Figure 22**b) and MIP (***Figure 25***b) signals was correlated with individual variation in two distinct aspects of behaviour supports this view. Moreover such a view is consistent with a demonstration that high value distractors, in a non-spatial task, sometimes only disrupted decision-making in the manner divisive normalization accounts predict after vmPFC lesion (Noonan et al., 2010). This could be explained if, in the absence of vmPFC, the divisive normalization effect of distractors on parietal activity had a greater impact on behaviour. Such findings may contribute to a growing understanding of failures of optimal decision-making by suggesting that situational variables that favour either parietal or vmPFC control of behaviour may determine patterns of deviation away from optimal decision-making.

Chapter 5. Contrasting Roles for Orbitofrontal Cortex, Anterior Cingulate Cortex, and Amygdala in Credit Assignment, Learning, and Exploration in Macaques.

Recent studies have challenged the view that orbitofrontal cortex (OFC) and amygdala mediate flexible reward-guided behaviour and suggested their roles are similar to those of other frontal areas, such as anterior cingulate cortex (ACC). We trained macaques to perform an object discrimination reversal task during functional magnetic resonance imaging sessions, and identified a lateral OFC (lOFC) region in which activity predicted adaptive win-stay/lose-shift behaviour. In contrast, ACC activity predicted future exploratory decisions regardless of reward outcome. Amygdala and lOFC activity was more strongly coupled on lose-shift trials. However, lOFC-amygdala coupling was also modulated as a function of the relevance of reward information in a manner that may ensure that only relevant reward information is used during credit assignment. Day-to-day fluctuations in signals and signal coupling were correlated with day-to-day fluctuation in performance. Our data demonstrate OFC, ACC, and amygdala each make unique contributions to flexible behaviour and credit assignment.

5.1 Introduction

The orbitofrontal cortex (OFC) together with the amygdala have long been thought to mediate changes in behaviour, particularly those that are guided by changes in the reward environment (Butter, 1969; Jones and Mishkin, 1972; Aggleton and Passingham, 1981; Spiegler and Mishkin, 1981; Izquierdo et al., 2004; Paton et al., 2006; Roberts, 2006;

Rudebeck and Murray, 2008; Morrison et al., 2011). In experimental settings such behavioural change has often been studied in the context of object discrimination reversal (ODR) tasks in which reward is first assigned to one stimulus but not another. After animals reach a high level of responding to the rewarded stimulus the reward assignment is switched and the reversal occurs; the previously rewarded stimulus is no longer rewarded and the previously unrewarded stimulus is now rewarded.

The links between OFC, amygdala, and reward-guided behavioural change have recently been challenged. Rudebeck and colleagues (2013b) reported that OFC lesions do not cause macaques to perform poorly on reversal tasks if the lesion is made by neurotoxin injection rather than by aspiration suggesting the reversal task impairments seen after OFC lesions are actually the consequence of deafferentation of an adjacent brain region but not of damage to the OFC *per se*. The identity of the critical brain region is unknown. Although OFC neuron activity has been reported during stimulus-reward reversal learning (Rolls, 2000; Morrison et al., 2011) it might be argued that had another frontal region been investigated then the proportion of neurons encoding critical task variables might have been higher. An additional challenge to the notion that OFC has a unique role in the flexible assignment of rewards to stimuli has been the report that lesions in another frontal cortical region, anterior cingulate cortex (ACC), impair stimulus-reward reversal learning (Chudasama et al., 2013).

A similar challenge has been mounted on the idea that the amygdala is important for ODR. Neurotoxic lesions in the amygdala have been reported not to affect (Izquierdo and Murray, 2007) or even to improve ODR performance (Rudebeck and Murray, 2008; Izquierdo et al., 2013).

Adding to the confusion are observations that while neurotoxic lesions do not cause reversal impairments in an old world primate like the macaque they can impair reversal

tasks in rodents (Schoenbaum et al., 2007) and new world monkeys such as marmosets (Roberts, 2006) . However, even in some of these species the role of the amygdala has been questioned. Stalnaker and colleagues (2007) showed that the reversal impairments caused by bilateral OFC lesions in the rat were abolished by bilateral lesions of the amygdala. The results suggest the surprising conclusion that OFC, at least in the rat, is counteracting or suppressing some feature of amygdala activity that disrupts rapid reward-guided behavioural change. Such a conclusion, however, is at odds with other claims based on neurophysiological data that amygdala and OFC carry related signals at short latency differences suggestive of inter-areal interaction and collaboration and the exchange of information during behavioural change (Morrison et al., 2011).

We attempted to reconcile these conflicting findings by recording activity throughout the whole brain using functional magnetic resonance imaging (fMRI) in four macaque monkeys while they performed a deterministic ODR task. In this way we hoped to identify frontal cortical regions carrying signals needed for reward-guided behavioural reversal in an old world primate in which behavioural change cannot be mediated by verbal or linguistic strategies that in human will depend on adjacent ventrolateral frontal cortical regions associated with language. In addition, because we recorded activity from the whole extent of OFC, amygdala and ACC simultaneously from all four animals, we could compare OFC and ACC and examine interactions between OFC and amygdala. In brief we identify a region in area 12/47 we refer to as lateral OFC (lOFC) that lies just lateral to the focus of many OFC investigations in the macaque that carries signals needed for behavioural change. We also found behavioural change signals in ACC and amygdala distinct to those in lOFC. Finally, we also identified two distinct types of lOFC-amygdala interaction that might be labelled positive and negative and which might be linked to exchange and suppression of signals respectively. We contend that the interaction between lOFC and amygdala changes

dynamically to ensure accurate stimulus-reward credit assignment and to avoid the integration of irrelevant reward information in estimates of stimulus-reward association.

5.2 Methods

Subjects

Four male rhesus monkeys (*Macaca mulatta*) were involved in the experiment. They weighed 10.4-11.9kg and were 7 years of age. They were group housed and kept on a 12-h light dark cycle, with access to water 12-16h on testing days and with free water access on non-testing days. All procedures were conducted under licences from the United Kingdom (UK) Home Office in accordance with the UK The Animals (Scientific Procedures) Act 1986. One testing session from one animal was excluded due to excessive motion during scanning.

Behavioural Training

Prior to actual data acquisition, all animals were trained to work in an MRI compatible chair in a sphinx position that was placed inside a custom mock scanner that simulated the MRI scanning environment. They were trained to use home-made infra-red touch sensors to respond to stimuli presented on a screen. They learned to perform the ODR task (sometimes called a stimulus reversal task). The mock scanner training was considered complete once the animals performed with greater than 75% accuracy for at least three consecutive training sessions. The animals underwent aseptic surgery to implant an MRI compatible head post. After a recovery period of at least two months, the animals were trained to perform the task inside the actual MRI scanner under head fixation. The imaging

data acquisition started once they performed at above 75% accuracy for at least another three consecutive sessions in the scanner.

Experimental Task

In our deterministic ODR task, subjects needed to choose repeatedly between two stimuli that were novel in each testing session (Figure 26a). Each trial began with a blank screen (inter-trial interval; 5-7s). Two stimuli were presented on the left and right sides (stimuli positions were randomized on every trial) on the screen, subjects had to choose an option by touching one of two infra-red sensors placed in front of their left and right hands that corresponded to the stimuli on the screen (decision phase, mean RT=1089ms after excluding trials with RT>10s). If the correct option was chosen, the unchosen option disappeared and the chosen option remained on the screen and a juice reward was delivered. If the incorrect option was chosen, both stimuli disappeared and no juice was delivered (outcome phase; 1.5s). Each reward was composed of two 0.6mL drops of blackcurrant juice delivered by a spout placed near the subject's mouth during testing. A given session ended when subjects performed 150 rewarded trials (on average 183.4 trials in total). The task used a deterministic reversal schedule such that each session began with one stimulus that always led to a reward and another stimulus that always led to no reward. The stimulus-reward contingencies reversed for the first time after 50 rewarded trials had been performed and then again after a further 50 rewarded trials were performed. No cue signalled the change in stimulus-reward assignment (Figure 26b). Each animal performed four to six sessions in the MRI scanner.

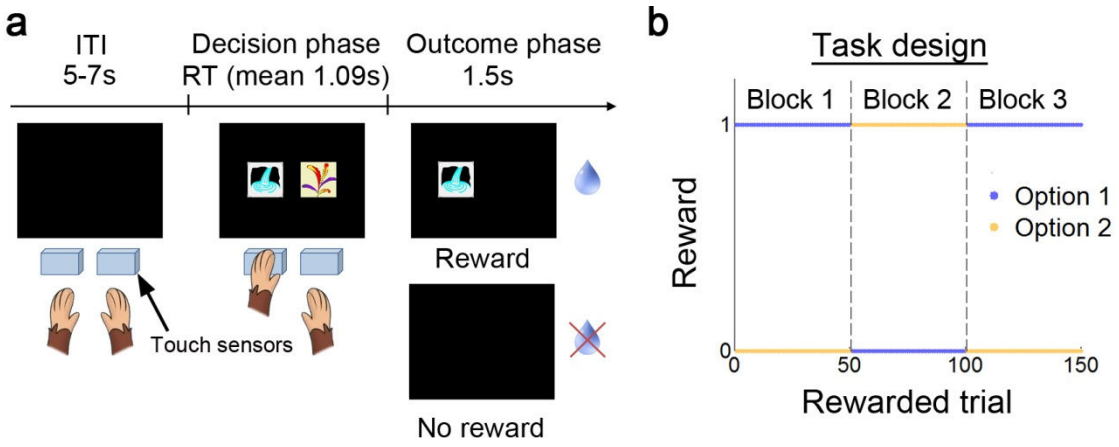


Figure 26. Object discrimination reversal (ODR) task. Each trial started with an inter-trial interval (ITI) showing a blank screen. Two options were then presented on the screen, monkeys chose one of the options by reaching the touch sensor placed in front of it (decision phase). If monkeys chose a correct option, the chosen option remained on the screen, whereas the unchosen option disappeared. Two drops of juice reward were delivered at the same time. However, if an incorrect option was chosen, both options disappeared and no juice was delivered (outcome phase). ***(b)*** *The task was designed with a two-option deterministic reversal schedule. Each session began with one correct option that led to a reward and one incorrect option that did not lead to a reward. After monkeys performed 50 rewarded trials, a new block started and the stimulus-reward contingencies reversed. Another new block/reversal started once more after another 50 rewarded trials were performed. There was no cue indicating a reversal. A testing session ended after monkeys made 150 correct choices.*

Behavioural analysis

Mean accuracies were calculated across subjects within the whole group (**Figure 27**a, c, d), across sessions within subject (**Figure 27**b) and across trials within session (**Figure 27**e, f). In **Figure 27**a, c, d, trials with the same trial number in a block were averaged across every testing session from all subjects. In **Figure 27**b, trials with the same block trial number were averaged only across testing sessions from the same subject. In the within session level (**Figure 27**e, f and subsequent fMRI analyses), accuracy of each trial was calculated by an

averaging window that included the correctness of the current trial as well as that of the next four trials.

In Figures 30a, b and 32a, b we show lOFC and ACC signals are related to learning phase accuracies and post-learning phase accuracies which serve as indices for learning rate and choice stochasticity respectively. In Figures 30c, d and 32c, d, we provided additional analyses confirming the different relationships between lOFC/ACC signals and behaviour by using learning rates and temperature parameters estimated by a computational model. The learning rate parameter α was estimated by the Rescorla-Wagner model (Rescorla and Wagner, 1972):

$$V_{t+1,s} = \begin{cases} V_{t,s} + \alpha\left(r_t - V_{t,s}\right) , & \text{if option } s \text{ was chosen} \\ V_{t,s} , & \text{if option } s \text{ was unchosen} \end{cases}$$

where $V_{t,s}$ and r_t are the value of option *s* and the choice outcome on trial *t* respectively. α is a learning rate free parameter.

The stochasticity parameter was estimated by applying a softmax function that models probabilities of choosing each option:

$$P_{t,s} = \frac{\exp(V_{t,s}/T)}{\sum_{s'=1}^{2} \exp(V_{t,s'}/T)}$$

where $P_{t,s}$ is the probability of choosing option *s* on trial *t*. *T* is a temperature free parameter describing the degree of stochasticity.

The free parameters *α* and *T* from the Rescorla-Wagner model and the softmax function respectively were fitted session-by-session by minimizing the negative log likelihood *L*:

$$L = -\sum_{t=1}^{N} \log(P_{t,c_t})$$

where *N* is the total number of trials and c_t is monkey's choice on trial *t*.

Imaging data acquisition

Imaging data were collected using a 3T MRI scanner and a four-channel phased-array receive coil in conjunction with a radial transmission coil (Windmiller Kolster Scientific, Fresno, CA). FMRI images and reference images for artefact corrections were collected while awake animals were head-fixed in a sphinx position in an MRI compatible chair. FMRI data were acquired using a gradient-echo T2* echo planar imaging (EPI) sequence with $1.5\times1.5\times1.5mm^3$ resolution, TR=2.28s, TE=30ms, flip angle=90°. Proton-density-weighted images using a gradient refocused echo (GRE) sequence (TR=10ms, TE=2.52ms, flip angle=25°) were acquired as reference for body motion artefact correction. T1-weighted MP-RAGE images ($0.5\times0.5\times0.5mm^3$ resolution, TR=2500ms, TE=4.01ms) were acquired in separate anesthetized scanning sessions (for details, see Sallet et al., 2013).

FMRI data preprocessing

FMRI data were corrected for body motion artefact by an offline-SENSE reconstruction method (Kolster et al., 2009) (Offline_SENSE GUI, Windmiller Kolster Scientific, Fresno, CA). The images were aligned to an EPI reference image slice-by-slice to account for body motion and then aligned to subject's structural volume to account for static field distortion(Kolster et al., 2014) (Align_EPI GUI and Align_Anatomy GUI, Windmiller Kolster Scientific, Fresno, CA). The aligned data were processed with high-pass temporal filtering (3-dB cutoff of 100s) and Gaussian spatial smoothing (full-width half maximum of 3mm). The data that were already registered to subject's structural space was registered to the CARET macaque F99 template using affine transformation.

fMRI data analysis

Whole-brain analysis was conducted using a univariate GLM approach with FMRIB's Software Library (FSL; Smith et al., 2004). We searched for brain regions that encoded future win-stay/lose-shift (*WSLS*) strategy: maintenance of the same choice on the next trial ("stay") after a reward outcome on the current trial ("win") and shifting to the alternative choice ("shift") after a no-reward ("lose") outcome on the current trial. To do this we applied a GLM to every testing session that included the following regressors: a binary regressor describing whether the monkey followed win-stay/lose-shift or win-shift/lose-stay (*WSLS regressor*) on the next trial, time-locked to the onset of the outcome phase on the current trial; a binary regressor (*shift*) describing whether the monkey stayed with the same option or shifted to the alternative on the next trial, also time-locked to the onset of the outcome phase on the current trial; a binary regressor (*reward*) indicating whether or not the monkey received any reward on the current trial, time-locked to the onset of the outcome period; and a binary regressor (*side*) describing whether the monkey made a left side or a right side response, time-locked to 0.5s prior to the onset of the outcome. Note that the *WSLS* effect can be considered as the interaction term between the main effects of *reward* and *shift* regressors. The inclusion of the *side* regressor should capture variance and noise in the BOLD signal unrelated to the stimulus-based choices that the monkeys were making but purely related to which hand the animal used.

Analyses were first conducted at the individual subject level. Average effects of the GLM across sessions within the same subject were calculated using a fixed-effects analysis. At the group level, analyses were performed using FMRIB's local analysis of mixed effects stage 1 and 2 (FLAME1+2) (Beckmann et al., 2003; Woolrich et al., 2004) and using standard cluster-based thresholding criteria of $z>2.3$ and $p<0.05$ cluster-corrected (Worsley et al., 1992).

We also conducted analyses on *a priori* defined ROIs by extracting the BOLD time course from two-voxel radius spherical masks placed over the lOFC (16, 8, -4 in CARET macaque F99 coordinates), cOFC (8, 17, 7), amygdala (14, -3, -13) and ACC (-1, 6, 17) signals. The ROIs were placed with respect to group coordinates in CARET macaque F99 space and were then transformed to subject space on each testing session using affine transformation. The extracted time courses were upsampled ten times by cubic spline interpolation. The time courses were then time-locked to the onset of the outcome phase and were analyzed by applying a GLM to every time point of each scanning session separately to obtain a time course of β weights. β weights of regressors across sessions of the same subject were averaged. The mean and standard error (denoted in the figures by lines and shadings respectively) of all the within-subject β weights were calculated across subjects for plotting the effect size time courses. In Figures 28c, 29b, 31b and 34b, the GLM included regressors describing *WSLS* (where the shift/stay response occurred on the next trial), *shift* and *reward* as a control regressor (not shown in the figure). In Figures 35 and 36, the GLM included only the *WSLS* regressor. In **Figure 38**, the GLM included a binary regressor describing whether the outcome on the previous trial had been a reward or not (*previous reward*) and also *WSLS* as a control regressor (not shown in the figure). In **Figure 33**, the ACC signals were time-locked at the decision phase and the GLM included the *shift*, *during shift* and *after shift* regressors that were binary regressors describing whether a choice shift was performed on the previous, current, or next trials respectively.

Functional connectivity analyses were performed between the lOFC and amygdala and the results of these analyses are shown in Figures 37, 39 and 40. In these analyses, the BOLD time course of the lOFC was used as the physiological regressor to predict the amygdala BOLD signal. In **Figure 37**, the psychological regressors were *shift*, *reward* and a regressor contrasting between the lose-shift and win-stay components of the WSLS strategy. The psychophysical interaction (PPI) regressor was computed by taking the product between

the lOFC time course and the lose-shift versus win-stay contrast. In **Figure 39**, the psychological regressors were *WSLS*, *reward* and *previous reward*. The PPI term was the product of the lOFC time course and the *previous reward* contrast. In ***Figure 40***, the psychological regressors were *WSLS*, *reward*, *previous reward* and the lose-shift versus win-stay contrast regressor. The PPI terms were the products of the lOFC time course with *previous reward* and the lose-shift versus win-stay contrast.

We also examined the relationship between the session-to-session variation in BOLD signal sizes and session-to-session variation in behaviour. The sizes of signal for each individual testing session that were used in the analysis were the peak signals taken within a specific time window. The *WSLS* signal sizes for the analyses shown in Figures 29 and 30 were the peak signals recorded between 2 and 7s. The *shift* signals in Figures 32 and 33 was the peak signal recorded between -3 and 1s. The *shift*, *during shift* and *after shift* signals in **Figure 33** were the peak signal recorded between -1 and 0s (with respect to the onset of the decision phase). The peak PPI size in the analysis in **Figure 37** was extracted between 0 and 5s. The PPI effect size in **Figure 39** was extracted between -1 and 2s. In order to remove between subject variance and to spare between session variance within the same subject each extracted neural signal size was normalized by subtracting the mean signal size across sessions within the same subject and divided by the standard deviation of the signal sizes for that subject. A similar normalization procedure was applied to behavioural indices: each session's behavioural index was normalized by subtracting the mean behavioural index across sessions for a given subject and then dividing by the standard deviation of the behavioural indices for that subject. T-tests were all two-tailed.

5.3 Results

Animals' behaviour

Four monkeys participated in our deterministic ODR task (Figure 26a), they had to discriminate which of two options led to a reward at the beginning of a session. The reward assignment reversed after the monkeys performed the 50th rewarded trial and then it reversed again after the 100th rewarded trial. The daily session stopped after the monkey performed 150 rewarded trials in total (Figure 26b). The animals encountered new stimuli at the beginning of each day of testing. On average, the monkeys performed 183.4 trials in each session and each monkey contributed 4-6 sessions in this dataset. To investigate the animals' behaviour, we split each session into three blocks (block 1: the initial learning period before any reversals; blocks 2 and 3: after the first and second reversal respectively) and calculated the average accuracy of each trial as a function of its position in the block (trial 1, trial 2, trial 3, and so on; **Figure 27**a). The monkeys typically showed poor accuracy on early trials but they were consistently above 50% correct after the 9th trial of a block (t_3>3.920, p<0.030; **Figure 27**a). The average performance of each monkey is shown in **Figure 27**b. When block 1 and blocks 2+3 were analyzed separately, the first trials of block 1 had a higher accuracy (38%; **Figure 27**c) due to random decisions than that of blocks 2+3 (0%) due to post-reversal decisions (**Figure 27**d). To illustrate performance on individual testing sessions the data were smoothed by calculating a running average over 5 trials that stopped at the last four trials of a block to avoid inclusion of trials in different blocks within a given average. **Figure 27**e, f show monkeys' performances in two example sessions.

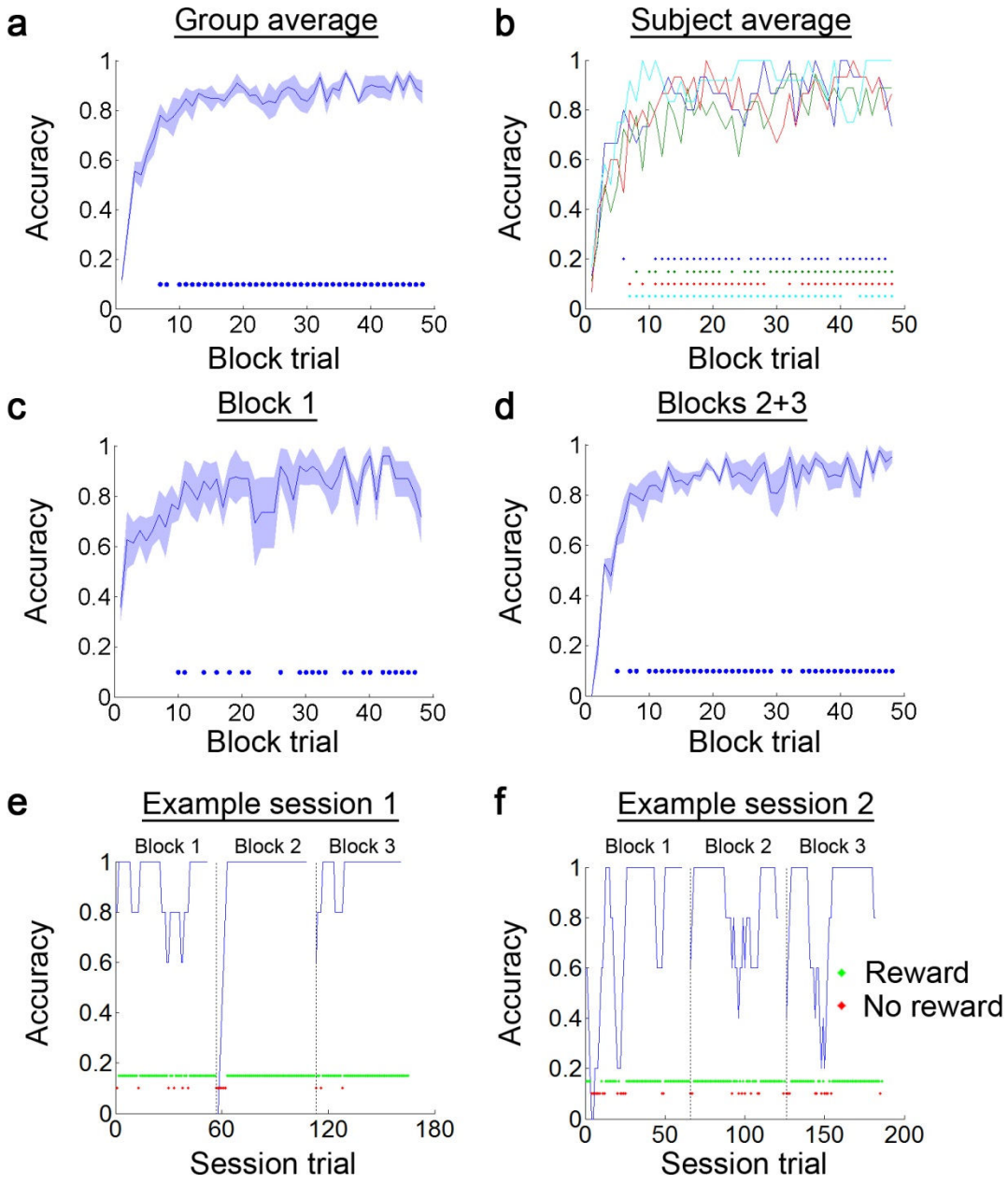


Figure 27. *Subjects' behaviour.* ***(a)*** *On average, all monkeys showed low accuracies on the early trials in a block.* ***(b)*** *Performance averaged across testing sessions within subject. Each line represents data from one subject. Monkeys showed similar performances in* ***(c)*** *block 1, which was the initial discrimination block, and* ***(d)*** *blocks 2+3, which were the blocks after the first and second reversals respectively. The raster plots in (a-d) indicate trials in a block with accuracies significantly higher than 0.5 ($p<0.05$).* ***(e, f)*** *Example sessions from two different subjects. Accuracies were calculated by using a moving average window of 5 trials. The dotted lines indicate reversals. The raster plots indicate the reward (green) and no reward (red) outcome events.*

lOFC encoded a win-stay/lose-shift signal

In this task, an optimal strategy is to stay with the same choice on the next trial after rewarded decisions but to shift to the alternative choice after non-rewarded decisions. In other words, monkeys should follow a win-stay/lose-shift (WSLS) rule. Our first analysis, therefore, examined activity across the whole brain to identify regions encoding the WSLS rule. It has recently been suggested that cOFC is not critical for reward-guided behavioural change (Rudebeck et al., 2013b) and by carrying out this analysis we hoped to identify which adjacent frontal area, if any, might possess activity related to the deployment of the WSLS rule needed for reward-guided behavioural change.

The effect of the WSLS rule on neural activity can actually be thought of as the effect of an interaction term on brain activity where the main effects are of *reward delivery* (versus non-delivery) and *choice shifting* (the next selected option is different from the current trial, versus a stay choice, where the next selected option is the same as on the current trial). We therefore also included these two main effects in our regression model. Whether reward was delivered or not had a broad impact on activity throughout the brain as has been previously reported (Vickery et al., 2011) and so, for the sake of clarity, we focus on WSLS and the main effect of choice shifting in the following fMRI analyses because these effects varied across regions and because they are the signals most directly related to the guidance of future behaviour.

First we conducted a whole brain analysis using standard fMRI blood oxygen level dependent (BOLD) imaging analysis procedures (see Methods). We identified a region in lOFC that carried a signal guiding WSLS behaviour in the next trial (cluster-based thresholding $z>2.3$, $p<0.05$ cluster-corrected; **Figure 28**a). The activation lay just lateral to

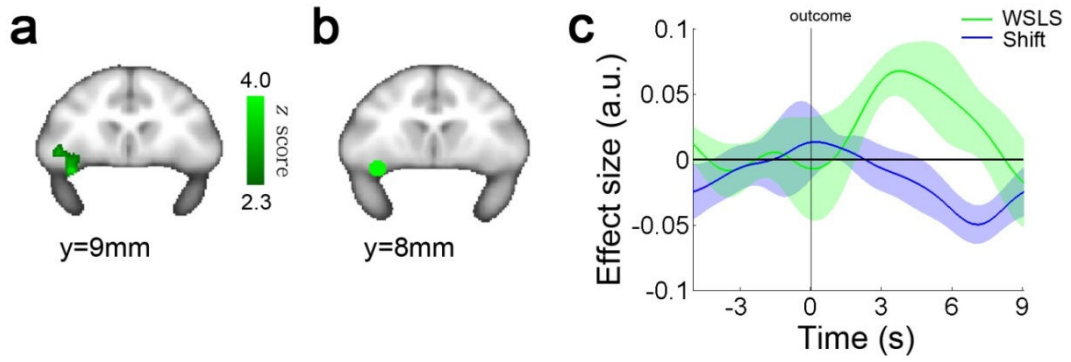


Figure 28. *Win-stay/lose-shift signal in the lOFC.* ***(a)*** *A whole brain analysis showing a signal in the lOFC lies within the orbital part of area 12/47 that predicted win-stay/lose-shift behaviour.* ***(b)*** *lOFC (16, 8, -4; green) BOLD activity was extracted for ROI analysis.* ***(c)*** *The lOFC showed WSLS activity ramped up after the onset of the outcome phase and peaked at around 4s (green). There was no significant shift signal with a similar peak latency during the outcome phase (blue).*

the lateral orbital sulcus and therefore just lateral to the region affected by the neurotoxic lesions made by Rudebeck and colleagues (2013b). It is therefore likely that the connections of this region would have been affected by the aspiration lesions that Rudebeck and colleagues showed did compromise reversal task performance.

To better understand and visualize activity in lOFC we extracted the time course of the regression coefficients relating BOLD activity to WSLS and the main effect of choice shifting from a region of interest (ROI) placed over the peak of the lOFC effect (**Figure 28**b). We avoided "double dipping" (Kriegeskorte et al., 2009) as we did not conduct further statistical analysis at this stage. In lOFC the WSLS signal ramped up after the onset of the outcome phase and peaked with a delay of around 4s (green trace, **Figure 28**c). The latency

is similar to that seen in other macaque BOLD imaging studies (Leite et al., 2002). No effect of behavioural shifting *per se* was visible with the same latency (blue trace, **Figure 28**c).

The whole brain fMRI test that we conducted follows the most widely used conventions in neuroimaging experiments. These conventions, however, are designed to deal with the difficulty of performing mass univariate statistical comparisons across multiple MRI voxels and are known to be conservative. It is well known, however, that such an approach can fail to identify meaningful and replicable effects not least if their spatial extent is limited. We therefore used an additional ROI approach to examine the relationship between BOLD and WSLS and the main effect of behavioural shifting in the central OFC (cOFC) region between the medial and lateral orbital sulci where neurotoxic lesions were made in previous studies (**Figure 29**a). We were able to identify a WSLS effect that approached a marginal level of significance (**Figure 29**b).

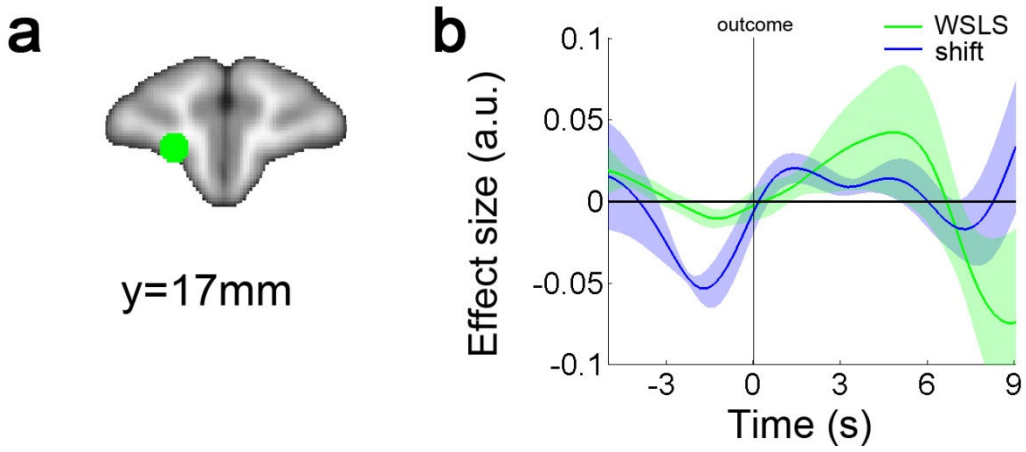


Figure 29. *WSLS signal in the cOFC.* ***(a)*** *The cOFC (8, 17, 7; green) BOLD activity was extracted for ROI analysis.* ***(b)*** *There was a marginal level of WSLS during the decision phase (green) and a negative shift signal prior and during the decision phase (blue).*

In summary, while there may be neurons in the cOFC region that are important for WSLS behaviour the lOFC may be an especially important when reward outcomes are used to guide adaptive behaviour. To investigate whether this was actually the case we next tested how the lOFC WSLS signal was related to task performance (note this test is statistically orthogonal to the original test used to identify the WSLS signal). We obtained the peak size of the WSLS signal from each animal on each testing session and the average choice accuracy from each animal on each testing session. Since the testing sessions were contributed by four different monkeys (4-6 sessions per monkey), we normalized the peak lOFC signal sizes and the behavioural accuracy measurements within each individual animal to remove between subject variance. The normalization of measurements (both of neural signals and of behavioural indices) was performed by subtracting the mean of the measurement within the same subject and then dividing by the standard deviation of the same measurements from the same subject, such that between subject variance was removed while between session variance remained.

In the behavioural task, learning the option values was important during the early trials of a block (learning phase: in the first 9 trials of each block before the group accuracy (**Figure 27**a) was consistently larger than 50%), whereas maintaining high choice consistency or low stochasticity was crucial for the monkeys after the learning phase (post-learning phase: after 30 trials of each block when group accuracy was consistently larger than 70%). Similar definitions of learning stages were employed to investigate performances in reversal learning (Jones and Mishkin, 1972; Murray and Izquierdo, 2007). We found that testing sessions with stronger lOFC WSLS signals were associated with higher accuracies in the learning phase (r=0.500, p=0.029; **Figure 30**a). The lOFC WSLS signal was not, however, related to accuracy in the post-learning phase (r=0.233, p=0.336; **Figure 30**b). These conclusions were confirmed when a complementary reinforcement learning model was applied (see *Behavioural analysis* in *Methods*); the lOFC WSLS signal was positively related to

the learning rate parameter (r=0.570, p=0.011) and unrelated to the stochasticity parameter (r=-0.071, p=0.772; **Figure 30**c, d).

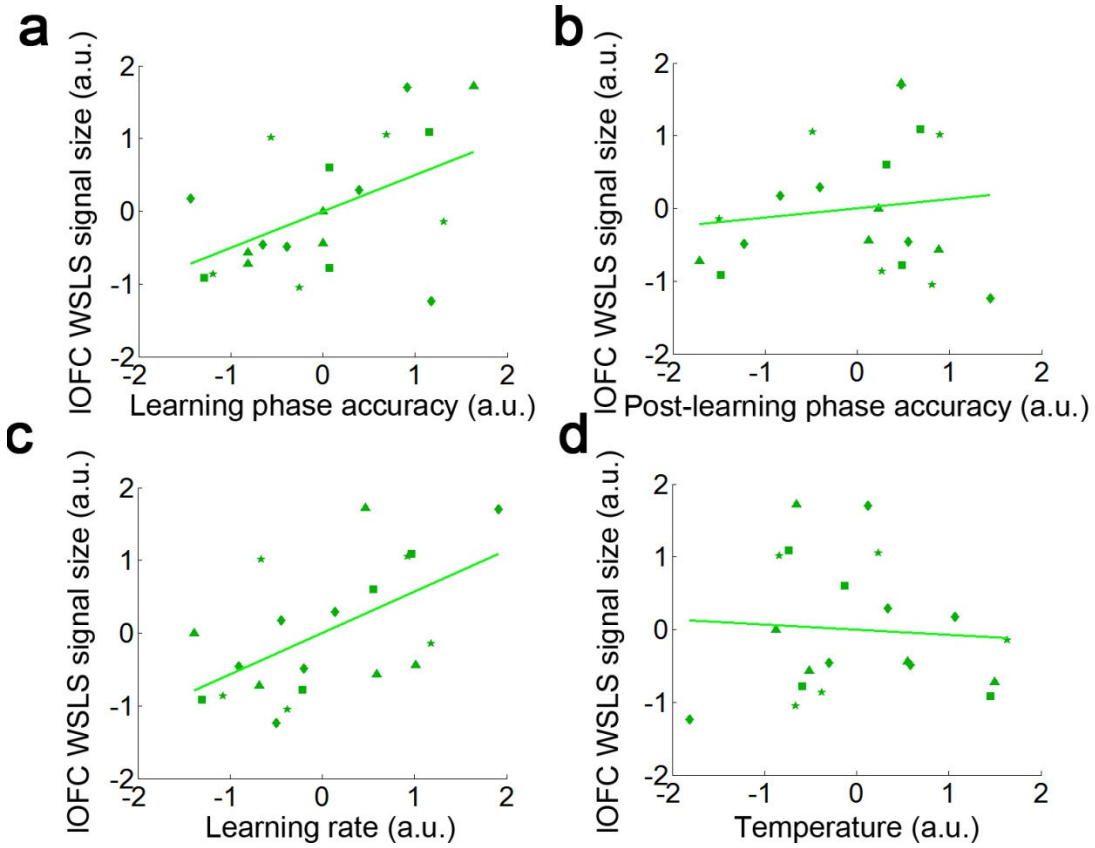


Figure 30. ***(a)*** *Testing sessions with larger WSLS signals in the lOFC were related to higher accuracies during the learning phase (first 9 trials in a block).* ***(b)*** *The sizes of the WSLS signal had no relationship with accuracies in the post-learning phase (after 30 trials in a block).* ***(c)*** *Testing sessions with larger WSLS signals in the lOFC were related to higher learning rates.* ***(d)*** *The sizes of the WSLS signal had no relationship with the temperatures, which are indices of choice stochasticity.*

Each type of marker symbol represents data from testing sessions of one animal.

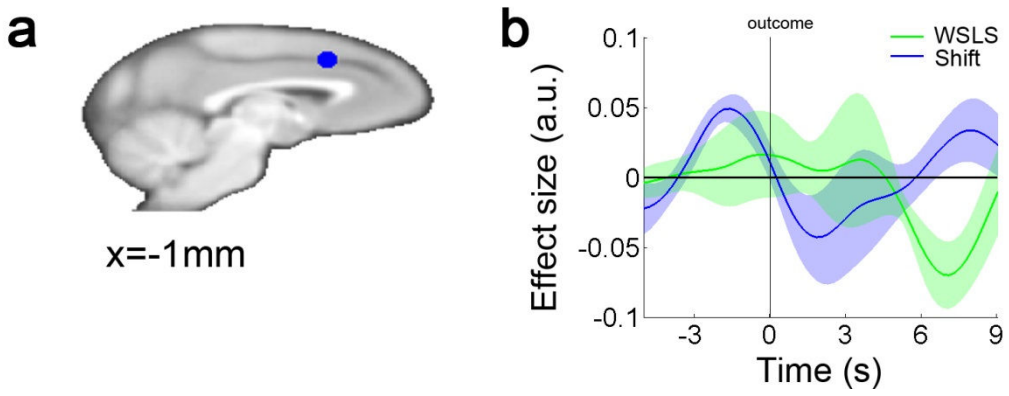


Figure 31. *Shift signal in the ACC.* ***(a)*** *The ACC (-1, 6, 17; blue) BOLD activity was extracted for ROI analysis.* ***(b)*** *There was a shift signal that ramped up prior and during the decision phase (blue). There was no WSLS signal during the decision or outcome phases (green).*

Comparing the roles of lOFC and ACC in guiding future behaviour

We next used similar ROI analysis procedures to examine signals in other brain regions also implicated in stimulus-reward reversal: the ACC (**Figure 31**a) (Chudasama et al., 2013) and amygdala (**Figure 34**a) (Schoenbaum et al., 2003; Saddoris et al., 2005; Schoenbaum et al., 2007; Morrison et al., 2011).

The ACC ROI was placed near the rostral cingulate motor zone that was included in lesions in which motor response reversal deficits have been reported (Rudebeck et al., 2008) and in a study in which ODR impairment was recently reported (Kennerley et al., 2006; Chudasama et al., 2013). Neubert and colleagues (submitted) have shown that, in the macaque, this region's activity coupling patterns resemble those of the human ACC region most closely linked to reward-guided decision-making and learning (Walton et al., 2004; Behrens et al., 2007; Rushworth et al., 2011; Amiez et al., 2012). If ACC has the same role as OFC in reward-guided behavioural change (Chudasama et al., 2013) then it should have the same WSLS signal as OFC. If instead ACC is solely concerned with some very different role, for example with response-reward association rather than stimulus-reward association as

has been argued (Rushworth et al., 2007; Rudebeck et al., 2008), then it might not be expected to carry any task-related signals.

The pattern of activity recorded in ACC accorded with neither expectation. Unlike OFC, ACC did not encode future WSLS behaviour during the outcome phase of the current trial, but it did carry a different task-related signal; there was a significant main effect of behavioural shifting. ACC activity was higher prior to and during the decision phase of a trial when a monkey shifted its choice from one stimulus to the other on the next trial regardless of whether reward was received (**Figure 31**b). In our stimulus-reward reversal task, choice shifting was only required right after the two reversals in stimulus-reward association and sometimes on the second trial of a session if the monkey did not choose the rewarded option on the first trial. For most of the time, the task required the monkeys to choose the same option repeatedly once the rewarded option was identified. Nevertheless the monkeys, as observed in many experiments, exhibited exploratory behaviour throughout the task even in post-learning phases. A second analysis confirmed a close link between the impending shift signal in ACC and exploratory behaviour. As we had done for the lOFC WSLS signal we examined the relationship between variation in ACC shift signal size and behavioural change. We applied an analogous normalization procedure to ACC signals and to measurements of average accuracy in the post-learning phase of each session. Stronger ACC shift signals were related to lower accuracies during the post-learning phase (**Figure 32**b; r=-0.480, p=0.038) but not in the learning phase (**Figure 32**a; r=0.121, p=0.621). This conclusion was confirmed by an independent but complementary test that showed that the ACC shift signal was related to computational parameters in a similar way – stronger signals were associated with higher stochasticities (r=0.480, p=0.038) but were unrelated to learning rates (r=0.168, p=0.492; **Figure 32**c, d). Note that these relationships are opposite to those seen in lOFC (**Figure 30**). Previous studies have shown that ACC activity varied as a function of whether an exploratory behaviour was performed (Hayden et al., 2011). We

found that ACC had a strong signal before a shift (t_{19}=5.703, p<0.001) and during a shift (t_{19}=3.378, p=0.003) but did not encode the shift on the trial after it had been effected (t_{19}=0.723, p=0.478; **Figure 33**). An ANOVA also showed that there was a significant difference in the size of these three signals ($F_{2,19}$=4.14, p=0.024).

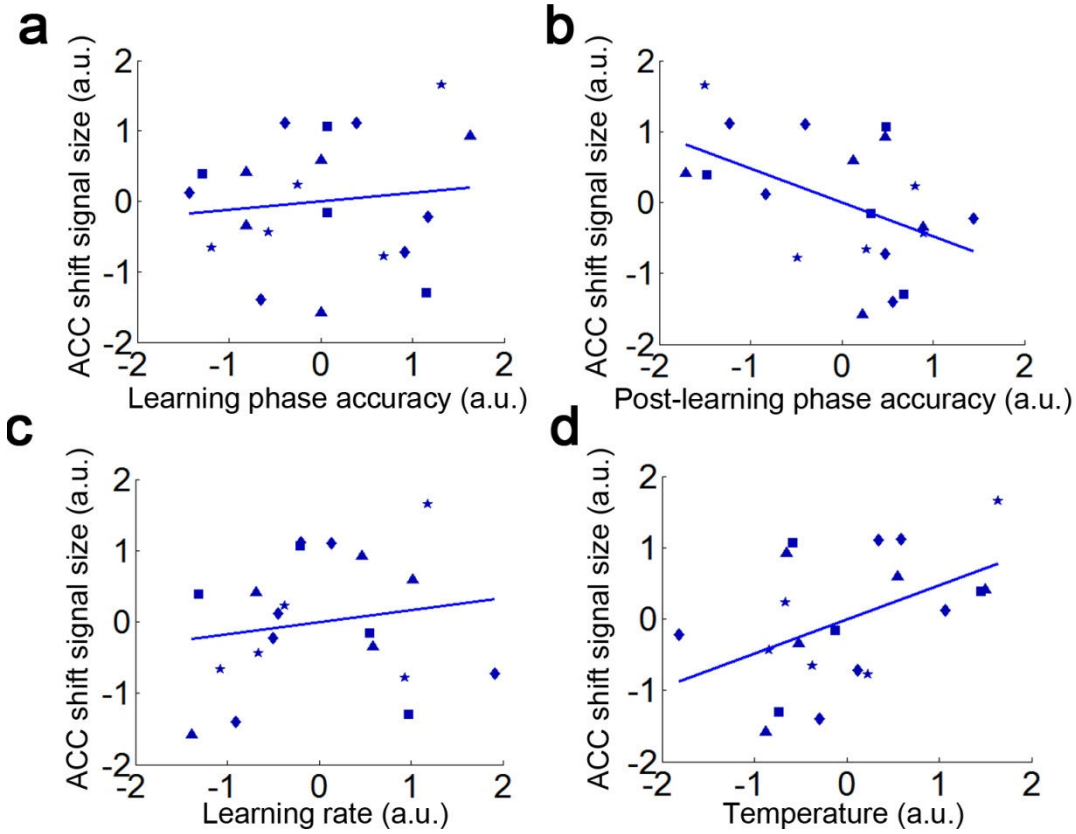


Figure 32. ***(a)*** *The sizes of the shift signals in the ACC had no relationship with accuracies in the learning phase.* ***(b)*** *Larger shift signals were related to higher accuracies during the post-learning phase. Each type of marker symbol in (c, d) represents data from one animal.* ***(c)*** *The sizes of the shift signals in the ACC had no relationship with the learning rates.* ***(d)*** *Larger shift signals were related to higher temperatures, which are indices of choice stochasticity. Each type of marker symbol in (a, b) represents data from one animal.*

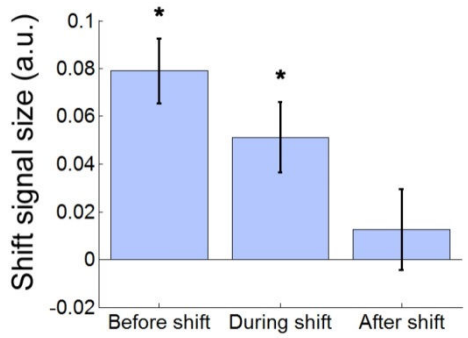


Figure 33. *The ACC shift signal remained high before (left) and during (middle) the execution of a choice shift and dropped after the choice shift was performed (right).*
** p<0.01*

Comparing the roles of lOFC and amygdala in guiding future behaviour

There is uncertainty about the role of primate amygdala in stimulus-reward reversal learning. Aspiration and radiofrequency lesions impair stimulus-reward reversal learning (Jones and Mishkin, 1972; Aggleton and Passingham, 1981; Spiegler and Mishkin, 1981) but such lesions also compromise fibres of passage through the amygdala and subjacent cortex. Confusingly neurotoxic lesions of the amygdala in rats have been reported as impairing (Schoenbaum et al., 2003) and improving (Izquierdo et al., 2013) stimulus-reward association learning while in old world monkeys they have been reported as leaving ODR performance unaffected (Izquierdo and Murray, 2007) or even improved (Rudebeck and Murray, 2008).

Some amygdala lesion-induced changes in OFC activity and other neurophysiological recordings made in the OFC and amygdala suggest the regions exchange information in order to bring about reward-guided behavioural change (Saddoris et al., 2005; Morrison et al., 2011; Rudebeck et al., 2013a). By contrast, the fact that OFC lesion-induced impairments in reward-guided behavioural shifting are reversed by subsequent amygdala lesions

(Stalnaker et al., 2007) and the suggestion that neurotoxic amygdala lesions might even improve ODR performance suggests opposition between the roles of the areas.

An ROI was placed over basolateral and lateral amygdala, which was the locus of previous reversal learning and associative learning experiments (**Figure 34**a) (Maren, 2005; Saddoris et al., 2005; Stalnaker et al., 2007; Johansen et al., 2010). We found that it carried a future WSLS signal at the outcome phase that was similar to one seen in lOFC. However, it also carried a shift signal (a main effect of behavioural change) that resembled the ACC shift signal. The amygdala shift signal, however, was most prominent at the outcome phase of the trial whereas the ACC shift signal was prominent at and prior to the decision phase (**Figure 34**b).

Because the amygdala's role in reward-guided behavioural reversal may be limited to certain stages or events during learning (Rudebeck and Murray, 2008) we analyzed the WSLS signals on trials when mean accuracy was lower than 0.7 (mean accuracy was calculated by using a moving window of five trials) and also during the five trials after reaching the 0.7 accuracy criterion. Presumably monkeys were still learning which stimulus

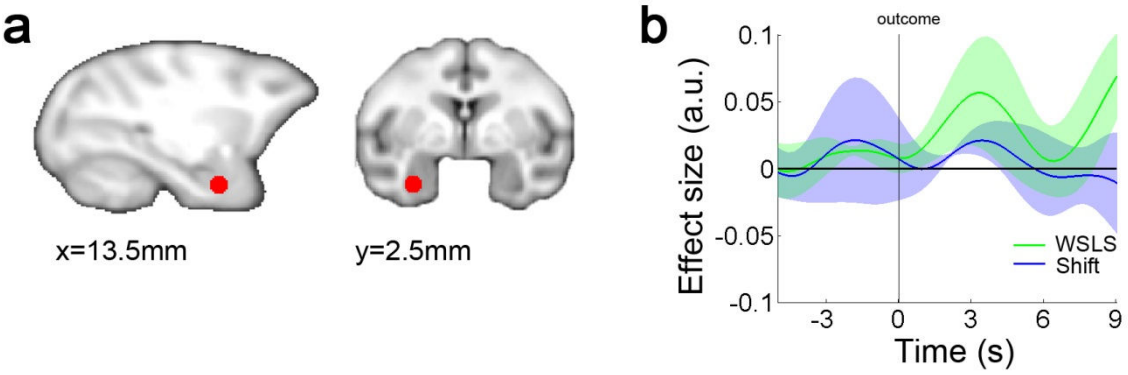


Figure 34. ***(a)*** *The amygdala ROI (14, -3, -13; red) in sagittal view (left panel) and coronal view for BOLD activity extraction.* ***(b)*** *The amygdala encoded both a WSLS signal (green) and a general shift signal (shift) during at the outcome phase.*

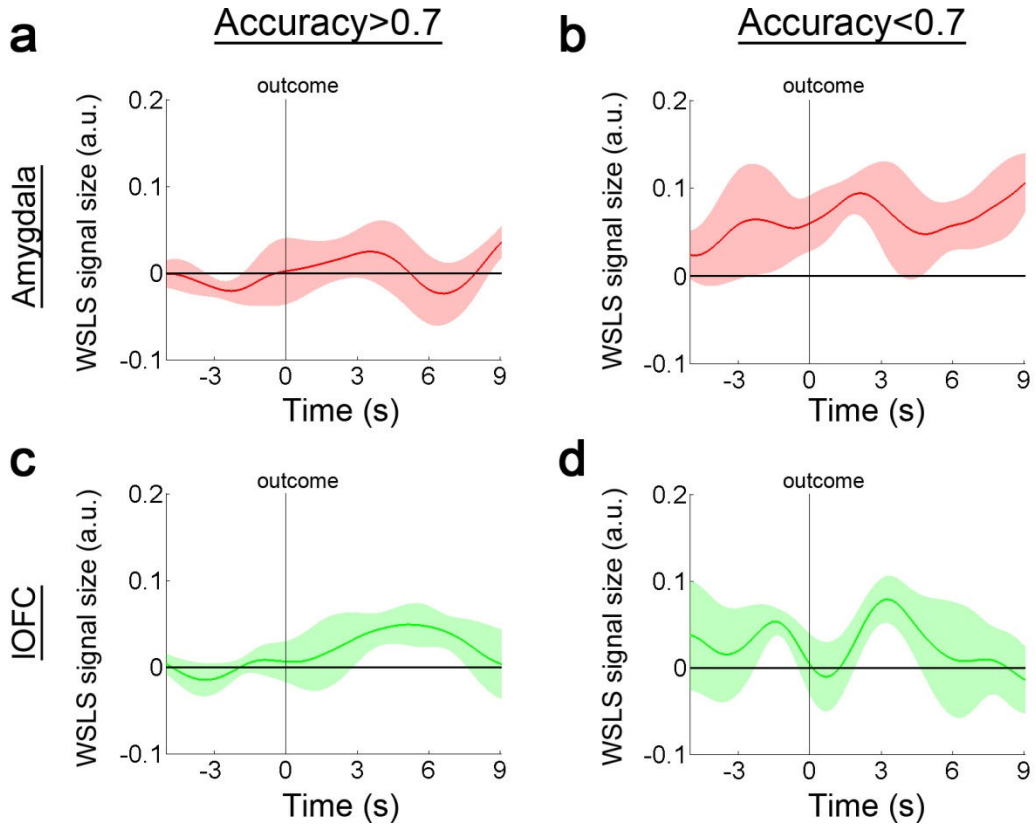


Figure 35. *The WSLS signal in amygdala was specific to low accuracy trials. We split the WSLS signals by whether a trial had accuracy below 0.7. The amygdala did not encode a WSLS signal when monkeys' accuracies were high* ***(a)*** *but did so when accuracies were low* ***(b)****. In contrast, the lOFC encoded WSLS signals no matter whether accuracies were high or low* ***(c, d)****.*

was correct on these trials. We compared the amygdala WSLS signal on these trials with the WSLS signal during phases of the task when performance exceeded 0.7 accuracy. Interestingly, we found that the amygdala had a stronger WSLS signal when accuracy was low than when accuracy was high (t_{19}=2.402, p=0.027; **Figure 35**a, b), whereas WSLS signals of similar sizes were present in both the low and high accuracy trials in the lOFC (t_{19}=1.191, p=0.248; **Figure 35**c, d).

Because there were more unrewarded trials in the low accuracy trials when the amygdala exhibited a WSLS signal and because the amygdala encoded both WSLS and shift

signals at the same time, we hypothesized that the amygdala WSLS signal might actually be specific to lose-shift trials. We tested this by analysing win and lose trials separately. We found that the amygdala was active on lose-shift but not on win-stay trials (**Figure 36**a, b). The lose-shift signal was also significantly larger than the win-stay signal (t_{19}=2.532, p=0.020). By contrast the lOFC exhibited both win-stay and lose-shift signals (**Figure 36**c, d; t_{19}=-1.289, p=0.213).

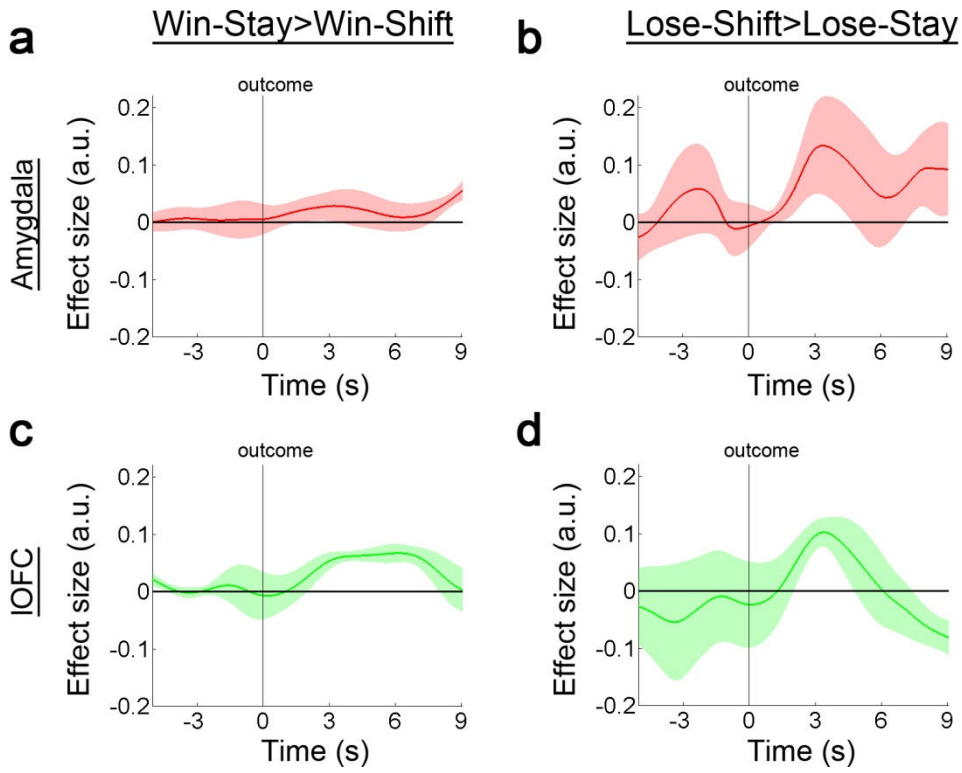


Figure 36. *Lose-shift but not win-stay signal was encoded in the amygdala. We split the WSLS signal by whether it was a win or lose trial. In other words, win-stay and lose-shift signals were investigated separately. The amygdala did not have a win-stay signal* ***(a)****, but encoded a lose-shift signal* ***(b)****. In contrast, both win-stay and lose-shift signals were seen in the lOFC* ***(c, d)****.*

Because the lOFC and amygdala exhibited a similar lose-shift signal but different win-stay signals, it is possible that the connectivity between these two regions could also be modulated as a function of the outcome of a choice. To test this we used a psychophysical interaction (PPI) analysis (Friston et al., 1997). In this PPI analysis we examined the impact on amygdala activity of the interaction of a physiological parameter, the lOFC activity (as indexed by the lOFC BOLD time series) and a psychological parameter indexed by a task variable (a contrast between the lose-shift contingency and the win-stay contingency). We focused on the interaction of these two influences (PPI effect) but took care to include both main effects in our analysis (O'Reilly et al., 2012). The lOFC and amygdala were more strongly coupled during lose-shift events than during win-stay events (**Figure 37**a). There was a relationship between the effect size of this change in coupling in different sessions and behavioural variation in the sessions; larger sizes of lose-shift coupling between lOFC and amygdala were present in testing sessions with higher proportions of lose-shift trials within lose trials (**Figure 37**c; r=0.575, p=0.010), but the degree of coupling was unrelated to the proportion of win-stay trials within win trials (**Figure 37**b; r=-0.075, p=0.760).

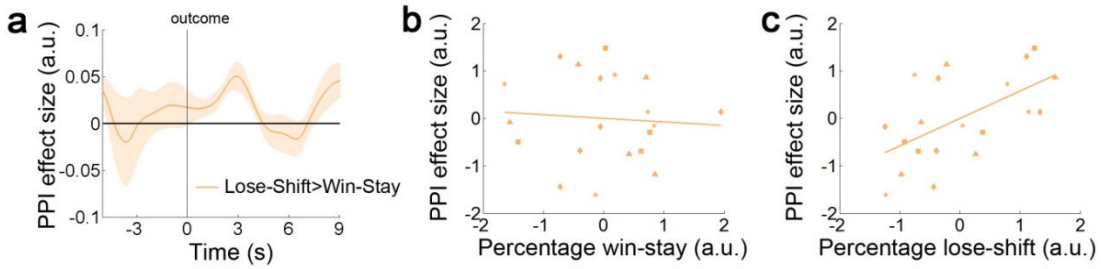


Figure 37. *PPI between lOFC and amygdala that guided behavioural change.* ***(a)*** *lOFC and amygdala showed stronger connectivity for performing a lose-shift behaviour than for performing a win-stay behaviour.* ***(b)*** *The effect sizes of the PPI were unrelated to how often the monkeys made stay choices after win trials.* ***(c)*** *However, larger PPI effect sizes were related to higher chances of lose-shift behaviours. Each type of marker symbol in (b-c) represents data from one animal.*

Non-contingent learning in the amygdala

So far the analyses have been consistent with a view of lOFC and amygdala as interacting in order to bring about behavioural change because the areas share a signal predicting choice shift after a failure to obtain reward. Variation in the size of the amygdala signal and in degree of modulation in lOFC-amygdala coupling predicted lose-shift behaviour. However, the view of the lOFC and amygdala as cooperating in the exchange of information to guide learning is at odds with demonstrations that amygdala lesions improved some aspects of ODR performance in macaques and rescued the ODR impairment caused by OFC lesions in rats (Stalnaker et al., 2007; Rudebeck and Murray, 2008).

One way to reconcile views of the amygdala as either helping or hindering reward-guided learning is by considering the possibility that it makes a particular type of contribution to reward-guided learning. In addition to learning precise contingent relationships between predictive stimuli and rewards it is clear that both in animals and humans a "spread of reward effect" occurs whereby the reward delivered after one choice "spreads" forwards to the next trial so that it also reinforces the choice made on the next trial (Thorndike, 1933; Walton et al., 2010). Such a learning mechanism is unproblematic in many situations if the same choice is repeated trial after trial. It is problematic when a learner is shifting rapidly between choices on consecutive trials because spread of reward effect means that the credit for an outcome on one trial may be partly misassigned to a different choice made on the subsequent trial that may actually have *followed* rather than *preceded* the outcome. If the amygdala still encodes information about rewards on trial t-1 that had actually been received in response to a choice, potentially a different choice, on trial t then it should be possible to observe an impact of the previous trial's reward on the amygdala BOLD signal.

Whether or not reward had been received on a previous trial did not influence decision and outcome related activity on subsequent trials in lOFC (**Figure 38**a). In the ACC, although there was a previous reward signal before the decision phase and after the end of the current trial, the signal was absent during and briefly after the outcome phase (**Figure 38**b). In contrast, the amygdala carried a previous reward signal throughout the course of the subsequent trial's decision and outcome phases (**Figure 38**c), suggesting that the amygdala might mediate the assignment of the reward from a previous trial to an option chosen on the current trial and hence mediate a spread of reward effect. In a PPI analysis in which we examined the interacting influences of lOFC activity and the previous trial's reward on amygdala activity, we found negative lOFC-amygdala coupling as a function of previous reward delivery (**Figure 39**a). In other words, the lOFC-amygdala coupling was weaker when the amygdala itself was showing a signal related to reward delivery on the previous trial.

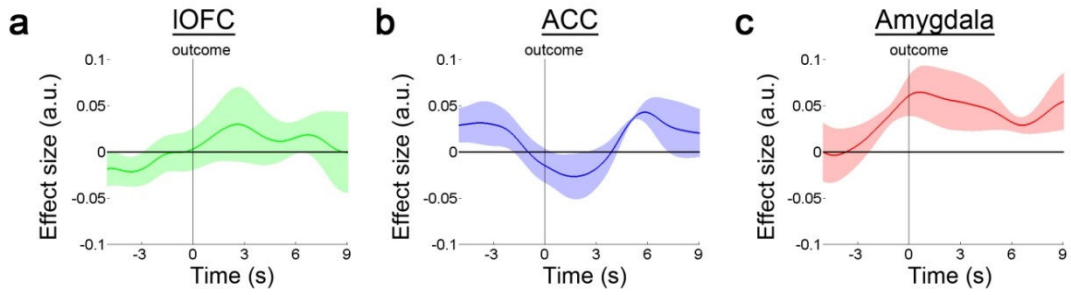


Figure 38. *Previous reward signals in the lOFC, ACC and amygdala.* ***(a, b)*** *The lOFC and ACC did not carry signals related to whether or not a reward had been delivered on the previous trial at the time of the outcome phase on the subsequent trial.* ***(c)*** *The amygdala encoded the outcome of the previous trial at the time of the decision and outcome phases of the current trial.*

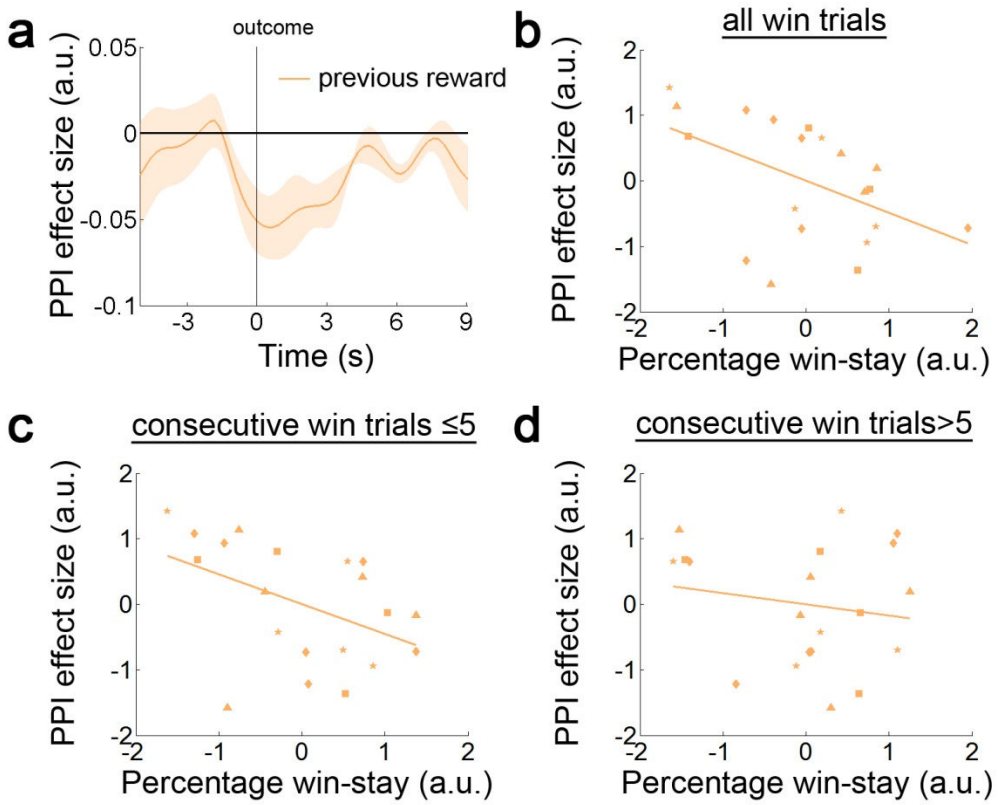


Figure 39. ***(a)*** *The lOFC-amygdala coupling was negatively modulated as a function of the previous outcome. Testing sessions with stronger negative modulation was related more stay decisions after a win trial* ***(b)****.* ***(c)*** *The relationship between the PPI effect size and percentage of win-stay was specific to the first 5 consecutive win trials.* ***(d)*** *The size of modulation in lOFC-amygdala connectivity as a function of previous outcome delivery was no longer related to how often monkeys repeated the same choice after the first 5 consecutive win trials. Each type of marker symbol in (b-d) represents data from one animal.*

One important consequence of non-contingent learning is that when consecutive choices alternate and only some are rewarded, it is more unlikely that an animal will stay with the rewarded option because credit for the reward may be misassigned to a subsequent incorrect choice. The "credit" for a non-reward may also be misassigned to a subsequent correct choice. Negative lOFC-amygdala coupling appears to reduce such credit assignment problems; in win trials, we found that stronger negative coupling between the lOFC and amygdala, as a function of previous reward, was related to more frequent future

win-stay decisions (**Figure 39**b; r=-0.491, p=0.033). When a series of win trials were performed consecutively, any effect of non-contingent learning should be most obvious on the first few consecutive win trials that occurred early in the series but it should diminish after repeatedly choosing the same option many times because even if there was non-contingent learning the credit for the previous reward could no longer be spread to another choice. We found that increased lOFC-amygdala negative coupling linked to a previous reward was related to a higher proportion of win-stay choices in the first five consecutive win trials (**Figure 39**c; r=-0.455, p=0.050; the correlation was much stronger after removing data from one outlier session, r=-0.616, p=0.005) but lOFC-amygdala negative coupling was not related to the proportion of win-stay choices made after the same win choice had already been repeated more than five times (**Figure 39**d; r=-0.168, p=0.491).

Finally, in order to investigate whether the lOFC-amygdala coupling could be simultaneously modulated as a function of a previous reward in a negative manner and as a function of lose-shift behaviour in a positive manner, we included both PPI regressors at the same time in one analysis. The results for the two PPI effects remained the same (***Figure 40***), suggesting both types of modulation co-existed.

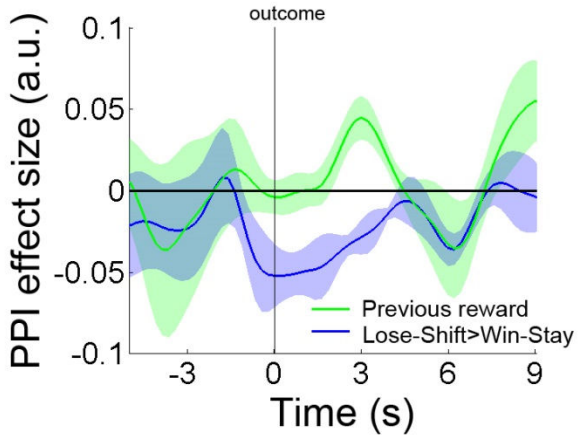


Figure 40. *Two types of activity coupling between lOFC and amygdala. When the models in Figures 37a and 39a were combined, similar patterns of PPI can be observed. There was a stronger lOFC-amygdala coupling when performing lose-shift behaviour than win-stay behaviour (green). At the same time, there was a negative effect on the lOFC-amygdala coupling as a function of whether a reward was delivered in the previous trial (blue).*

5.4 Discussion

The amygdala and OFC have been linked to the flexible use of reward information to guide behaviour. However, the nature of the link has recently become controversial. For many years flexible reward-guided behaviour has been investigated using the ODR task in which animals have to learn first that one stimulus is associated with reward but that subsequently only an alternative stimulus is associated with reward. The identification of amygdala and OFC with ODR reflects a long history of demonstrations that lesions impair ODR in macaques (Butter, 1969; Jones and Mishkin, 1972; Izquierdo et al., 2004; Rudebeck and Murray, 2008) and more recent demonstrations that the responses of individual neurons in macaque OFC and amygdala track changes in the reward associations of stimuli (Paton et al., 2006; Morrison et al., 2011).

Unfortunately this attractively coherent picture of OFC and amygdala function has been called into question by recent demonstrations that OFC and amygdala lesions in

macaques do not disrupt ODR when the lesions are made by neurotoxin injection (Izquierdo and Murray, 2007; Kazama and Bachevalier, 2009; Rudebeck et al., 2013b). Confusingly neurotoxic lesions made in the OFC of new world monkeys, such as marmosets (Dias et al., 1996a, 1997; Rygula et al., 2010), and of rodents (Schoenbaum et al., 2003; Saddoris et al., 2005; Schoenbaum et al., 2007) impair ODR. The degree of homology between the OFC in old world primates such as macaques and humans on the one hand and rodents and new world primates such as marmosets has been questioned (Passingham and Wise, 2012). Nevertheless, in both rodents and macaques there is now evidence that selective lesions of the amygdala lead to improvements, not impairments, in ODR performance (Stalnaker et al., 2007; Rudebeck and Murray, 2008; Izquierdo et al., 2013).

By training macaques to perform ODR in the MRI scanner we attempted to obtain a new perspective into the neural basis of flexible reward-guided behaviour in primates. We found a region in orbital area 12/47, the most lateral part of OFC (lOFC; **Figure 28**a, b), that showed a particularly strong WSLS signal as monkeys used the current outcome of a choice to adaptively guide future behaviour (**Figure 28**c). The effect cannot be related to a verbal mediation strategy dependent on adjacent ventral frontal brain regions associated with language because the lOFC data come from macaques which are non-linguistic. Stronger signals in this region were related to testing sessions with higher learning rates (**Figure 30**a, c). The location of the lOFC region we identified has important implications for understanding some of the past controversy and reconciling apparently conflicting patterns of results. The region lies just lateral to the cOFC region (areas 11 and 13) where the effects of neurotoxic lesions were studied (Kazama and Bachevalier, 2009; Rudebeck et al., 2013b). It would, however, be likely to have been partially deafferented by aspiration lesions in the cOFC region; connections running to and from this region in the amygdalofugal pathway and uncinate fascicle are immediately subjacent to the cOFC (Croxson et al., 2005; Lehman et al., 2011; Jbabdi et al., 2013). Although the amygdala's connections to medial OFC are often

emphasized, it is well established that the lOFC region that we identified as carrying a WSLS is also interconnected with the amygdala (Carmichael and Price, 1995a). Further, and unlike more medial OFC, lOFC is also interconnected with the inferior temporal and perirhinal cortex areas that are critically important for the visual processing of stimuli that are to be associated with reward (Carmichael and Price, 1995b; Kondo et al., 2005) in a task such as ODR.

We do not argue that the lOFC is the only OFC region important for flexible reward-guided behaviour. Our contention is, however, that the lOFC has a special importance that may explain why lesions that do not include it or deafferent it should not cause impairments. Indeed in an early study by Butter (1969), poor object reversal learning was not seen in monkeys with posteromedial OFC or anterior OFC lesions, but only observed in monkeys with total OFC or lOFC lesions (including the same part area 12/47 that we identify here). The human homolog of the same lOFC region was also found to be active in human fMRI ODR experiments (O'Doherty et al., 2001; Ghahremani et al., 2010).

At the same time that the importance of OFC for ODR has been questioned it has been suggested that other frontal regions, such as ACC, may play the same role that was previously associated with OFC (Chudasama et al., 2013). Although other studies have suggested that OFC and ACC contributions to reward-guided decision making are distinct in other tasks (Rudebeck et al., 2008; Kennerley et al., 2009; Luk and Wallis, 2013) Chudasama and colleagues have presented evidence suggesting that both lOFC and ACC are important for ODR. In the current study we found that like lOFC the ACC carried a signal that was related to impending behavioural change during ODR. In contrast to OFC, however, ACC did not carry a WSLS signal but instead carried a signal that predicted future choice shifting regardless of the outcome of the current choice (**Figure 31**a, b) and was related to exploratory or stochastic behaviour (**Figure 32**b, d). Such a pattern of activity is consistent

with an ACC role in determining the balance between continuation and maintenance of behaviour on the one hand and exploratory behaviour on the other (Daw et al., 2006; Kennerley et al., 2006; Amiez et al., 2012) and in taking choices that are alternatives to the recent standard or default pattern (Hayden et al., 2011; Kolling et al., 2012; Kolling et al., 2014).

We found that the amygdala also carried task-relevant signals that guided adaptive behavioural change. These signals, however, were only prominent at early stages of learning (**Figure 35**a, b) and when the monkeys failed to obtain a reward (**Figure 36**a, b). Moreover, we were able to show that activity in the lOFC and amygdala was more strongly coupled as a function of a lose-shift signal observed when animals were going to shift to an alternative option after failing to obtain rewards (**Figure 37**a, c). Although the relatively poor temporal resolution means that BOLD coupling analyses are unable to infer the direction of signal propagation, it is likely that the lose-shift signal in the lOFC originates, at least in part, from the amygdala. Morrison and colleagues (2011) recorded both OFC and amygdala neuronal activity with the high temporal precision that electrophysiological recording techniques afford and were able to show that the amygdala exerted a stronger influence on OFC activity, as opposed to an OFC-to-amygdala influence, when option value updating was required.

Our experimental design also allowed us to understand why amygdala lesions might, in some cases, improve ODR performance (Stalnaker et al., 2007; Rudebeck and Murray, 2008). This is because we also found a second type of signal in the amygdala that might be disruptive to task performance and which might be suppressed by the OFC. It is becoming clear that in a number of brain areas reward memories are represented over multiple time-scales – a short time-scale that contains information about the most recent reward only and over long time-scales that integrate over both the recent and remote histories of reward events (Bernacchia et al., 2011; Wittmann et al., submitted). Multiple time-scale reward

information can be useful for assessing whether, on average, the environment is getting better or worse but it can be problematic if long time-scale reward memories interfere with the process whereby specific reward events are associated with specific preceding events. Under such circumstances learning is slowed down by a "spread of reward" effect (Thorndike, 1933), such that the credit for a previous reward delivered after a previous choice is mistakenly assigned to the current choice. We showed that the amygdala contained a signal that encoded the outcome of the previous trial even though this was no longer relevant to assessing the correctness of the current trial's response (**Figure 38**c). The lOFC did not carry a similar signal (**Figure 38**a) and in fact coupling between lOFC and amygdala became weaker when the amygdala carried a signal that was related to the irrelevant previous outcome (**Figure 39**a). Note that the negative lOFC-amygdala coupling that occurs as a function of prior rewards is of the opposite sign to the positive lOFC-amygdala coupling that occurs as a function of lose-shift signals. Because both patterns of lOFC-amygdala coupling are not simply correlations in overall activity levels, but instead arise in relation to particular event-related signals they can occur simultaneously. This suggests that the lOFC could also have a role in selecting reward memories in a relevant time-scale (i.e. memory that only involved the immediate reward in this experiment) at the expense of other irrelevant time scales (i.e. memories that involved the outcome of the previous trial in this experiment). Consistent with this interpretation we found that more negative modulation in lOFC-amygdala coupling, as a function of the previous reward, was related to more win-stay behaviour (**Figure 39**b, c). This supports the view that the lOFC could interact with other brain regions to select relevant information and reject irrelevant information to generate adaptive behaviours (Chau et al., 2014). If this view is correct then it would also predict that lOFC lesions would lead animals to be more likely to make fallacious links between a current choice and a previous reward than control animals and indeed this is exactly what has been observed (Noonan et al., 2010; Walton et al., 2010).

In summary, we show that lOFC, ACC, and amygdala all carry signals related to reward-guided behavioural change but that there are fundamental differences in the nature of the signals carried by each area. LOFC activity is consistently related to the use of WSLS rules that facilitate good task performance but in some other cases the signals are detrimental to good deterministic ODR performance. For example, ACC activity is related to exploratory behaviour and an aspect of amygdala activity is related to previous reward outcomes that are irrelevant for assessing the success of the current choice. Variations in the strengths of signals in each area and in inter-areal signal coupling are associated with variations in different aspects of performance.

Chapter 6. General Discussions

In this thesis, I reported experiments that investigated decision making and reward-guided behavioural adaptation using model simulations and functional magnetic resonance imaging (fMRI). In a first series of experiments, I examined the differential computations of ventromedial prefrontal cortex (vmPFC) and the medial intraparietal region (MIP) in relation to irrational choice patterns exemplified in multiple alternative decisions in humans and in biophysical model predictions (Chapters 2-4). In a second major study, I investigated a recent controversy concerning the exact neural substrates that subserve behavioural adaptation. By performing fMRI in macaques, I identified a lateral orbitofrontal cortex (lOFC) region that is important for using reward information to flexibly guide behavioural change (Chapter 5).

6.1 Multialternative Decision Making

I investigated how rational choice could be violated during multiple alternative decision making by using a task in which choices were made between two options in the presence of a third option distractor. Biophysical model simulations (Chapter 2), behavioural testing (Chapter 3) and neuroimaging in human (Chapter 4) were performed. I showed that a biophysical model of decision making through mutual inhibition predicts that a key decision variable – the signal encoding the difference in value between two choices – is *reduced* in the presence of a low value third alternative. This occurred as the poor third alternative imposes only weak inhibition on both better and worse potential choices, leaving them both relatively active (Figures 4, 6c). Both behaviour (Figures 11, 13) and value difference signals in vmPFC (**Figure 19**) in humans corresponded to model predictions. The effects were

consistent no matter whether the third option was an unchooseable distractor or a chooseable option with poor values.

The effect of the distractor on the value difference signal recorded in the vmPFC and choice accuracy in behaviour were opposite to the predictions of the divisive normalization model. However, when the BOLD data was re-analyzed based on the spatial location of the stimuli, I found that the value signal in MIP relating to the value of the chosen option on the contralateral response side was modulated by the total value of options in the ipsilateral side (**Figure 24**). In other words, the chosen value signal in the contralateral side, thought to be within the response field of the MIP, was normalized by the total option value in the ipsilateral side, thought to be outside the response field of the MIP: a hallmark of divisive normalisation. The dissociable effects of the alternative options on the vmPFC and MIP signals could be due to the organization of their neural networks – there is a somatotopic map in the MIP and this may be a prerequisite for divisive normalization to occur, whereas no evidence for topography is found in the vmPFC.

In Chapter 3, additional experiments showed that the value of a third, non-transient, choosable option exerts two kinds of influences on behaviour depending on the context. First, there was a specific behavioural effect on "incongruent" trials, where to make an appropriate choice, subjects had to integrate across decision dimensions as one option was associated with a greater reward magnitude whereas the other option with a greater reward probability. The presence of a *poorer* third option on incongruent trials was related to *more* suboptimal decisions, as predicted by the biophysical model (**Figure 17**). A previous study has shown that the vmPFC is particularly active on incongruent trials (Hunt et al., 2012). Second, on "congruent" trials where one option is associated with both a greater reward magnitude *and* probability than the next best alternative, the presence of a *poorer* third option was related to *less* suboptimal decisions, which is consistent with divisive

normalization predictions made in the parietal cortex (Louie et al., 2011). It is possible that the third option effect on incongruent trials was driven by the vmPFC, whereas the third option effect on congruent trials was driven by the MIP.

Furthermore, it is argued that there are multiple decision making systems that operate in parallel to select the course of behaviour (Kable and Glimcher, 2009; Rushworth et al., 2012). The decisions of monkeys with vmPFC lesions, but not normal controls, are influenced by the third option in a manner predicted by divisive normalization (Noonan et al., 2010). In our study, individual variations in both vmPFC (**Figure 22**) and MIP (***Figure 25***) were correlated with accuracy and reaction time respectively. Taken together, our study contributes to an understanding of how vmPFC and PPC contribute to different aspects of behaviour under different contexts.

6.2 Behavioural Adaptation

The OFC and amygdala have long been thought to have a role in reward-guided behaviour. However, this view has become controversial after recent studies showed that neurotoxic ablations of the OFC and amygdala in macaques do not cause impairment in object reversal task performance (Izquierdo and Murray, 2007; Kazama and Bachevalier, 2009; Rudebeck et al., 2013b). Adding to the confusion regarding the actual role of these regions is that object reversal impairments can be found in marmosets and rodents with OFC neurotoxic lesion (Dias et al., 1996b; Schoenbaum et al., 2003; Saddoris et al., 2005; Schoenbaum et al., 2007; Rygula et al., 2010). Rodents or macaques with amygdala lesion even showed improvements in object reversal performance in some studies (Stalnaker et al., 2007; Rudebeck and Murray, 2008; Izquierdo et al., 2013).

I attempted to obtain a new perspective into the neural mechanism of behavioural adaptation by recording neural activity from the whole brain while macaques were performing in an object reversal task. I found a region in the lOFC (**Figure 28**a, b) that was particularly active when monkeys used the information of the current outcome to guide future choice (**Figure 28**c). Stronger signals in this region were related to testing sessions with higher learning rates (**Figure 30**a, c). It should be noted that I do not argue that the lOFC is the only important division in the OFC for guiding behaviour. My notion is that the lOFC has a special role such that impairment in object reversal learning observed in many cOFC lesion studies that used an aspiration technique could be a consequence of deafferentation of the lOFC region.

The anterior cingulate cortex (ACC) is also implicated in reward-guided decision making (Rudebeck et al., 2008; Kennerley et al., 2009; Luk and Wallis, 2013). In contrast to the lOFC, I showed that the ACC signal was related to future choice shifting regardless of the outcome of the current choice (**Figure 31**) and was related to exploratory or stochastic behaviour (**Figure 32**b, d). Such a pattern of activity is consistent with an ACC role in determining the balance between exploitative and exploratory behaviour and for computing an optimal choice shifting strategy (Daw et al., 2006; Kennerley et al., 2006; Hayden et al., 2011; Amiez et al., 2012; Kolling et al., 2012; Kolling et al., 2014).

Like the lOFC, the amygdala also carried signals that guided object reversal learning. These signals, however, were only prominent when there was a stronger demand for learning (**Figure 35**a, b) and when the monkeys failed to obtain a reward (**Figure 36**a, b). An advantage of performing whole brain activity recording is that it enables investigation of the interactions between brain regions under different task demands. By performing psychophysical interaction (PPI) analysis, it was possible to show that the lOFC and amygdala exhibited stronger coupling as a function of a lose-shift signal observed when animals were

going to shift to an alternative option after failing to obtain rewards (**Figure 37**). It is possible that a lose-shift signal is emerged from the amygdala and is passed on to the lOFC for guiding behavioural adaptation.

6.3 A Neural mechanism for Selecting Task Relevant Parameters

It is noticeable that in the human decision making experiment and the monkey object reversal experiment, the lOFC was functionally connected with other brain regions as a function of task relevant parameters. In Chapter 5, the information encoded by monkeys' amygdala signals included memory of outcomes related to the current choice *and* to the previous choice (**Figure 38**c). During deterministic object reversal learning, it is important to update an association between the *current* outcome and a choice, but not between the *previous* outcome and this choice, to allow appropriate, precise credit assignment to take place. The lOFC and amygdala were *decoupled* as a function of the memory of the previous outcome. Therefore, it is possible that the lOFC has a role avoiding the task irrelevant reward memory during credit assignment. Similarly in Chapter 4, human subjects computed a relevant value difference between the available options (HV-LV) and an irrelevant value difference between an available option and the distractor (HV-D) (Figure 23). The lOFC and vmPFC strongly coupled as a function of HV-LV and decoupled as a function of HV-D. Although functional connectivity analysis cannot be used to infer causality, it is possible that the lOFC exerted a selection pressure to filter out the task irrelevant information. There is multitude of evidence supporting this view. The lOFC is active when there is a demand for attention selection or set shifting in humans (Dias et al., 1996a; Monchi et al., 2001; Arana et al., 2003; Ghahremani et al., 2010; Hampshire et al., 2012). Old world monkeys with lOFC lesion or new world monkeys with OFC lesions are impaired in forgoing a previously

rewarded option that is no longer rewarded now (Butter, 1969; Iversen and Mishkin, 1970; Roberts and Wallis, 2000; Man et al., 2009).

6.4 Future Directions

In Chapter 3, I showed that whether a third option has an impact on decision making in a way that follows the biophysical model prediction or the divisive normalization model prediction depends on whether there is a need for integrating decision attributes (**Figure 17**). On the other hand, the fMRI analyses Figures 19 and 24 show that vmPFC and MIP follow the biophysical model and divisive model predictions respectively. Although choice accuracies were strongly related to vmPFC activities in general (**Figure 22**), it is possible that accuracies could become related to MIP activities when the analysis was performed focusing on trials that follow divisive normalization prediction (i.e. when there is no need to integrate decision attributes for making a decision).

A recent study reported that there are different neural systems related to different types of decisions and the inferior frontal gyrus has a role in selecting between those neural systems to drive behaviour depending on task demands (Kolling et al., 2014). In the future, an analysis could be performed to investigate whether the inferior frontal gyrus has a similar role of gating between vmPFC and MIP signals during multialternative decision making and whether the switch operates according to the context (i.e. whether there is a need for integrating decision attributes).

Our data in Chapter 5 showed that the amygdala carried a signal related to past outcome history and this signal was functionally decoupled from the lOFC. This suggests that the lOFC could have a role in avoiding any non-contingent learning signals during credit assignment. Further experiments and analyses are required for an in-depth examination for the actual neural mechanism of the spread of reward effects. For example, an analysis could

be performed to search for neural representations of specific stimulus-reward associations and to test how the representations could be strengthened by non-contingent reward signal. This kind of test will involve non-conventional methods such as multivariate voxel pattern analysis, because different stimulus-reward associations could be represented in the same brain region and generate the same univariate signal (Klein-Flugge et al., 2013). Once the multivariate signals for the specific stimulus-reward associations are identified, I could investigate the neural mechanisms of non-contingent learning by testing whether these signals are enhanced by the amygdala signal that is related to past outcome.

An experiment for studying spread of reward effects requires dissociation between contingent and non-contingent learning. One problem with using the object reversal task to study non-contingent learning is that it requires subjects to select repeatedly the same option for most of the time between reversal periods. This is problematic because the contingent learning and the non-contingent learning effects of a given reward are both assigned to the same option if a choice is repeated between consecutive trials. A powerful experiment for investigating non-contingent learning should require animals to sample different options continuously throughout the whole task so that the contingent and non-contingent learning effects can be decorrelated. For example, an experiment could involve three objects with different reward probabilities within a task while on any given trial only two of the objects are presented for a making decision. This could avoid animals choosing the same best option at all times – they have to choose the second best option when the best option is unavailable.

Bibliography

Aggleton JP, Passingham RE (1981) Syndrome Produced by Lesions of the Amygdala in Monkeys (Macaca Mulatta). Journal of Comparative and Physiological Psychology 95:961-977.

Amiez C, Sallet J, Procyk E, Petrides M (2012) Modulation of feedback related activity in the rostral anterior cingulate cortex during trial and error exploration. Neuroimage 63:1078-1090.

Arana FS, Parkinson JA, Hinton E, Holland AJ, Owen AM, Roberts AC (2003) Dissociable contributions of the human amygdala and orbitofrontal cortex to incentive motivation and goal selection. J Neurosci 23:9632-9638.

Ashbridge E, Walsh V, Cowey A (1997) Temporal aspects of visual search studied by transcranial magnetic stimulation. Neuropsychologia 35:1121-1131.

Ballard K, Knutson B (2009) Dissociable neural representations of future reward magnitude and delay during temporal discounting. Neuroimage 45:143-150.

Baxter MG, Parker A, Lindner CC, Izquierdo AD, Murray EA (2000) Control of response selection by reinforcer value requires interaction of amygdala and orbital prefrontal cortex. J Neurosci 20:4311-4319.

Bays PM, Singh-Curry V, Gorgoraptis N, Driver J, Husain M (2010) Integration of goal- and stimulus-related visual signals revealed by damage to human parietal cortex. J Neurosci 30:5968-5978.

Bechara A, Damasio H, Damasio AR (2000) Emotion, decision making and the orbitofrontal cortex. Cereb Cortex 10:295-307.

Bechara A, Damasio AR, Damasio H, Anderson SW (1994) Insensitivity to future consequences following damage to human prefrontal cortex. Cognition 50:7-15.

Bechara A, Tranel D, Damasio H, Damasio AR (1996) Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. Cereb Cortex 6:215-225.

Beckmann CF, Jenkinson M, Smith SM (2003) General multilevel linear modeling for group analysis in FMRI. Neuroimage 20:1052-1063.

Behrens TE, Woolrich MW, Walton ME, Rushworth MF (2007) Learning the value of information in an uncertain world. Nat Neurosci 10:1214-1221.

Bernacchia A, Seo H, Lee D, Wang XJ (2011) A reservoir of time constants for memory traces in cortical neurons. Nat Neurosci 14:366-372.

Blatt GJ, Andersen RA, Stoner GR (1990) Visual receptive field organization and cortico-cortical connections of the lateral intraparietal area (area LIP) in the macaque. The Journal of comparative neurology 299:421-445.

Bonin V, Mante V, Carandini M (2005) The suppressive field of neurons in lateral geniculate nucleus. J Neurosci 25:10844-10856.

Boorman ED, Rushworth MF, Behrens TE (2013) Ventromedial Prefrontal and Anterior Cingulate Cortex Adopt Choice and Default Reference Frames during Sequential Multi-Alternative Choice. The Journal of Neuroscience 33:2242-2253.

Boorman ED, Behrens TE, Woolrich MW, Rushworth MF (2009) How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. Neuron 62:733-743.

Bouret S, Richmond BJ (2010) Ventromedial and Orbital Prefrontal Neurons Differentially Encode Internally and Externally Driven Motivational Values in Monkeys. J Neurosci 30:8591-8601.

Burke SN, Thome A, Plange K, Engle JR, Trouard TP, Gothard KM, Barnes CA (2014) Orbitofrontal cortex volume in area 11/13 predicts reward devaluation, but not reversal learning performance, in young and aged monkeys. J Neurosci 34:9905-9916.
Busemeyer JR, Townsend JT (1993) Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. Psychol Rev 100:432-459.
Butter CM (1969) Perseveration in extinction and in discrimination reversal tasks following selective frontal ablations in Macaca mulatta. Physiology & Behavior 4:163-171.
Camille N, Griffiths CA, Vo K, Fellows LK, Kable JW (2011) Ventromedial frontal lobe damage disrupts value maximization in humans. J Neurosci 31:7527-7532.
Caplin A, Dean M (2008a) Axiomatic methods, dopamine and reward prediction error. Curr Opin Neurobiol 18:197-202.
Caplin A, Dean M (2008b) Dopamine, reward prediction error, and economics. Q J Econ 123:663-701.
Carandini M, Heeger DJ (2012) Normalization as a canonical neural computation. Nat Rev Neurosci 13:51-62.
Carmichael ST, Price JL (1994) Architectonic subdivision of the orbital and medial prefrontal cortex in the macaque monkey. The Journal of comparative neurology 346:366-402.
Carmichael ST, Price JL (1995a) Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. The Journal of comparative neurology 363:615-641.
Carmichael ST, Price JL (1995b) Sensory and premotor connections of the orbital and medial prefrontal cortex of macaque monkeys. The Journal of comparative neurology 363:642-664.
Carmichael ST, Price JL (1996) Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. The Journal of comparative neurology 371:179-207.
Chau BK, Kolling N, Hunt LT, Walton ME, Rushworth MF (2014) A neural mechanism underlying failure of optimal choice with multiple alternatives. Nat Neurosci 17:463-470.
Chudasama Y, Daniels TE, Gorrin DP, Rhodes SE, Rudebeck PH, Murray EA (2013) The role of the anterior cingulate cortex in choices based on reward value and reward contingency. Cereb Cortex 23:2884-2898.
Churchland AK, Ditterich J (2012) New advances in understanding decisions among multiple alternatives. Curr Opin Neurobiol 22:920-926.
Churchland AK, Kiani R, Shadlen MN (2008) Decision-making with multiple alternatives. Nat Neurosci 11:693-702.
Clark L, Bechara A, Damasio H, Aitken MRF, Sahakian BJ, Robbins TW (2008) Differential effects of insular and ventromedial prefrontal cortex lesions on risky decision-making. Brain 131:1311-1322.
Clarke HF, Cardinal RN, Rygula R, Hong YT, Fryer TD, Sawiak SJ, Ferrari V, Cockcroft G, Aigbirhio FI, Robbins TW, Roberts AC (2014) Orbitofrontal dopamine depletion upregulates caudate dopamine and alters behavior via changes in reinforcement sensitivity. J Neurosci 34:7663-7676.
Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N (2012) Neuron-type-specific signals for reward and punishment in the ventral tegmental area. Nature 482:85-88.
Cohen YE, Andersen RA (2002) A common reference frame for movement plans in the posterior parietal cortex. Nat Rev Neurosci 3:553-562.
Croxson PL, Johansen-Berg H, Behrens TE, Robson MD, Pinsk MA, Gross CG, Richter W, Richter MC, Kastner S, Rushworth MF (2005) Quantitative investigation of connections of the prefrontal cortex in the human and macaque using probabilistic diffusion tractography. J Neurosci 25:8854-8866.

Damasio AR (1994) Descartes' Error : Emotion, Reason and the Human Brain. New York: Grosset/Putnam.
Damasio AR (1996) The somatic marker hypothesis and the possible functions of the prefrontal cortex. Philos Trans R Soc Lond B Biol Sci 351:1413-1420.
Damasio AR, Tranel D, Damasio H (1990) Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. Behav Brain Res 41:81-94.
Damasio H, Grabowski T, Frank R, Galaburda AM, Damasio AR (1994) The return of Phineas Gage: clues about the brain from the skull of a famous patient. Science 264:1102-1105.
Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. Nature 441:876-879.
De Martino B, Fleming SM, Garrett N, Dolan RJ (2013) Confidence in value-based choice. Nat Neurosci 16:105-110.
de Wit S, Corlett PR, Aitken MR, Dickinson A, Fletcher PC (2009) Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. J Neurosci 29:11330-11338.
Deichmann R, Gottfried JA, Hutton C, Turner R (2003) Optimized EPI for fMRI studies of the orbitofrontal cortex. Neuroimage 19:430-441.
Depireux DA, Simon JZ, Klein DJ, Shamma SA (2001) Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. J Neurophysiol 85:1220-1234.
Dias R, Robbins TW, Roberts AC (1996a) Dissociation in prefrontal cortex of affective and attentional shifts. Nature 380:69-72.
Dias R, Robbins TW, Roberts AC (1996b) Primate analogue of the Wisconsin Card Sorting Test: effects of excitotoxic lesions of the prefrontal cortex in the marmoset. Behav Neurosci 110:872-886.
Dias R, Robbins TW, Roberts AC (1997) Dissociable forms of inhibitory control within prefrontal cortex with an analog of the Wisconsin Card Sort Test: restriction to novel situations and independence from "on-line" processing. The Journal of Neuroscience 17:9285-9297.
Dorris MC, Glimcher PW (2004) Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. Neuron 44:365-378.
Essex BG, Clinton SA, Wonderley LR, Zald DH (2012) The impact of the posterior parietal and dorsolateral prefrontal cortices on the optimization of long-term versus immediate value. J Neurosci 32:15403-15413.
Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1-47.
Fellows LK (2006) Deciding how to decide: ventromedial frontal lobe damage affects information acquisition in multi-attribute decision making. Brain 129:944-952.
Fellows LK, Farah MJ (2007) The role of ventromedial prefrontal cortex in decision making: judgment under uncertainty or judgment per se? Cereb Cortex 17:2669-2674.
FitzGerald TH, Seymour B, Dolan RJ (2009) The role of human orbitofrontal cortex in value comparison for incommensurable objects. J Neurosci 29:8388-8395.
Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6:218-229.
Furman M, Wang XJ (2008) Similarity effect and optimal control of multiple-choice decision making. Neuron 60:1153-1168.
Ghahremani DG, Monterosso J, Jentsch JD, Bilder RM, Poldrack RA (2010) Neural components underlying behavioral flexibility in human reversal learning. Cereb Cortex 20:1843-1852.

Gold JI, Shadlen MN (2002) Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. Neuron 36:299-308.
Gold JI, Shadlen MN (2007) The neural basis of decision making. Annu Rev Neurosci 30:535-574.
Goldberg ME, Colby CL, Duhamel JR (1990) Representation of visuomotor space in the parietal lobe of the monkey. Cold Spring Harbor symposia on quantitative biology 55:729-739.
Gould IC, Nobre AC, Wyart V, Rushworth MF (2012) Effects of decision variables and intraparietal stimulation on sensorimotor oscillatory activity in the human brain. J Neurosci 32:13805-13818.
Green L, Myerson J (2004) A discounting framework for choice with delayed and probabilistic rewards. Psychological bulletin 130:769-792.
Grefkes C, Fink GR (2005) The functional organization of the intraparietal sulcus in humans and monkeys. Journal of Anatomy 207:3-17.
Grol MJ, de Lange FP, Verstraten FA, Passingham RE, Toni I (2006) Cerebral changes during performance of overlearned arbitrary visuomotor associations. J Neurosci 26:117-125.
Haber SN, Kunishio K, Mizobuchi M, Lynd-Balta E (1995) The orbital and medial prefrontal circuit through the primate basal ganglia. J Neurosci 15:4851-4867.
Haber SN, Kim KS, Mailly P, Calzavara R (2006) Reward-related cortical inputs define a large striatal region in primates that interface with associative cortical connections, providing a substrate for incentive-based learning. J Neurosci 26:8368-8376.
Hampshire A, Owen AM (2006) Fractionating attentional control using event-related fMRI. Cereb Cortex 16:1679-1689.
Hampshire A, Chaudhry AM, Owen AM, Roberts AC (2012) Dissociable roles for lateral orbitofrontal cortex and lateral prefrontal cortex during preference driven reversal learning. Neuroimage 59:4102-4112.
Hampton AN, Bossaerts P, O'Doherty JP (2006) The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. J Neurosci 26:8360-8367.
Hare TA, Camerer CF, Rangel A (2009) Self-control in decision-making involves modulation of the vmPFC valuation system. Science 324:646-648.
Hare TA, Malmaud J, Rangel A (2011) Focusing Attention on the Health Aspects of Foods Changes Value Signals in vmPFC and Improves Dietary Choice. J Neurosci 31:11077-11087.
Hart AS, Rutledge RB, Glimcher PW, Phillips PE (2014) Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. J Neurosci 34:698-704.
Hayden BY, Pearson JM, Platt ML (2011) Neuronal basis of sequential foraging decisions in a patchy environment. Nat Neurosci 14:933-939.
Heeger DJ (1992) Normalization of cell responses in cat striate cortex. Visual neuroscience 9:181-197.
Heeger DJ, Simoncelli EP, Movshon JA (1996) Computational models of cortical visual processing. Proc Natl Acad Sci U S A 93:623-627.
Hollerman JR, Schultz W (1998) Dopamine neurons report an error in the temporal prediction of reward during learning. Nat Neurosci 1:304-309.
Huber J, Payne JW, Puto C (1982) Adding Asymmetrically Dominated Alternatives - Violations of Regularity and the Similarity Hypothesis. J Consum Res 9:90-98.
Hunt LT, Kolling N, Soltani A, Woolrich MW, Rushworth MF, Behrens TE (2012) Mechanisms underlying cortical activity during value-guided choice. Nat Neurosci 15:470-476, S471-473.

Husain M, Rorden C (2003) Non-spatially lateralized mechanisms in hemispatial neglect. Nat Rev Neurosci 4:26-36.
Iversen SD, Mishkin M (1970) Perseverative interference in monkeys following selective lesions of the inferior prefrontal convexity. Exp Brain Res 11:376-386.
Izquierdo A, Murray EA (2004) Combined unilateral lesions of the amygdala and orbital prefrontal cortex impair affective processing in rhesus monkeys. J Neurophysiol 91:2023-2039.
Izquierdo A, Murray EA (2007) Selective bilateral amygdala lesions in rhesus monkeys fail to disrupt object reversal learning. J Neurosci 27:1054-1062.
Izquierdo A, Suda RK, Murray EA (2004) Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. J Neurosci 24:7540-7548.
Izquierdo A, Darling C, Manos N, Pozos H, Kim C, Ostrander S, Cazares V, Stepp H, Rudebeck PH (2013) Basolateral amygdala lesions facilitate reward choices after negative feedback in rats. J Neurosci 33:4105-4109.
Jbabdi S, Lehman JF, Haber SN, Behrens TE (2013) Human and monkey ventral prefrontal fibers use the same organizational principles to reach their targets: tracing versus tractography. J Neurosci 33:3190-3201.
Jenkinson M (2003) Fast, automated, N-dimensional phase-unwrapping algorithm. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 49:193-197.
Jenkinson M, Smith S (2001) A global optimisation method for robust affine registration of brain images. Medical image analysis 5:143-156.
Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17:825-841.
Jocham G, Hunt LT, Near J, Behrens TE (2012) A mechanism for value-guided choice based on the excitation-inhibition balance in prefrontal cortex. Nat Neurosci 15:960-961.
Johansen JP, Hamanaka H, Monfils MH, Behnia R, Deisseroth K, Blair HT, LeDoux JE (2010) Optical activation of lateral amygdala pyramidal cells instructs associative fear learning. Proc Natl Acad Sci U S A 107:12692-12697.
Johnson PB, Ferraina S, Bianchi L, Caminiti R (1996) Cortical networks for visual reaching: physiological and anatomical organization of frontal and parietal lobe arm regions. Cereb Cortex 6:102-119.
Jones B, Mishkin M (1972) Limbic lesions and the problem of stimulus--reinforcement associations. Experimental neurology 36:362-377.
Kable JW, Glimcher PW (2007) The neural correlates of subjective value during intertemporal choice. Nat Neurosci 10:1625-1633.
Kable JW, Glimcher PW (2009) The neurobiology of decision: consensus and controversy. Neuron 63:733-745.
Kahneman D, Tversky A (1979) Prospect theory - analysis of decision under risk. Econometrica 47:263-291.
Kahnt T, Park SQ, Haynes JD, Tobler PN (2014) Disentangling neural representations of value and salience in the human brain. Proc Natl Acad Sci U S A 111:5000-5005.
Kaping D, Vinck M, Hutchison RM, Everling S, Womelsdorf T (2011) Specific contributions of ventromedial, anterior cingulate, and lateral prefrontal cortex for attentional selection and stimulus valuation. PLoS biology 9:e1001224.
Kazama A, Bachevalier J (2009) Selective aspiration or neurotoxic lesions of orbital frontal areas 11 and 13 spared monkeys' performance on the object discrimination reversal task. J Neurosci 29:2794-2804.

Kennerley SW, Wallis JD (2009) Reward-dependent modulation of working memory in lateral prefrontal cortex. J Neurosci 29:3259-3270.
Kennerley SW, Dahmubed AF, Lara AH, Wallis JD (2009) Neurons in the frontal lobe encode the value of multiple decision variables. J Cogn Neurosci 21:1162-1178.
Kennerley SW, Walton ME, Behrens TE, Buckley MJ, Rushworth MF (2006) Optimal decision making and the anterior cingulate cortex. Nat Neurosci 9:940-947.
Klein-Flugge MC, Barron HC, Brodersen KH, Dolan RJ, Behrens TE (2013) Segregated encoding of reward-identity and stimulus-reward associations in human orbitofrontal cortex. J Neurosci 33:3202-3211.
Knutson B, Taylor J, Kaufman M, Peterson R, Glover G (2005) Distributed neural representation of expected value. J Neurosci 25:4806-4812.
Kolling N, Wittmann M, Rushworth MF (2014) Multiple Neural Mechanisms of Decision Making and Their Competition under Changing Risk Pressure. Neuron 81:1190-1202.
Kolling N, Behrens TE, Mars RB, Rushworth MF (2012) Neural mechanisms of foraging. Science 336:95-98.
Kolster H, Janssens T, Orban GA, Vanduffel W (2014) The Retinotopic Organization of Macaque Occipitotemporal Cortex Anterior to V4 and Caudoventral to the Middle Temporal (MT) Cluster. J Neurosci 34:10168-10191.
Kolster H, Mandeville JB, Arsenault JT, Ekstrom LB, Wald LL, Vanduffel W (2009) Visual field map clusters in macaque extrastriate visual cortex. J Neurosci 29:7031-7039.
Kondo H, Saleem KS, Price JL (2005) Differential connections of the perirhinal and parahippocampal cortex with the orbital and medial prefrontal networks in macaque monkeys. The Journal of comparative neurology 493:479-509.
Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci 12:535-540.
Kringelbach ML, Rolls ET (2004) The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. Prog Neurobiol 72:341-372.
Leathers ML, Olson CR (2012) In monkeys making value-based decisions, LIP neurons encode cue salience and not action value. Science 338:132-135.
Lebreton M, Jorge S, Michel V, Thirion B, Pessiglione M (2009) An automatic valuation system in the human brain: evidence from functional neuroimaging. Neuron 64:431-439.
Lee J, Maunsell JH (2009) A normalization model of attentional modulation of single unit responses. PLoS One 4:e4651.
Lee J, Maunsell JH (2010) Attentional modulation of MT neurons with single or multiple stimuli in their receptive fields. J Neurosci 30:3058-3066.
Lehman JF, Greenberg BD, McIntyre CC, Rasmussen SA, Haber SN (2011) Rules ventral prefrontal cortical axons use to reach their targets: implications for diffusion tensor imaging tractography and deep brain stimulation for psychiatric illness. J Neurosci 31:10392-10402.
Leite FP, Tsao D, Vanduffel W, Fize D, Sasaki Y, Wald LL, Dale AM, Kwong KK, Orban GA, Rosen BR, Tootell RB, Mandeville JB (2002) Repeated fMRI using iron oxide contrast agent in awake, behaving macaques at 3 Tesla. Neuroimage 16:283-294.
Lim SL, O'Doherty JP, Rangel A (2011) The decision value computations in the vmPFC and striatum use a relative value code that is guided by visual attention. J Neurosci 31:13214-13223.
Loewenstein GF (1988) Frames of Mind in Intertemporal Choice. Manage Sci 34:200-214.
Logan GD (2002) An instance theory of attention and memory. Psychological review 109:376-400.

Louie K, Glimcher PW (2010) Separating Value from Choice: Delay Discounting Activity in the Lateral Intraparietal Area. The Journal of Neuroscience 30:5498-5507.
Louie K, Glimcher PW (2012) Efficient coding and the neural representation of value. Ann N Y Acad Sci 1251:13-32.
Louie K, Grattan LE, Glimcher PW (2011) Reward value-based gain control: divisive normalization in parietal cortex. J Neurosci 31:10627-10639.
Louie K, Khaw MW, Glimcher PW (2013) Normalization is a general neural mechanism for context-dependent decision making. Proceedings of the National Academy of Sciences.
Luce RD (1959) Individual Choice Behavior: A Theoretical Analysis. New York: Wiley.
Luk CH, Wallis JD (2013) Choice Coding in Frontal Cortex during Stimulus-Guided or Action-Guided Decision-Making. J Neurosci 33:1864-1871.
Lynch JC, Graybiel AM, Lobeck LJ (1985) The differential projection of two cytoarchitectonic subregions of the inferior parietal lobule of macaque upon the deep layers of the superior colliculus. The Journal of comparative neurology 235:241-254.
Mackey S, Petrides M (2010) Quantitative demonstration of comparable architectonic areas within the ventromedial and lateral orbital frontal cortex in the human and the macaque monkey brains. Eur J Neurosci 32:1940-1950.
Malhotra P, Coulthard EJ, Husain M (2009) Role of right posterior parietal cortex in maintaining attention to spatial locations over time. Brain 132:645-660.
Man MS, Clarke HF, Roberts AC (2009) The role of the orbitofrontal cortex and medial striatum in the regulation of prepotent responses to food rewards. Cereb Cortex 19:899-906.
Man MS, Dalley JW, Roberts AC (2010) Opposing effects of 5,7-DHT infusions into the orbitofrontal cortex and amygdala on flexible responding. Cereb Cortex 20:1668-1675.
Maren S (2005) Synaptic mechanisms of associative memory in the amygdala. Neuron 47:783-786.
Mars RB, Jbabdi S, Sallet J, O'Reilly JX, Croxson PL, Olivier E, Noonan MP, Bergmann C, Mitchell AS, Baxter MG, Behrens TE, Johansen-Berg H, Tomassini V, Miller KL, Rushworth MF (2011) Diffusion-weighted imaging tractography-based parcellation of the human parietal cortex and comparison with human and macaque resting-state functional connectivity. J Neurosci 31:4087-4100.
Mattingley JB, Husain M, Rorden C, Kennard C, Driver J (1998) Motor role of human inferior parietal lobe revealed in unilateral neglect patients. Nature 392:179-182.
Maunsell JH, Van Essen DC (1983) Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. J Neurophysiol 49:1127-1147.
Mirenowicz J, Schultz W (1996) Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. Nature 379:449-451.
Monchi O, Petrides M, Petre V, Worsley K, Dagher A (2001) Wisconsin Card Sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. J Neurosci 21:7733-7741.
Monosov IE, Hikosaka O (2012) Regionally distinct processing of rewards and punishments by the primate ventromedial prefrontal cortex. J Neurosci 32:10318-10330.
Morrison SE, Salzman CD (2009) The convergence of information about rewarding and aversive stimuli in single neurons. J Neurosci 29:11471-11483.
Morrison SE, Saez A, Lau B, Salzman CD (2011) Different time courses for learning-related changes in amygdala and orbitofrontal cortex. Neuron 71:1127-1140.
Murray EA, Izquierdo A (2007) Orbitofrontal cortex and amygdala contributions to affect and action in primates. Ann N Y Acad Sci 1121:273-296.

Murray EA, O'Doherty JP, Schoenbaum G (2007) What We Know and Do Not Know about the Functions of the Orbitofrontal Cortex after 20 Years of Cross-Species Studies. J Neurosci 27:8166-8169.
Nassar MR, Wilson RC, Heasly B, Gold JI (2010) An Approximately Bayesian Delta-Rule Model Explains the Dynamics of Belief Updating in a Changing Environment. The Journal of Neuroscience 30:12366-12378.
Nassar MR, Rumsey KM, Wilson RC, Parikh K, Heasly B, Gold JI (2012) Rational regulation of learning dynamics by pupil-linked arousal systems. Nat Neurosci 15:1040-1046.
Neubert FX, Mars R, Sallet J, Rushworth M (submitted) Connectivity profiles reveal relationship of brain areas for reward-guided learning and decision making in human and monkey frontal cortex.
Noonan MP, Mars RB, Rushworth MF (2011) Distinct roles of three frontal cortical areas in reward-guided behavior. J Neurosci 31:14399-14412.
Noonan MP, Kolling N, Walton ME, Rushworth MF (2012) Re-evaluating the role of the orbitofrontal cortex in reward and reinforcement. Eur J Neurosci 35:997-1010.
Noonan MP, Walton ME, Behrens TE, Sallet J, Buckley MJ, Rushworth MF (2010) Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. Proc Natl Acad Sci U S A 107:20547-20552.
O'Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C (2001) Abstract reward and punishment representations in the human orbitofrontal cortex. Nat Neurosci 4:95-102.
O'Reilly JX, Woolrich MW, Behrens TE, Smith SM, Johansen-Berg H (2012) Tools of the trade: psychophysiological interactions and functional connectivity. Social cognitive and affective neuroscience 7:604-609.
Ongur D, Price JL (2000) The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. Cereb Cortex 10:206-219.
Ongur D, An X, Price JL (1998) Prefrontal cortical projections to the hypothalamus in macaque monkeys. The Journal of comparative neurology 401:480-505.
Padoa-Schioppa C (2009) Range-adapting representation of economic value in the orbitofrontal cortex. J Neurosci 29:14004-14014.
Padoa-Schioppa C (2011) Neurobiology of economic choice: a good-based model. Annu Rev Neurosci 34:333-359.
Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. Nature 441:223-226.
Padoa-Schioppa C, Assad JA (2008) The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. Nat Neurosci 11:95-102.
Palminteri S, Boraud T, Lafargue G, Dubois B, Pessiglione M (2009) Brain hemispheres selectively track the expected value of contralateral options. J Neurosci 29:13465-13472.
Passingham RE, Wise SP (2012) The neurobiology of the prefrontal cortex : anatomy, evolution, and the origin of insight, 1st Edition. Oxford, United Kingdom: Oxford University Press.
Paton JJ, Belova MA, Morrison SE, Salzman CD (2006) The primate amygdala represents the positive and negative value of visual stimuli during learning. Nature 439:865-870.
Peters J, Buchel C (2009) Overlapping and distinct neural systems code for subjective value during intertemporal and risky decision making. J Neurosci 29:15727-15734.
Pettibone JC, Wedell DH (2007) Testing alternative explanations of phantom decoy effects. Journal of Behavioral Decision Making 20:323-341.
Philiastides MG, Biele G, Heekeren HR (2010) A mechanistic account of value computation in the human brain. Proc Natl Acad Sci U S A 107:9430-9435.

Plassmann H, O'Doherty J, Rangel A (2007) Orbitofrontal Cortex Encodes Willingness to Pay in Everyday Economic Transactions. J Neurosci 27:9984-9988.
Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex. Nature 400:233-238.
Preuss TM (1995) Do rats have prefrontal cortex? The rose-woolsey-akert program reconsidered. J Cogn Neurosci 7:1-24.
Rangel A, Camerer C, Montague PR (2008) A framework for studying the neurobiology of value-based decision making. Nat Rev Neurosci 9:545-556.
Ratcliff R, McKoon G (2008) The diffusion decision model: theory and data for two-choice decision tasks. Neural computation 20:873-922.
Reddi BA, Asrress KN, Carpenter RH (2003) Accuracy, information, and response time in a saccadic decision task. J Neurophysiol 90:3538-3546.
Rescorla R, Wagner A (1972) A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Classical Conditioning (Black AH, Prokasy WF, eds). New York: Appleton-Century-Crofts.
Reynolds JH, Heeger DJ (2009) The normalization model of attention. Neuron 61:168-185.
Rich EL, Wallis JD (2014) Medial-lateral organization of the orbitofrontal cortex. J Cogn Neurosci 26:1347-1362.
Robbins H (1952) Some Aspects of the Sequential Design of Experiments. B Am Math Soc 58:527-535.
Roberts AC (2006) Primate orbitofrontal cortex and adaptive behaviour. Trends Cogn Sci 10:83-90.
Roberts AC, Wallis JD (2000) Inhibitory control and affective processing in the prefrontal cortex: neuropsychological studies in the common marmoset. Cereb Cortex 10:252-262.
Roberts AC, Tomic DL, Parkinson CH, Roeling TA, Cutter DJ, Robbins TW, Everitt BJ (2007) Forebrain connectivity of the prefrontal cortex in the marmoset monkey (Callithrix jacchus): an anterograde and retrograde tract-tracing study. The Journal of comparative neurology 502:86-112.
Roe RM, Busemeyer JR, Townsend JT (2001) Multialternative decision field theory: a dynamic connectionist model of decision making. Psychol Rev 108:370-392.
Rolls ET (2000) The orbitofrontal cortex and reward. Cereb Cortex 10:284-294.
Rolls ET, Critchley HD, Mason R, Wakeman EA (1996) Orbitofrontal cortex neurons: role in olfactory and visual association learning. J Neurophysiol 75:1970-1981.
Romo R, Brody CD, Hernandez A, Lemus L (1999) Neuronal correlates of parametric working memory in the prefrontal cortex. Nature 399:470-473.
Rudebeck PH, Murray EA (2008) Amygdala and orbitofrontal cortex lesions differentially influence choices during object reversal learning. J Neurosci 28:8338-8343.
Rudebeck PH, Murray EA (2011a) Balkanizing the primate orbitofrontal cortex: distinct subregions for comparing and contrasting values. Ann N Y Acad Sci 1239:1-13.
Rudebeck PH, Murray EA (2011b) Dissociable effects of subtotal lesions within the macaque orbital prefrontal cortex on reward-guided behavior. J Neurosci 31:10569-10578.
Rudebeck PH, Mitz AR, Chacko RV, Murray EA (2013a) Effects of amygdala lesions on reward-value coding in orbital and medial prefrontal cortex. Neuron 80:1519-1531.
Rudebeck PH, Saunders RC, Prescott AT, Chau LS, Murray EA (2013b) Prefrontal mechanisms of behavioral flexibility, emotion regulation and value updating. Nat Neurosci 16:1140-1145.
Rudebeck PH, Behrens TE, Kennerley SW, Baxter MG, Buckley MJ, Walton ME, Rushworth MF (2008) Frontal cortex subregions play distinct roles in choices between actions and stimuli. J Neurosci 28:13775-13785.

Rushworth MF, Behrens TE (2008) Choice, uncertainty and value in prefrontal and cingulate cortex. Nat Neurosci 11:389-397.
Rushworth MF, Nixon PD, Passingham RE (1997a) Parietal cortex and movement. II. Spatial representation. Exp Brain Res 117:311-323.
Rushworth MF, Nixon PD, Eacott MJ, Passingham RE (1997b) Ventral prefrontal cortex is not essential for working memory. J Neurosci 17:4829-4838.
Rushworth MF, Behrens TE, Rudebeck PH, Walton ME (2007) Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. Trends Cogn Sci 11:168-176.
Rushworth MF, Kolling N, Sallet J, Mars RB (2012) Valuation and decision-making in frontal cortex: one or many serial or parallel systems? Curr Opin Neurobiol 22:946-955.
Rushworth MF, Noonan MP, Boorman ED, Walton ME, Behrens TE (2011) Frontal cortex and reward-guided learning and decision-making. Neuron 70:1054-1069.
Rushworth MF, Buckley MJ, Gough PM, Alexander IH, Kyriazis D, McDonald KR, Passingham RE (2005) Attentional selection and action selection in the ventral and orbital prefrontal cortex. J Neurosci 25:11628-11636.
Rutledge RB, Dean M, Caplin A, Glimcher PW (2010) Testing the reward prediction error hypothesis with an axiomatic model. J Neurosci 30:13525-13536.
Rygula R, Walker SC, Clarke HF, Robbins TW, Roberts AC (2010) Differential contributions of the primate ventrolateral prefrontal and orbitofrontal cortex to serial reversal learning. J Neurosci 30:14552-14559.
Rygula R, Clarke HF, Cardinal RN, Cockcroft GJ, Xia J, Dalley JW, Robbins TW, Roberts AC (2014) Role of Central Serotonin in Anticipation of Rewarding and Punishing Outcomes: Effects of Selective Amygdala or Orbitofrontal 5-HT Depletion. Cereb Cortex.
Saddoris MP, Gallagher M, Schoenbaum G (2005) Rapid associative encoding in basolateral amygdala depends on connections with orbitofrontal cortex. Neuron 46:321-331.
Sallet J, Mars RB, Noonan MP, Neubert FX, Jbabdi S, O'Reilly JX, Filippini N, Thomas AG, Rushworth MF (2013) The organization of dorsal frontal cortex in humans and macaques. J Neurosci 33:12255-12274.
Schoenbaum G, Saddoris MP, Stalnaker TA (2007) Reconciling the roles of orbitofrontal cortex in reversal learning and the encoding of outcome expectancies. Ann N Y Acad Sci 1121:320-335.
Schoenbaum G, Roesch MR, Stalnaker TA, Takahashi YK (2009) A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. Nat Rev Neurosci 10:885-892.
Schoenbaum G, Setlow B, Nugent SL, Saddoris MP, Gallagher M (2003) Lesions of orbitofrontal cortex and basolateral amygdala complex disrupt acquisition of odor-guided discriminations and reversals. Learn Mem 10:129-140.
Schultz W (1997) Dopamine neurons and their role in reward mechanisms. Curr Opin Neurobiol 7:191-197.
Schultz W (2012) Updating dopamine reward signals. Curr Opin Neurobiol.
Seo H, Lee D (2008) Cortical mechanisms for reinforcement learning in competitive games. Philos Trans R Soc Lond B Biol Sci 363:3845-3857.
Seo H, Barraclough DJ, Lee D (2009) Lateral intraparietal cortex and reinforcement learning during a mixed-strategy game. J Neurosci 29:7278-7289.
Sereno AB, Maunsell JHR (1998) Shape selectivity in primate lateral intraparietal cortex. Nature 395:500-502.
Sereno MI, Huang RS (2014) Multisensory maps in parietal cortex. Curr Opin Neurobiol 24:39-46.
Sereno MI, Pitzalis S, Martinez A (2001) Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. Science 294:1350-1354.

Shadlen MN, Newsome WT (1996) Motion perception: seeing and deciding. Proc Natl Acad Sci U S A 93:628-633.
Shadlen MN, Newsome WT (2001) Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J Neurophysiol 86:1916-1936.
Shadlen MN, Kiani R (2013) Decision making as a window on cognition. Neuron 80:791-806.
Shadlen MN, Britten KH, Newsome WT, Movshon JA (1996) A computational analysis of the relationship between neuronal and behavioral responses to visual motion. J Neurosci 16:1486-1510.
Silver MA, Kastner S (2009) Topographic maps in human frontal and parietal cortex. Trends Cogn Sci 13:488-495.
Simon HA (1956) Rational choice and the structure of the environment Psychological review 63:129-138.
Simonson I (1989) Choice Based on Reasons - the Case of Attraction and Compromise Effects. J Consum Res 16:158-174.
Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM (2004) Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 Suppl 1:S208-219.
Spiegler BJ, Mishkin M (1981) Evidence for the sequential participation of inferior temporal cortex and amygdala in the acquisition of stimulus-reward associations. Behav Brain Res 3:303-317.
Stalnaker TA, Franz TM, Singh T, Schoenbaum G (2007) Basolateral amygdala lesions abolish orbitofrontal-dependent reversal impairments. Neuron 54:51-58.
Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH (2013) A causal link between prediction errors, dopamine neurons and learning. Nat Neurosci 16:966-973.
Strait CE, Blanchard TC, Hayden BY (2014) Reward Value Comparison via Mutual Inhibition in Ventromedial Prefrontal Cortex. Neuron 82:1357-1366.
Sugrue LP, Corrado GS, Newsome WT (2004) Matching behavior and the representation of value in the parietal cortex. Science 304:1782-1787.
Sugrue LP, Corrado GS, Newsome WT (2005) Choosing the greater of two goods: neural currencies for valuation and decision making. Nat Rev Neurosci 6:363-375.
Summerfield C, Tsetsos K (2012) Building Bridges between Perceptual and Economic Decision-Making: Neural and Computational Mechanisms. Frontiers in neuroscience 6:70.
Sutton RS, Barto AG (1998) Reinforcement learning : an introduction. Cambridge, Mass.: MIT Press.
Thorndike EL (1933) A Proof of the Law of Effect. Science 77:173-175.
Thorpe SJ, Rolls ET, Maddison S (1983) The orbitofrontal cortex: neuronal activity in the behaving monkey. Exp Brain Res 49:93-115.
Tranel D, Damasio H, Damasio AR (1995) Double Dissociation between Overt and Covert Face Recognition. J Cogn Neurosci 7:425-432.
Tremblay L, Schultz W (1999) Relative reward preference in primate orbitofrontal cortex. Nature 398:704-708.
Tversky A (1969) Intransitivity of Preferences. Psychological Review 76:31-&.
Tversky A (1972) Elimination by Aspects - Theory of Choice. Psychological review 79:281-&.
Tversky A, Kahneman D (1991) Loss Aversion in Riskless Choice - a Reference-Dependent Model. Q J Econ 106:1039-1061.
Tversky A, Simonson I (1993) Context-Dependent Preferences. Manage Sci 39:1179-1189.

Usher M, McClelland JL (2001) The time course of perceptual choice: the leaky, competing accumulator model. Psychol Rev 108:550-592.
Usher M, McClelland JL (2004) Loss aversion and inhibition in dynamical models of multialternative choice. Psychol Rev 111:757-769.
Valentin VV, Dickinson A, O'Doherty JP (2007) Determining the neural substrates of goal-directed learning in the human brain. J Neurosci 27:4019-4026.
Vallar G (1998) Spatial hemineglect in humans. Trends Cogn Sci 2:87-97.
Vickery TJ, Chun MM, Lee D (2011) Ubiquity and Specificity of Reinforcement Signals throughout the Human Brain. Neuron 72:166-177.
von Neumann J, Morgenstern O (1944) Theory of Games and Economic Behavior. Princeton: Princeton University Press.
Walker SC, Robbins TW, Roberts AC (2009) Differential contributions of dopamine and serotonin to orbitofrontal cortex function in the marmoset. Cereb Cortex 19:889-898.
Wallis JD (2012) Cross-species studies of orbitofrontal cortex and value-based decision-making. Nat Neurosci 15:13-19.
Wallis JD, Kennerley SW (2010) Heterogeneous reward signals in prefrontal cortex. Curr Opin Neurobiol 20:191-198.
Walton ME, Devlin JT, Rushworth MFS (2004) Interactions between decision making and performance monitoring within prefrontal cortex. Nat Neurosci 7:1259-1265.
Walton ME, Behrens TE, Buckley MJ, Rudebeck PH, Rushworth MF (2010) Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. Neuron 65:927-939.
Wang XJ (2002) Probabilistic decision making by slow reverberation in cortical circuits. Neuron 36:955-968.
Wang XJ (2012) Neural dynamics and circuit mechanisms of decision-making. Curr Opin Neurobiol 22:1039-1046.
Wei Z, Wang XJ, Wang DH (2012) From distributed resources to limited slots in multiple-item working memory: a spiking network model with normalization. J Neurosci 32:11228-11240.
Wittmann M, Kolling N, Akaishi R, Chau BK, Brown JW, Nelissen N, Rushworth M (submitted) Predicting how your luck will change: decision making driven by reward memories with multiple time constants in anterior cingulate cortex
Woolrich M (2008) Robust group analysis using outlier inference. Neuroimage 41:286-301.
Woolrich MW, Behrens TE, Beckmann CF, Jenkinson M, Smith SM (2004) Multilevel linear modelling for FMRI group analysis using Bayesian inference. Neuroimage 21:1732-1747.
Worsley KJ, Evans AC, Marrett S, Neelin P (1992) A three-dimensional statistical analysis for CBF activation studies in human brain. J Cereb Blood Flow Metab 12:900-918.
Wunderlich K, Dayan P, Dolan RJ (2012) Mapping value based planning and extensively trained choice in the human brain. Nat Neurosci 15:786-791.