Modelling human decision under risk and uncertainty

Laurence Hunt

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the University of Oxford

Wadham College **Trinity Term 2011**

Table of Contents

Modelling human decision under risk and uncertainty

Laurence T. Hunt

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the University of Oxford.

Wadham College, Oxford **Trinity Term 2011**

Abstract

Humans are unique in their ability to flexibly and rapidly adapt their behaviour and select courses of action that lead to future reward. Several 'component processes' must be implemented by the human brain in order to facilitate this behaviour. This thesis examines two such components; (i) the neural substrates supporting action selection during value-guided choice using magnetoencephalography (MEG), and (ii) learning the value of environmental stimuli and other people's actions using functional magnetic resonance imaging (fMRI). In both situations, it is helpful to formally model the underlying component process, as this generates predictions of trial-to-trial variability in the signal from a brain region involved in its implementation.

In the case of value-guided action selection, a biophysically realistic implementation of a drift diffusion model is used. Using this model, it is predicted that there are specific times and frequency bands at which correlates of value are seen. Firstly, there are correlates of the overall value of the two presented options, and secondly the difference in value between the options. Both correlates should be observed in the local field potential, which is closely related to the signal measured using MEG. Importantly, the *content* of these predictions is quite distinct from the *function* of the model circuit, which is to transform inputs relating to the value of each option into a categorical decision.

In the case of social learning, the same reinforcement learning model is used to track both the value of two stimuli that the subject can choose between, and the advice of a confederate who is playing alongside them. As the confederate advice is actually delivered by a computer, it is possible to keep prediction error and learning rate terms for stimuli and advice orthogonal to one another, and so look for neural correlates of both social and non-social learning in the same fMRI data. Correlates of intentional inference are found in a network of brain regions previously implicated in social cognition, notably the dorsomedial prefrontal cortex, the right temporoparietal junction, and the anterior cingulate gyrus.

This work was funded by a 4-year studentship from the Wellcome Trust.

Acknowledgements

It is hard to know where to begin when thanking someone like Tim Behrens.

This thesis would not have been possible without the support of many scientists who have been kind with their time, advice and code. They did most of the real work. Matthew Rushworth's lab meetings and wisdom have been invaluable throughout. I often wish I had been a better listener to what he had to say from the start of my DPhil, but I hope he can forgive me and I will try and listen harder. Mark Woolrich made MEG work with his beamforming code, and also gave the project a good firm shake just when it needed it. Vladimir Litvak was immensely supportive of my stumbling attempts to provide code for SPM, and then taught me how to use the code once I'd written it. Gareth Barnes' MEG meetings at the FIL convinced me that there might (somewhere) be something interesting worth studying in the data. Alireza Soltani provided us with the code for the model in Chapter 5, showed us that it was making some interesting predictions, and even feigned interest when we added one additional prediction on top of this. Nils Kolling and George Wallis both showed wisdom beyond their years. Thomas Klausberger showed great kindness, and Peter Somogyi taught me much about how to try to become a scientist, during a difficult first term. Dave, Duncan, Marilyn, Sue, Paul and everyone who keeps FMRIB going, Sven, Anling and everyone who's getting OHBA going, Steve Knight at OCMR, you all do a remarkable job – thank you.

Much more important than neuroscience is porridge and winning games of squash. Maryann Noonan and Jerome Sallet were heroic providers on each front, a sign of true friendship if ever there was one. Ali Drewitt taught me much about vegetables, which I have since forgotten. Ilan Cooper manages to live in DC and still be on the end of the phone whenever I need him. There are too many others to thank, but I should mention (in the order that they pop into my head) Gerhard, Jesper, Max, Natalie Barker, Jill, Charlie Stagg, Tamar, Erie, Stam, Saad, Nick, Mark Walton, Mike Israel, Kathryn, Alex Savio, the FMRIB five-a-side football crew and all team Rushworth.

I thank all my family – Mum, Dad, Adam, Harriet, Roger, Alison, Will, Kate, Blacka, Mosey – whose love and support keep me going. It has been a tough few years but we're getting there slowly, I think (excepting Mo, who has already left the scene…).

Pip knows how much this work depends upon her efforts, recently quipping that she should really be joint first author of the thesis. It is a true shame that no mechanism exists in Oxford for this, but if I were ever to become senior enough to sit on a committee somewhere, I hereby promise to push for this progressive measure for you.

Tim, as ever, gets the final say. He will ho-hum, jolly good and tum te tum now he realises this is the case (assuming he got this far and didn't start a consolatory game of iSnooker after the opening sentence). He is a brilliant, creative scientist, has been an outstanding mentor throughout my DPhil, and I honestly don't think I would be pursuing a career in science without him. He may not believe the last part of that sentence, but it is true, and I draw on the analogy that teenagers often start to grow angry with their parents shortly before they leave home, and leave it at that.

Long abstract

Chapter 1 introduces the concept of goal-directed decision making, adopts a *component process* account of this behaviour, and distinguishes it from other behaviours that are not goal-directed. I introduce some of the brain regions that have been implicated in goal-directed choice, drawing upon evidence in monkey, man and rat. A widespread network of brain regions shows correlates of value during decision, and signatures of learning during reward feedback; this includes orbitofrontal and anterior cingulate cortex, portions of striatum and parietal cortex. A notable point is the diversity of signals recorded, and the diversity of brain regions from which they can be detected. This brings up the question of whether there is large redundancy in neural coding of values, and which regions of the brain are fundamentally involved in the *comparison* of values. It is also apparent that learning mechanisms for reward may be applicable in a more general setting than previously thought.

Chapter 2 introduces some modeling that might allow us to isolate brain regions fundamental to *value comparison*. These models originate from a 'drift diffusion' model, which has been widely used in sensory decision paradigms, but implement the diffusion process in a biophysically plausible manner, with nonlinear attractor dynamics. In this thesis, these models are used to make predictions of what a brain region performing value comparison might look like, and where in the brain matches such a signature. I also introduce *reinforcement learning* models, which are used to track the value of an action or a particular stimulus as it varies through time. In this thesis, these models are used to track the intention of another individual during a social interaction, and make predictions of neural activity from a brain region involved in this process of intentional inference.

Chapter 3 focuses on methodological developments, particularly in the field of magnetoencephalography (MEG), that have allowed for investigation of the physiological correlates of decision processes in healthy human subjects. I address theoretical and practical problems encountered in acquiring MEG data, the current state of the art in terms of MEG sensor technology, and solutions to the problems faced during source reconstruction. I also address key developments of particular importance for this thesis. I also briefly discuss functional MRI, which is used in a later study. Here, there is a more mature consensus upon acquisition and analysis techniques, and relatively standard techniques were adopted in this thesis.

Chapter 4 introduces the two experimental paradigms that are used in the subsequent MEG and fMRI experiments respectively. Both experiments involve choosing between two options for monetary reward. It is important to accurately model the

subjective value of different options, as neural activity is typically found to reflect subjective, rather than objective, value during choice. In one paradigm, *Prospect theory* is found to fit behaviour well, even though there is some previous evidence that this might be unexpected when repeatedly sampling the same options. The reasons underlying this controversy are explored. In a second paradigm, a social interaction is found to be accurately modeled using a reinforcement learning model which is used to infer the *intention* of a social partner during a monetary gambling task. This is one of the first applications of such a model to intentional inference, and suggests that such models may generalise to many forms of inference, in different frames of reference.

Chapter 5 presents results suggesting that a biophysical model that has previously been used to make predictions of single-unit data in perceptual choice can be extended to make predictions of magnetoencephalography (MEG) data in value-guided choice. These predictions relate to the specific time windows and frequency bands in which certain correlates of value will be seen. The predictions are then tested in MEG data recorded during value-guided choice. Many regions are found to show value correlates, but only two cortical subregions (superior parietal and ventromedial prefrontal cortex) match well with the model predictions. This suggests that models previously used for perceptual choice can be extended into the value-guided domain, and can also be used to discriminate the functional role of different cortical regions into different component processes of decision making.

Chapter 6 examines neural activity during intentional inference at two separate phases of the decision process – when one is receiving *feedback* about the behaviour of another individual, and when one is making decisions based upon *advice* from this confederate. This socially-derived information is combined with reward-based information in order to guide behaviour. At the time of making the decision, information from both social and non-social sources is combined in the ventromedial pre-frontal cortex, in order to guide future choices. However, when learning about social and non-social information, similar computations are carried out in distinct neural substrates. Specifically, the prediction error on reward information is found in ventral striatum, whilst the prediction error on social information is found in the theory of mind network, encompassing dorsomedial pre-frontal cortex and the temporoparietal junction. The learning rate on reward and social information dissociates into sulcal and gyral portions, respectively, of the anterior cingulate cortex.

Chapter 7 draws together the different streams of research encompassed in this thesis, and highlights two general principles that of importance for future studies. Firstly, a distinction can be drawn between what is 'represented' in terms of the *content* of a neural

signal and what its functional role might be, which is important when considering neuroimaging (and single unit) data recorded during value-guided choice. Secondly, in social cognition, several different 'frames of reference' might be adopted when analyzing the data, and this might distinguish the role of different brain regions during a social interaction.

Chapter 1: Introduction

This thesis addresses mechanisms used by the human brain to support value-guided action selection. A component process account of decision making is adopted, allowing us to test the mechanistic role of different brain structures in separable components of choice. Much has already been established concerning the functional neuroanatomy of value-guided choice; a network of cortical and sub-cortical structures is recruited during valuation of different options, and also during the learning of these values via reinforcement. These structures include orbitofrontal, anterior cingulate and parietal cortex, subcortical structures in the striatum and amygdala, and dopaminergic projections to widespread regions of the forebrain. In this chapter, I introduce the component process account of decision making, review the functional neuroanatomy of value-guided choice, and highlight several open questions addressed in the thesis.

> Let us weigh the gain and the loss in wagering that God is. Let us estimate these two chances. If you gain, you gain all; if you lose, you lose nothing. Wager, then, without hesitation that He is. "That is very fine. Yes, I must wager; but I may perhaps wager too much." Let us see. Since there is an equal risk of gain and of loss, if you had only to gain two lives, instead of one, you might still wager. But if there were three lives to gain, you would have to play (since you are under the necessity of playing), and you would be imprudent, when you are forced to play, not to chance your life to gain three at a game where there is an equal risk of loss and gain.
>
> **Blaise Pascal, 1669,** Pensées

Deciding how best to act can determine more than just one's chances of eternal salvation in the afterlife; it can also determine our likelihood of survival in the present life – and with it, our chances of reproductive success. The human brain can be considered nature's ultimate extension of this principle, allowing us to act in a way that is both adapted and adaptive to the demands of our environment, and to outcompete other species and individuals with inferior action planning and selection capabilities. Thus, selection of the most appropriate action is not only *a* reason to have a brain, but *the* reason to have a brain: witness the vast abundance of organisms with no need to act adaptively and rapidly in response to their environment, and the vast absence of brains in such organisms.

This thesis is concerned with the mechanisms used by the human brain when deciding how to act. Central to the mechanisms underlying decision-making will be the idea that the human brain might learn, compute and compare the *subjective values* of different actions. We will investigate ways in which these values can be estimated from human choices, candidate mechanisms by which they might be learnt and compared in the brain, and use neuroimaging data, collected as human subjects make value-guided

choices, to test these mechanistic hypotheses. The mechanistic explanations depend upon mathematical modeling that also serves a key role in analysing the neural activity that we record.

1.1 Value-guided choice: background and general framework

1.1.1 Some definitions and distinctions

We must begin with some definitions. A crucial concept in our investigation of value-guided choice will be that of *reward*. Reward is an operational term, which refers to something that the subject wants to obtain (Schultz, Dayan and Montague, 1997). It can include external stimuli, objects or money, the performance of a certain act, or an internal state. Typically, an organism will approach reward and work in order to obtain it. In machine learning, reward is a scalar quantity and the agent's *sole purpose* is to maximise the long-term reward that is obtained (Sutton and Barto, 1998). Biologically, it is not unreasonable to align reward with pleasure, although it should be borne in mind that the hedonic satisfaction associated with a stimulus might not always scale with the degree to which an animal will approach or work for that stimulus (Berridge, 1996). Reward is perhaps instead more closely aligned to the economic concept of *utility*. Whilst economists might originally have wanted a 'hedonimeter' with which such utilities could be measured (Colander, 2007), it is now typically assumed that they can instead be revealed from the 'outward phenomena to which they give rise' (Marshall, 1920)– namely, choice behaviour.

It is also important to distinguish between the different kinds of actions that an animal can take. It might be tempting to argue that every action ever taken is, in some sense, a 'decision' made by the animal. However, a wealth of evidence from psychology and behavioural neuroscience suggests that all actions are not created equal (Balleine, Daw and O'Doherty, 2008). Instead, actions can be subdivided into those that are *reflexive* or *habitual* – those elicited based only upon environmental stimuli – and those that are *goal-directed* – which take into account the reward contingent upon certain actions, and are sensitive to manipulations of the value of a reward. One of the classic tests to distinguish the two categories are that the latter, but not the former, are sensitive to *reinforcer devaluation* – namely, satiation of a specific reward or pairing with an aversive stimulus (Holland and Rescorla, 1975). Reinforcer devaluation should reduce the level of responding if the action is goal-directed, but not if the action is habitual. At times, the habitual and value-guided 'systems' (assuming that they are truly dissociable from one another) may come into conflict, and mechanisms might be needed for resolving this conflict (Daw, Niv and Dayan, 2005).

A further distinction that should be drawn is between goal-directed and *Pavlovian* responding. As will be discussed further in **chapter 2**, Ivan Pavlov studied animal's responses to 'unconditioned stimuli' (normally rewarding), such as the presentation of food, and to other stimuli predictive of these rewarding stimuli. The responses made by the animal to the stimuli did not affect the probability that the animals would obtain the food. One such response is a dog salivating, a response readily elicited when food reward is presented. Is salivation a goal-directed action? It is not. The key distinction is that although the action is prompted by presentation of the food, the animal cannot alter this response *in order to obtain* the food. This can be shown in an experiment in which the animal does not receive any food if it salivates on that particular trial (Sheffield, 1965). The animal continues to salivate after several hundred trials of such training – it cannot withhold salivation in order to get more food. Pavlovian, habitual and goal-directed actions are thus largely thought of as separate from one another, with only goal-directed actions satisfying the criteria necessary for a '*bona fide* decision.' However, the degree to which they are truly separated from one another in the brain is still a matter of ongoing debate.

In this thesis, we will primarily be concerned with actions that we consider to be goal-directed – a term we will use interchangeably with 'value-guided', 'value-based' or 'reward-guided.' The underpinnings of value-guided choice lie not only in psychology and cognitive neuroscience, but also in economics, which has been concerned with predicting the future behaviour of consumers in the decisions that they make, and in computer science, which has developed algorithms that allow for the learning and comparison of rewarding actions. It also, at times, draws upon principles from statistics and behavioural ecology.

1.1.2 What computations might a goal-directed agent need?

We can think of goal-directed choice as being broken down into several component processes that are necessary and sufficient for its proper execution. One useful scheme for this component process account is presented in figure 1. Whilst the exact extent to which these processes are separable in the brain is unclear (there may be an overlap, for instance, between action selection and which actions are valued), it is nonetheless useful when considering which part of the framework is being tested in each of the experiments that we design.

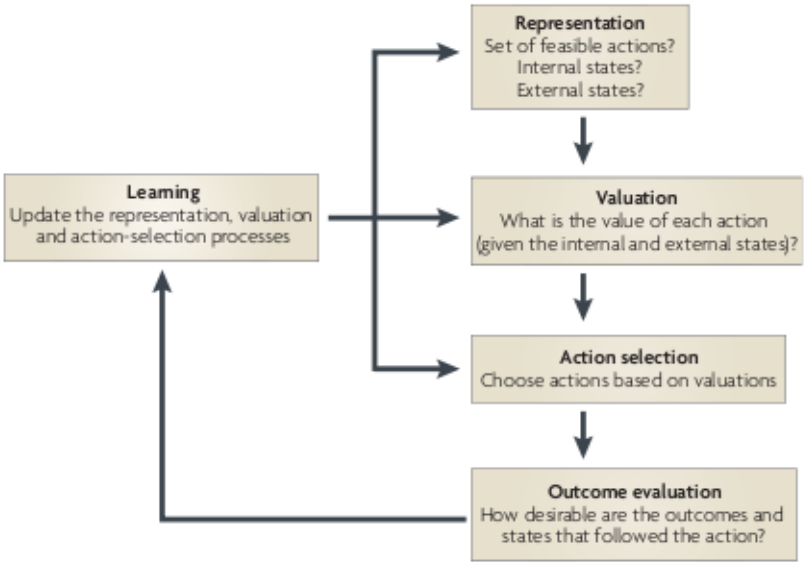

Figure 1. A component process account of goal-directed decision making. 5 component processes are identified: (i) representation of appropriate actions; (ii) valuation of these actions; (iii) selection of the most appropriate action; (iv) evaluation of outcomes; (v) learning based upon differences between evaluated outcomes and initial valuation. This thesis addresses stages (ii), (iii) and (v), but does not address the representation of available outcomes or outcome evaluation. Adapted from (Rangel, Camerer and Montague, 2008).

1.1.2.1 Valuation

The value of an action typically consists of an estimate of the long-term reward that will be obtained from taking that action. This estimate may be influenced by several considerations. First, it may incorporate knowledge of the likely outcomes that are contingent upon certain actions, and the reward value associated with those outcomes. Second, it might incorporate knowledge of likely 'state transitions' – the probability of reaching a certain state (in which other actions can be taken) as a consequence of taking a particular action – equivalent to knowing something about the *structure* of the environment. A classic example of such knowledge would be the latent learning exhibited by rats trained by Tolman, who rapidly found their way to a reward in a maze after they had previously been given a chance to explore the maze's layout in the absence of reward (Tolman, 1948). Third, it may incorporate information about the value of different stimuli that may be encountered, that are not themselves rewarding but have been predictive of reward in the past. A key idea here is that incommensurable stimuli will need to be translated into a form of *common currency* in which stimulus values can be compared (O'Doherty, 2004, FitzGerald, Seymour and Dolan, 2009). Finally, it may be the case that cognitive processing, including the use of propositional logic and conceptual knowledge, will help to shape the expected value of a given course of action (Rangel, Camerer and Montague, 2008, Kumaran et al., 2009).

Our understanding of the learning and estimation of values can often be helped by using a mathematical model to try to capture the *subjective values* used to guide choice behaviour. These models may, for instance, try to capture cognitive biases used by subjects when using probabilistic information or information about rewards. Similarly, they may try to capture the effects of delay on reward, eliciting so-called 'temporal discounting'. In **chapter 2**, we will examine in more detail some of the mathematical models that have been developed to capture these processes. Whilst the models have already been established as describing choice *behaviour* well in both human and animal subjects, recently researchers have begun to investigate which *brain regions* support valuation, and whether activity in these regions can also be captured by the same mathematical models used to describe behaviour.

1.1.2.2 Action selection

Whilst quite a considerable amount is known about the computations and neural substrates supporting *valuation* of different choices, far less is known about the processes underlying value-guided *action selection* in the brain. For instance, it is unclear in which 'decision space' value-guided action selection needs to take place. Should we first compute the value associated with each action, and then choose amongst the different action values? Or should we instead form a representation of the behavioural goals, choose which goal is most desirable, and select an action plan that leads to that goal (Padoa-Schioppa, 2010)? Might we even use attention to bias which actions are considered more valuable during the course of a decision (Krajbich, Armel and Rangel, 2010)? These questions are only beginning to be explored.

It is equally unclear what *selection mechanism* is used by the brain to select between different options. One idea, which we will explore in some detail in this thesis, is that value-guided action selection may depend upon integrative mechanisms similar to those needed for decision-making in a perceptual discrimination task. A popular mechanism in such tasks is one in which evidence is compared via *competing accumulators* racing towards a *decision threshold*, at which point a decision is made. In **chapter 2** we will explore the mathematical formulations of such a mechanism, and how these might be implemented by a neural circuit.

1.1.2.3 Learning

Of all the computational processes associated with goal-directed choice, learning is perhaps the best understood. A key idea common to all forms of learning, including reflexive, Pavlovian and value-guided, is that *surprising* events are critical to driving

changes in future expectations. The degree of surprise will depend upon forming some kind of *prediction* at each point in time, and how much the currently observed outcome deviates from this prediction. This idea has been formalised in numerous mathematical models and used to successfully explain neural activity; again we will examine these more closely in **chapter 2**.

A further important consideration is the degree to which estimates should be updated by surprising events – that is, the *value of new information* (Behrens et al., 2007, Rushworth and Behrens, 2008). In many cases, it will also be unclear which stimulus or action led to a particular reward, and so the role of *credit assignment* will become important to guide successful learning. The neural substrates supporting these last two computations is somewhat less clear, although is beginning to be elucidated. Finally, it is also unclear whether similar mechanistic processes used for learning about the reward structure of the environment might be used for learning more complex structure in the environment, such as making predictions about the behaviour of social partners.

1.1.3 Summary

We have briefly considered what is meant by a value-guided choice, and that it can be described in terms of underlying component processes. It has been mentioned that each of the three components of valuation, decision making and learning has been influenced by adopting a mathematical formalism to describe behaviour and neural activity. We will consider these in detail later, but it is first important to review the functional neuroanatomy supporting these component processes.

1.2 The functional neuroanatomy of value-guided choice

The past fifteen years has seen a rapid expansion in our understanding of the neural substrates that support value-guided choice. Investigators have used techniques such as single-unit recording and human neuroimaging to study the *correlates* of value in the intact brain. This has been combined with interference techniques including lesion studies, transcranial magnetic stimulation, and microstimulation to investigate the *causal* role of brain circuits in decision-making. Results from both correlative and causal approaches support the idea that a distributed network of brain regions signal metrics related to reward prediction, reward-guided learning, and action selection. Some of the most important regions of the human brain are shown in figure 2. In this section, an overview is provided of the functional roles of these different brain regions during value-guided choice and learning by reinforcement.

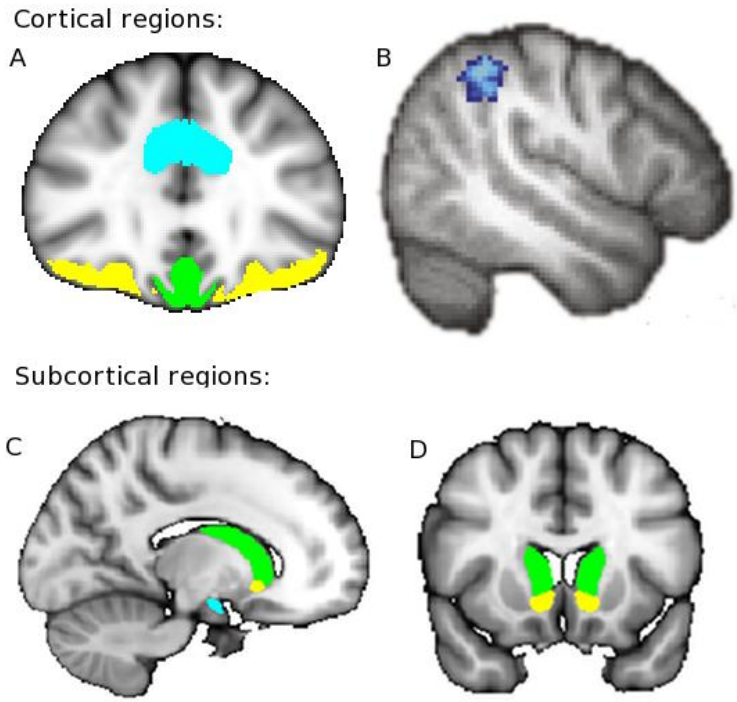

Figure 2. Some regions of the human brain implicated in reward-guided decision and action. See text for details. A/B: Cortical regions. (A) Anterior cingulate sulcus (light blue); ventromedial prefrontal cortex (green); lateral orbitofrontal cortex (yellow). MNI Y=31mm. (B) Mid-intraparietal sulcus; close to likely human homologue of intraparietal neurons recorded during reward-guided action selection in monkeys. Adapted from (Boorman et al., 2009). MNI X=50mm. C/D: Subcortical regions. (C) Dorsal striatum (green); nucleus accumbens (yellow); amygdala (blue). MNI X=-14mm. (D) Dorsal striatum (green); nucleus accumbens (yellow). MNI Y=14mm.

1.2.1 Choices and value in the brain

The precursor to reward-guided action selection is the computation of value. Several brain regions appear to encode relevant variables, but notably there is often heterogeneity, either across studies or within a single study, with respect to precisely what metric is encoded in a given brain region. It is unclear whether this heterogeneity is a reflection of the need to encode many different decision variables within a single brain region, or whether it is an emergent property of an underlying process such as value comparison.

1.2.1.1 Lateral intraparietal cortex and correlates of economic value

The lateral intraparietal cortex (LIP) of the macaque monkey contains neurons that contribute to the generation of saccadic eye movements. LIP neurons typically have a 'preferred direction', and their firing rate increases prior to saccades made in this direction (Gnadt and Andersen, 1988). In the late 1990s, a debate raged about the precise role of these neurons in cognitive processing. One school of thought tied LIP neural activity to *stimulus* processing, arguing that activity reflected the amount of

attention allocated to a particular location in space (Gottlieb, Kusunoki and Goldberg, 1998). An alternative school argued that LIP more closely reflected *motor* preparation, coding for the *intention* to move to a spatial location (Snyder, Batista and Andersen, 1997). With cleverly designed experiments, each school was able to provide evidence in support of its own philosophy of LIP function, and against that of the alternative.

It was in this context that Michael Platt and Paul Glimcher proposed that LIP activity might instead code for a variable that was not so closely tied to stimulus or response processing, but was instead some form of intermediary between the two – a *decision variable* (Platt and Glimcher, 1999). A few pioneering studies had examined neural activity in relation to reward expectancy independent of stimulus or response properties before, e.g. (Watanabe, 1996). For the first time, however, Platt and Glimcher related the firing rates of LIP neurons to variables more commonly used in economic descriptions of value – namely, the probability and magnitude of reward received from making a particular saccade. By varying across blocks the amount of juice, or the probability of receiving juice, associated with a saccade made towards a particular spatial location, Platt and Glimcher showed a clear linear correlation between the *expected value* of the planned saccade and the firing rates of LIP neurons (figure 3). Whilst this analysis did not entirely resolve the intention/attention debate (Andersen and Buneo, 2002, Bisley and Goldberg, 2003), it nevertheless was the first to test the concept of economic value being coded in the brain (Shizgal, 1997), which was to inspire a whole field of subsequent research. A complementary literature had previously related LIP activity to the representation of a decision variable, but in the domain of perceptual (rather than value-guided) decision making (Shadlen and Newsome, 1996). This perceptual literature was closely tied to the mathematical modelling of decision processes, and so is reviewed in more detail in **chapter 2**.

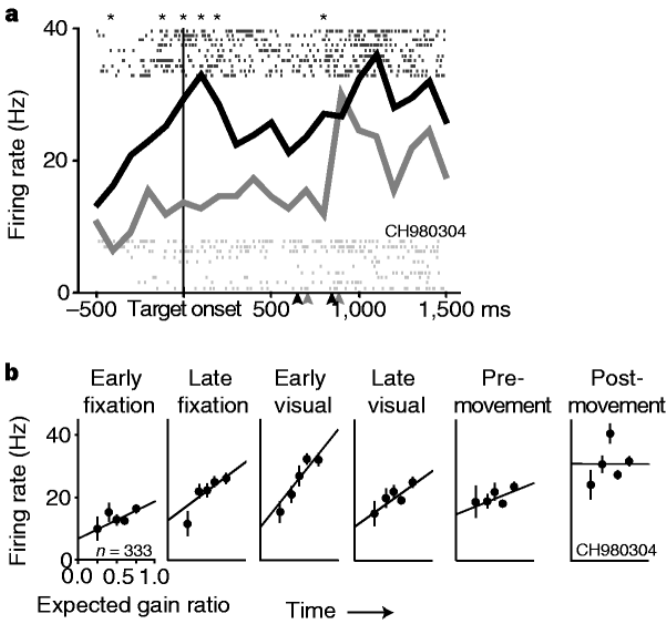

Figure 3. Firing rate of LIP neurons covaries with expected gain of juice reward during free choice. (A) Raster plot of firing rates during high expected gain ratio (juice delivered on 0.75 of trials; dark raster/lines) and low expected gain ratio (juice delivered on 0.25 of trials; light raster/lines). (B) Firing rates of cell during different phases of the experiment leading up to saccade execution.

A host of subsequent single-unit electrophysiology papers built on the findings of Platt and Glimcher, investigating the coding of subjective values in LIP neural firing rates. Sugrue and colleagues used a task in which rewards were probabilistically allocated to one of two options, and the options then became 'baited' – the rewards remained hidden behind the options until the monkey chose that option. Animal and human behaviour in such a task tracks a 'matching law', in which they distribute their responses in proportion to the frequency of reward being delivered on each option (Herrnstein, 1961). LIP activity tracked the local probability of each option being rewarded (Sugrue, Corrado and Newsome, 2004). A similar local tracking of reward probability by LIP neurons was also found in a study by Seo and colleagues, who used a reinforcement learning model (see **chapter 2**) to track the probability of reward in a 'matching pennies' task (Seo, Barraclough and Lee, 2009). Interestingly, Seo *et al*. found that as many neurons represented the summed value of both options ('overall value') as represented the difference in value between chosen and unchosen options – one of the first hints of heterogeneity in the encoding of reward expectations during choice. Glimcher's subsequent studies demonstrated that neural activity in LIP tracks subjective, rather than objective, value more faithfully (Dorris and Glimcher, 2004), and that activity evolves from initially representing subjective value to subsequently representing the probability of saccading towards that option (Louie and Glimcher, 2010).

Surprisingly, in spite of the widespread neurophysiological data supporting the role of LIP in coding the value of particular actions, only a handful of human functional MRI studies have found evidence of value coding in the intraparietal sulcus (IPS), the location of the likely human homologue of LIP (and of the nearby equivalent for arm movements, the parietal reach region). These fMRI studies have tended to find that BOLD signal in the IPS correlates *negatively* with the value of the chosen option (Boorman et al., 2009, Gershman, Pesaran and Daw, 2009), and is particularly active on trials where the subject switches which action is chosen (Boorman et al., 2009, Glascher, Hampton and O'Doherty, 2009). It also shows some degree of lateralisation for actions made by the contralateral hand (Gershman, Pesaran and Daw, 2009). The negative correlation with value is somewhat surprising in light of the single unit studies. However, it should be remembered that the fMRI signal will contain a mixture of selective and non-selective neuronal populations, rather than selectively focussing on responses towards a particular receptive field. Thus, a trial in which there is more conflict between possible actions – because the chosen value is lower – might paradoxically lead to an increase in gross activity, as more action representations are recruited during the choice. IPS BOLD fMRI signal readily scales with the degree of response uncertainty in a non-value guided decision-making task (Huettel, Song and McCarthy, 2005), consistent with there being greater signal on trials where multiple action representations are simultaneously recruited.

Similarly, the effect of lesions to the human intraparietal sulcus on value-guided choice remains relatively unexplored. This is perhaps because parietal cortex damage is far more commonly associated with the condition of hemispatial neglect, in which a unilateral portion of visual space receives reduced attention (Bisiach and Luzzatti, 1978). Whilst it might be tempting to interpret neglect as a deficit in valuation of a certain location in space, lesions causing neglect tend to be centred on the more ventral and lateral angular gyrus, rather than the IPS (Husain and Rorden, 2003).

1.2.1.2 Cingulate cortex and dorsal striatum also code for action values

Several other regions of the brain also appear to carry information about the expected reward associated with making a particular action when a decision is being made. Two regions that have received particular attention are the striatum, in particular the dorsomedial part, and the anterior cingulate cortex, in particular its sulcal portion (ACCs) (Rushworth et al., 2004, Balleine, Delgado and Hikosaka, 2007). In the rat, lesions to the dorsomedial striatum produce insensitivity to reinforcer devaluation on instrumental value-guided choice tasks (Yin et al., 2005). A similar effect can also be

achieved with lesions to the rat pre-limbic cortex (Corbit and Balleine, 2003, Killcross and Coutureau, 2003); this projects strongly to dorsomedial striatum, and its dorsal portion bears some similarities to the macaque ACCs in its anatomical connectivity (Kunishio and Haber, 1994, Rushworth et al., 2004). In the ACCs of the macaque monkey, both lesions and injection of the $GABA_A$ receptor agonist muscimol cause a deficit in appropriate selection of reward-guided actions (Shima and Tanji, 1998, Hadland et al., 2003). By contrast, lesions to the dorsolateral striatum and infralimbic cortex do not produce deficits in goal-directed behaviour, but instead produce a disruption of the formation of habitual behaviours (Killcross and Coutureau, 2003, Yin, Knowlton and Balleine, 2004).

These effects of causal manipulations of dorsomedial striatum and ACCs on goal-directed action selection are corroborated by single-unit recording studies in both structures. Early work demonstrated that ACCs neurons increased their firing rates as reward expectancy increased in a task requiring several serial responses to obtain reward, and this was associated with a decrease in the frequency of erroneous actions as reward approached (Shidara and Richmond, 2002). A key finding was that in a paradigm in which one of two cues instructed a go/no-go response for possible reward, ACC neurons would primarily code for the *response*, the possibility of reward, or the interaction of these two factors (Matsumoto, Suzuki and Tanaka, 2003). By contrast, very few ACC neurons coded for *stimulus* identity or the interaction of stimulus with reward. However, this pattern was reversed in dorsolateral PFC, where many neurons coded stimulus-reward but not action-reward contingencies (Matsumoto, Suzuki and Tanaka, 2003). Subsequent direct manipulations of reward probability and magnitude have shown that ACCs neuronal firing rates correlate with the expected value of reward on action based decision tasks (Amiez, Joseph and Procyk, 2006, Kennerley et al., 2009). ACCs neurons sensitive to reward magnitude also fire selectively for *specific* actions (in this case saccade directions), suggesting that ACCs 'multiplexes' information about action and reward during value-guided choice (Hayden and Platt, 2010).

A similar multiplexing of information about specific actions and predicted reward can also be found in the caudate nucleus of the striatum. The caudate contains neurons selective to particular saccade directions, neurons sensitive to reward expectation, and some neurons that reflect the interaction of these two factors (Kawagoe, Takikawa and Hikosaka, 1998). By explicitly manipulating the reward probability associated with different actions, Samejima and colleagues demonstrated that striatal neurons signaled the probability of a particular action (in this case a joystick movement) being rewarded (Samejima et al., 2005). These findings closely

linked striatal neuronal activity to the signaling of action values, as has been demonstrated previously for LIP. Interestingly, in both the study of Samejima and a subsequent study, relatively few caudate neurons were found to code for the value of the *chosen* action (Cai, Kim and Lee, 2011). This is in contrast to a study of striatal neurons by Lau and Glimcher, which found a heterogeneous response, with approximately equal numbers of neurons sensitive to action values and chosen values (Lau and Glimcher, 2008).

In human imaging studies, investigations of regions encoding action values may again be hampered by the mixed selectivity of neurons encoding different actions in the same brain region. One possible approach to circumvent this issue may be to use separate effectors, with different selective brain regions, for different alternatives (Wunderlich, Rangel and O'Doherty, 2009). Nevertheless, striatal and cingulate activity shows patterns of activity that is suggestive of the encoding of action-outcome contingencies. Activity in human ACC increases when multiple response plans come into conflict (Botvinick et al., 1999, Botvinick, Cohen and Carter, 2004). As in the IPS, this might be interpreted as the representation of multiple action-outcome contingencies being simultaneously activated, rather than implying a role in conflict detection *per se* (Rushworth et al., 2007b). The ACC interestingly does not, unlike the IPS, scale linearly with response uncertainty during non value-guided choice (Huettel, Song and McCarthy, 2005). It does, however, show a negative correlate of the chosen value during reward-guided choice (Boorman, personal communication). Notably, in an action-based task in which only *one* speeded response is required, but the reward magnitude and probability associated with fast responding is varied, both ACCs and striatal BOLD fMRI signal scale *positively* with the expected value of the action (Knutson et al., 2005).

1.2.1.3 Stimulus values: orbitofrontal cortex and amygdala

A complementary literature has focused on responses to stimuli, rather than actions, that are predictive of future reward. Early recording studies of single unit activity in orbitofrontal cortex (OFC) suggested that neuronal activity would signal the reward outcome associated with presentation of a specific stimulus. If, during a reversal learning paradigm, the reward value of a stimulus changed due to the occurrence of a reversal, the firing rates of the neurons selective for that stimulus also typically reversed, even though the stimulus presented was the same (Thorpe, Rolls and Maddison, 1983, Rolls et al., 1996). In these early recording studies, it was unclear whether the neurons were truly responding to stimulus-reward associations or to reward delivery, as the two were coincident with one another. Subsequent studies

avoided this confound by placing a delay between these two events, and showed that OFC single unit activity reflected the reward predictive properties of stimuli prior to delivery of reward (Schoenbaum, Chiba and Gallagher, 1998, Tremblay and Schultz, 1999, Hikosaka and Watanabe, 2000, Roesch and Olson, 2004). OFC neuronal activity has also been recently shown to linearly reflect the subjective value of different options in an economic choice task (Padoa-Schioppa and Assad, 2006). Here, its activity is notably more closely tied to stimulus values rather than the value of a specific action, but again these responses are heterogeneous, with some neurons responding to stimulus 'offer values', some to the 'chosen value', and some to just the anticipated juice taste. Finally, OFC firing rates are also sensitive to the devaluation of reinforcers associated with particular stimuli (Rolls, Sienkiewicz and Yaxley, 1989). Reward predictive cells associated with specific stimuli can also be found in basolateral amygdala (Schoenbaum, Chiba and Gallagher, 1998), an area with reciprocal connections to OFC. Amygdala neurons code for both stimulus identity and reward value at the time of stimulus presentation, and then primarily code for value during the delay when the stimulus is removed (Paton et al., 2006).

These results suggest that both OFC and amygdala play an important role in encoding stimulus-reward contingencies during value-guided choice. Lesions to both structures render monkeys and rats less sensitive to reinforcer devaluation following stimulus-outcome learning (Gallagher, McMahan and Schoenbaum, 1999, Izquierdo, Suda and Murray, 2004, Izquierdo and Murray, 2007), and crossed disconnection lesions between OFC and amygdala are sufficient to produce similar deficits (Baxter et al., 2000). Notably, these lesions can also affect the normal deployment of transitivity between different food rewards in monkeys (Baylis and Gaffan, 1991) and humans (Fellows and Farah, 2007). Often the effects in these lesion studies are complemented by findings in deficits of reward learning; these are reviewed later in this chapter.

The effects of OFC lesions on reinforcer devaluation are notably restricted to stimulus-outcome, but not action-outcome, contingencies (Ostlund and Balleine, 2007). Here, an important contrast can be drawn between the roles of orbitofrontal and anterior cingulate portions of prefrontal cortex. Lesions to the former affect the learning (and deployment) of probabilistic stimulus-reward contingencies, whereas lesions to the latter affect action-reward contingencies (Rudebeck et al., 2008). This is in close agreement with the distribution of stimulus- and action-selective cells in each of the two structures.

1.2.1.4 Medial and lateral orbitofrontal cortex show distinct patterns of connectivity and reward sensitivity

The dissociation between orbitofrontal cortex and anterior cingulate cortex in carrying information about stimuli and actions is perhaps reflective of the distinct patterns of anatomical connectivity exhibited by these two regions (figure 4). The ACC has dense reciprocal connections with motor cortex (Morecraft and Van Hoesen, 1992) and spinal cord (Dum and Strick, 1991), suggestive of a critical role in guiding action selection. By contrast, the OFC (in particular its lateral subdivision) receives highly processed sensory information from neurons in visual, auditory, and other sensory cortices (Carmichael and Price, 1995), but has little direct connectivity to regions associated with motor output. Importantly, a further subdivision can be drawn between medial and lateral portions of the OFC. These two subdivisions form separable networks with strong within-network connectivity, but relatively sparse between-network connectivity (figure 4; (Carmichael and Price, 1996, Price, 2007)). The medial subdivision also receives relatively little in the way of processed sensory input (Carmichael and Price, 1995), but has strong outputs to visceral control structures in hypothalamus and midbrain (An et al., 1998, Ongür, An and Price, 1998), as well as some connectivity to cingulate cortex (Carmichael and Price, 1995). Similar patterns of connectivity can also be found using diffusion weighted imaging in humans (Croxson et al., 2005).

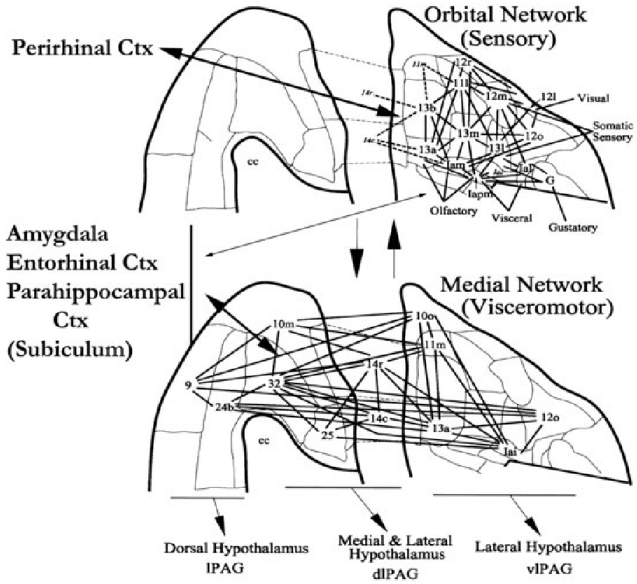

Figure 4. Ventral prefrontal cortex subdivided on the basis of internal and external connections. The ventral prefrontal cortex can be subdivided into an orbital network which receives processed sensory information, and a medial network which is more strongly connected to visceral control structures. Both networks have stronger within-network than between-network connectivity. Adapted from (Price, 2007).

Interestingly, correlates of value in fMRI experiments of goal-directed choice tend to be focused on the medial wall, in ventromedial prefrontal cortex (VMPFC) (Hampton, Bossaerts and O'Doherty, 2006, Kable and Glimcher, 2007, Plassmann, O'Doherty and Rangel, 2007, Behrens et al., 2008, Boorman et al., 2009, FitzGerald, Seymour and Dolan, 2009). This is part of the medial network identified by Price and colleagues (Price, 2007, Rushworth and Behrens, 2008). VMPFC is activated irrespective of whether the decision is over actions or stimuli (Glascher, Hampton and O'Doherty, 2009). The nature of the signal observed in VMPFC shows heterogeneity between studies; in some fMRI studies it has been found to signal a difference between chosen and unchosen values (Serences, 2008, Boorman et al., 2009), whilst in others it has appeared to signal the overall value of available reward (Blair et al., 2006), or the value of just the chosen option (Kable and Glimcher, 2007). However, VMPFC has been the subject of relatively few published single-unit recording studies thus far. Instead, the studies reviewed above in section 1.2.1.3 are all taken from the more lateral portion of OFC, which falls within Price's lateral connectional network.

A direct comparison of single unit activity in the two regions was recently performed by Bouret and colleagues, who found that lateral OFC neurons more frequently coded the reward magnitude associated with a *particular stimulus* during a cued response task, whereas more VMPFC neurons coded the reward magnitude available during *self-initiated, uncued, action* (Bouret and Richmond, 2010). Single unit activity in VMPFC rarely codes for stimulus value during the stimulus-based economic choice task used previously to study OFC once monkeys have been overtrained (Padoa-Schioppa, unpublished observations) but may code for value in a task in which multiple dimensions have to be considered whilst the monkey is still learning the task (Hayden, unpublished observations). One interpretation of these data is that whereas lateral OFC may directly code (and guide decisions over) stimulus-reward contingencies, VMPFC may frame decisions in the context of current behavioural goals.

The proposed distinction is consistent with a recent double dissociation identified between the roles of lateral and medial portions of OFC in value-guided choice. Whereas lesions to the lateral OFC affected the precise stimulus to which credit for reward delivery was assigned (Walton et al., 2010), lesions to the medial OFC affected selection of the most appropriate behavioural goal (Noonan et al., 2010). The role of VMPFC in selecting over behavioural goals is also supported by lesion studies in human patients; here, although VMPFC-lesioned patients can effectively choose between options with multiple attributes, their process of deciding is markedly different from healthy controls (Fellows, 2006). VMPFC-lesioned subjects switch to a strategy of

accumulating information about each alternative separately, rather than comparing alternatives across attributes. In the absence of a VMPFC, subjects might be less able to compare options by selecting over internal goals, but will still be able to make a decision by slowly constructing an action value for each available option.

1.2.1.5. Summary

A distributed network of cortical and subcortical structures is implicated in signalling reward expectancies and guiding valuation during goal-directed choice. Distinctions can be drawn between regions that appear to code value-related activity more in the frame of reference of actions, such as parietal/cingulate cortices and the dorsal striatum, and regions that are more in the frame of reference of stimuli, such as orbitofrontal cortex and amygdala. Stimulus value-related modulations in activity can also be found in other, less 'cognitive' brain regions, including primary visual cortex (Shuler and Bear, 2006, Serences, 2008), but it is unclear whether neurons in these regions code for reward expectancies *per se* or for the allocation of attention coincident with reward expectation. Some brain regions, such as ventromedial prefrontal cortex, may not be so closely tied to coding in the frame of reference of stimuli or actions, but may instead frame decisions over current behavioural goals.

The *mechanisms* by which decisions are made on the basis of reward expectancies remain highly unclear. One key finding that may be of relevance is that, within the same brain region, different studies appear to emphasise different aspects of value-related signalling. Some neurons (or BOLD fMRI signals) appear to code for stimulus or action values, others appear to code for chosen values, and yet others code for overall ('state') values or the difference in value between chosen and unchosen options. As discussed above, this heterogeneity has been shown in parietal cortex, orbitofrontal cortex and ventromedial prefrontal cortex, and so may be a general principle of activity in any region of the brain implicated in value-guided choice.

How might we explain the diversity of different value-related signals observed? It may be important to place the results into a mechanistic framework in which action selection occurs, that could simultaneously explain the presence of all of these different signals in neural activity. In **chapter 2**, we explore mathematical frameworks that might be appropriate for value-guided choice behaviour, with the intention of ultimately testing whether these diverse representations might emerge from within these frameworks.

1.2.2 Reward outcomes and value learning in the brain

In many situations, the value of different courses of action must be estimated via trial-and-error learning. The birth of theories of associative learning can be traced back to the work of scientists at the turn of the 20th century, notably Edward Thorndike and Ivan Pavlov, who studied the means by which animals learn predictions of rewarding stimuli (Thorndike, 1911, Pavlov, 1927). In his 'law of effect', Thorndike proposed that learning consists of the formation of associations between particular stimuli and responses when those responses elicited rewarding outcomes. In a simple demonstration, he found that after placing a hungry cat in a 'puzzle box', the cat would try many different actions until he eventually pressed a lever that released him from the box. During repeated trials, the cat would gradually become faster to press the lever and escape the box each time. Thus, he argued, the reward of escaping had reinforced the association between the stimulus of the box and the response of pressing the lever. This process was later termed 'operant' (or stimulus-response) conditioning, and can be contrasted with the 'classical' (or stimulus-reinforcer) conditioning studied by Pavlov, in which an affectively neutral stimulus could acquire value via repeated pairings with a positive or negative reinforcer. The classic example first studied by Pavlov was the acquisition of positive value by a bell (the 'conditioned stimulus', or CS) when closely followed by delivery of food (the 'unconditioned stimulus', or US). Prior to pairing, the US alone would cause the dog to salivate, but the CS would elicit no behavioural response. After multiple CS-US presentations, however, the CS would begin to elicit a response of salivation, suggesting that the animal had learnt an association between its presentation and the delivery of food. Both Thorndike and Pavlov used these examples as demonstrations against previous explanations of learning, which claimed animals could learn the puzzle box via 'insight', or would begin to salivate via 'psychic' expectation of the delivery of food. The processes underlying classical and operant conditioning (and situations in which Pavlov's and Thorndike's theories failed to explain certain phenomena) formed the basis of much behaviourist research during the course of the 20th century (Watson, 1913, Tolman, 1948, Ferster and Skinner, 1957, Konorski, 1967, Pearce and Bouton, 2001). These behaviourist studies were gradually complemented by investigations of the neural substrates that supported reward-guided learning.

1.2.2.1 Dopamine acts as a reward signal that reinforces stimuli and actions preceding its release

A key insight into the neural substrates of associative learning came from *intracranial self-stimulation* studies (Olds and Milner, 1954). These found that electrical stimulation delivered to specific brain regions could serve as a reinforcing stimulus, in many cases a more powerful one than naturally occurring rewards such as food. One area identified as being a particularly effective stimulation site was the medial forebrain bundle (MFB), which contains dopaminergic axons projecting from the ventral tegmentum to regions of the forebrain, including striatum and prefrontal cortex. It was found that 6-hydroxydopamine-induced lesions of the dopaminergic system and dopamine receptor antagonists both reduce the reinforcing properties of MFB intracranial self-stimulation (Fouriezos and Wise, 1976, Fibiger et al., 1987), and antagonists also reduced responding when specifically injected into the nucleus accumbens (Mogenson et al., 1979). Amphetamine, which increases synaptic availability of dopamine, augmented reinforcement associated with self-stimulation (Colle and Wise, 1988) and could itself act as a reinforcer (Beninger and Hahn, 1983). Together, these findings led to the proposal that dopamine release, in particular to nucleus accumbens, subserved the reinforcing properties of MFB stimulation, and so that dopamine might also play an important role in signaling rewards occurring naturally in the environment (Wise and Rompre, 1989).

This hypothesis was supported by studies that investigated the effects of manipulations of the dopamine system on learning guided by natural rewards. Pioneering studies showed that delivery of dopaminergic antagonists caused rats' responses for food rewards to be altered (in addition to the Parkinsonian locomotor deficits associated with such drugs). After a period of instrumental conditioning off-drug, injection of the dopaminergic antagonist pimozide caused rats to behave similarly to control rats placed in extinction; this was true even though food reward delivery for the drugged rats remained contingent on lever pressing, suggesting that dopamine was needed to mediate the rewarding effects of food delivery (Wise et al., 1978). This finding was extended to emphasise a distinction between dopaminergic release to the dorsal striatum, which when impaired primarily affected sensorimotor learning, and to the ventral striatum, which affected responses to primary rewards and the learning of stimulus-reward associations (reviewed in (White, 1989, Robbins and Everitt, 1992)). Similar deficits in reward-guided associative learning can also be found in Parkinsonian patients (Knowlton, Mangels and Squire, 1996, Frank, Seeberger and O'Reilly R, 2004, Rutledge et al., 2009).

Direct evidence for the role of dopamine release in the signaling of rewards was provided using single unit recordings of presumed dopaminergic neurons in the monkey ventral tegmental area. These neurons showed robust responses when the monkey touched morsels of food that were hidden from view, but not to touching rewards that the monkey had already seen (Romo and Schultz, 1990). It was later shown that dopaminergic responses to rewards depended crucially on their unpredictability (Apicella et al., 1992), a finding that could later be replicated in humans using functional MRI and PET imaging techniques (Berns et al., 2001, Zald et al., 2004). This led to the hypothesis that dopamine release reflected not reward *per se*, but *errors in prediction of reward* (Schultz, Dayan and Montague, 1997, Schultz, 1998). This hypothesis allowed activity to be explained in the light of mathematically formal theories of learning developed in behaviourist psychology and machine learning; the basis of these theories is discussed in **chapter 2**.

1.2.2.2 Lesion studies reveal cortical substrates of reward-guided learning

Dopaminergic neurons are distinguished by their projections to striatum, with axonal ramification such that each individual dopamine neuron has half a million striatal release sites (Schultz, 1998). This is complemented by a widespread projection to cortex, which targets areas that are closely tied to motor output and cognition, but has relatively sparse innvervation of early sensory areas (Lewis et al., 1987). Focussing specifically on projections to frontal cortex, arborisation in supplementary motor area, area 6 and ACC is particularly dense, and there is weaker (but nonetheless dense) projection to structures such as ventromedial frontal and orbitofrontal cortex (figure 5; (Williams and Goldman-Rakic, 1993)).

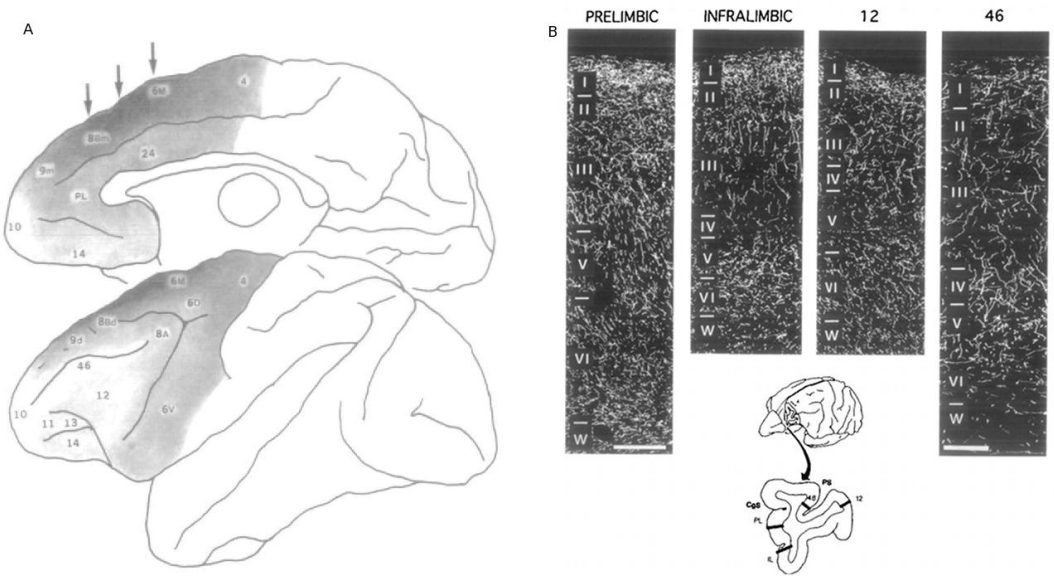

Figure 5. Dopaminergic innervation of prefrontal cortex, investigated using a monoclonal antibody to dopamine (adapted from Williams and Goldman-Rakic, 1993). (A) Schematic diagram showing density of dopaminergic innervation of frontal cortex. Arrows point towards areas of peak innervation. (B) A single section of prefrontal cortex (anterior to the genu of the corpus callosum) reveals that even in areas such as area 12 (ventro-lateral PFC), there is strong dopaminergic innervation, albeit weaker than areas on the medial surface (such as prelimbic cortex).

It is clear that mesolimbic projections (to nucleus accumbens) and nigrostriatal projections are important for instrumental conditioning (Robbins and Everitt, 1992, Salamone and Correa, 2002). However, it can also be seen that the dopaminergic predictive reward signal will influence activity in many of the cortical structures discussed in section 1.2.1 as coding for action and stimulus values, and so may be in a position to influence synaptic plasticity in these regions too. Two of the structures that have been implicated in value learning with reference to their connections to the dopaminergic system are anterior cingulate cortex (Holroyd and Coles, 2002, Brown and Braver, 2005) and orbitofrontal cortex (Schoenbaum et al., 2009). Direct manipulations of dopaminergic input to these structures on reward-guided learning have only been studied in a few cases (e.g. Walton et al., 2005), and so there is limited evidence for or against the role of mesocortical dopamine in value learning. However, both of these structures have been strongly implicated in associative learning, irrespective of whether this depends upon dopaminergic input, based upon evidence from lesion studies, neuroimaging and single unit electrophysiology.

Damage to OFC has been documented in several high-profile case studies as preserving normal cognitive, sensory and motor function, but causing a loss of motivation, reflected in patients' inability to run their everyday lives (Harlow, 1848, Eslinger and Damasio, 1985, Damasio, 1994). Such motivational deficits might be explained in light of patients' inability to assign value correctly, or to learn new values

appropriately. Monkeys with circumscribed lesions to orbitofrontal cortex are impaired on flexibly reassigning stimulus-reward contingencies during visual object discrimination learning, or 'reversal learning' (Mishkin, 1964, Jones and Mishkin, 1972, Dias, Robbins and Roberts, 1996, Meunier, Bachevalier and Mishkin, 1997, Izquierdo, Suda and Murray, 2004). OFC reversal deficits are specific to learning stimulus-reward, but not action-reward, contingencies (in contrast to the ACC (Rudebeck et al., 2008)). Similar perseveration deficits in reversal learning tasks can also be found in human subjects with OFC damage (Rolls et al., 1994, Bechara et al., 1997, Fellows and Farah, 2003, 2005).

The role of the OFC in reversal learning have traditionally been proposed as either processing negative reinforcement following unrewarded choices in the task, causing an inability to learn from errors in its absence (Kringelbach and Rolls, 2004), or suppressing a previously rewarded response, causing a tendency to perseverate in its absence (Elliott, Dolan and Frith, 2000). However, recent data have highlighted that OFC-lesioned monkeys are in fact equally sensitive as controls to negative feedback in the absence of reversal (Walton et al., 2010). Following a reversal, they also tend to switch actions even more frequently than controls (rather than 'perseverate'), in spite of being rewarded for some of their choices on the alternate options (Rudebeck and Murray, 2008, Walton et al., 2010). It has been pointed out that their deficits are instead better described as an impairment in *credit assignment* - correctly assigning reward to the appropriate stimulus choice that caused it. The credit assignment hypothesis of OFC function contends that reversal deficits arise because when the OFC-lesioned subject switches to choosing the rewarded option following a reversal, rewards on this option are inappropriately assigned to the unrewarded alternative that has been chosen on recent trials. Evidence for this hypothesis can be found in the choice behaviour of OFC-lesioned monkeys performing a modified reversal task, by running a sensitive logistic regression analysis in which the conjunction of reward and recent choice history is considered (Walton et al., 2010). In light of such a view of OFC function, we might expect a representation of the specific stimulus identity on a given trial to be held in OFC at the time of feedback, and for this to be combined with information concerning errors in prediction of reward. Evidence from anatomical studies (reviewed above) and electrophysiology (reviewed below) suggests that this is the case.

In contrast with OFC, the effect of lesions to anterior cingulate cortex have received relatively less attention, perhaps because their effects on classic tests of working memory and object reversal learning are relatively slight (Meunier, Bachevalier and Mishkin, 1997, Rushworth et al., 2003). However, these too have recently been

found to affect reward-guided learning, selectively when it is with respect to which *action* is more likely to yield reward, rather than which object or stimulus (Shima and Tanji, 1998, Hadland et al., 2003, Kennerley et al., 2006, Rudebeck et al., 2008). In one study, monkeys were trained that either turning or lifting a joystick would yield a food reward; the rewarded action would remain constant for 25 trials, and reward would then unexpectedly reverse to the alternative action. ACC-lesioned monkeys persistently performed worse than controls on the task in terms of error rates. However, their deficit was more specific than this. A logistic regression analysis was used to determine the influence of recent outcomes on future choice behaviour. This analysis revealed that control monkeys would integrate over the outcomes of several recent trials to determine their future behaviour, consistent with a process of learning via reinforcement (figure 6). By contrast, ACC-lesioned monkeys were influenced by only the most recent trial outcome, suggesting that they were unable to learn action values using a reinforcement learning strategy (and might instead have depended upon a less stable strategy, such as using working memory for the outcome and choice of the most recent trial) (figure 6; (Kennerley et al., 2006)).

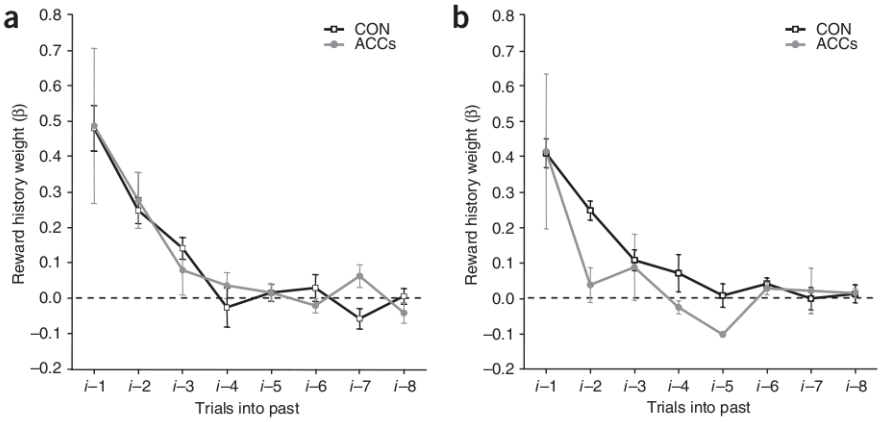

Figure 6. Influence of lesions to anterior cingulate sulcus (ACCs) on weight attributed to recent outcomes on previous trials, estimated using logistic regression. (A) Pre-lesion behaviour. (B) Post-lesion behaviour, for lesioned (ACCs) and control (CON) monkeys. Notice that post-lesion, ACCs lesioned monkeys do not integrate over multiple trials in a reinforcement-learning style fashion, and instead attribute weight only to the most recent outcome.

1.2.2.3 Electrophysiological and functional imaging studies of reward learning in ACC and OFC

Whilst findings from lesion studies are strongly indicative of a role for OFC and ACC in reward-guided learning, it is always possible that the observed deficits may arise from impairment in appropriate value-guided choice, rather than learning *per se*; it may be that a separate neural circuit is involved in the learning of these values, and that the role of OFC and ACC lies in implementing these cached values. It is often difficult to tease

these two hypotheses apart in lesion data, but in electrophysiological and functional imaging studies this task becomes more straightforward – as neural responses to decision and reward can be *temporally* separated from one another. Evidence from both these modalities strongly implicates OFC and ACC in the learning of stimulus and action values on the basis of reward feedback.

There are a large number of reward-coding neurons that can be found in OFC at the time of feedback (Rosenkilde, Bauer and Fuster, 1981) in addition to cells encoding information about choices made earlier in the trial (Tsujimoto, Genovesio and Wise, 2009). This conjunction of reward and choice information is precisely what would be required of a region performing contingent learning of reward associations. The evidence for coding of *errors* in prediction of reward at the time of feedback is mixed (Kennerley et al., 2009, Takahashi et al., 2009, Sul et al., 2010), perhaps because relatively few studies have used optimal experimental designs to tease apart reward and prediction error activity (Sul et al., 2010). In any case, prediction error signals will likely be available to the OFC via dopaminergic projections (either directly or via striatum). Human OFC is also found to respond to the delivery of food, liquid, or monetary reward (Elliott, Friston and Dolan, 2000, O'Doherty et al., 2000, O'Doherty et al., 2001, Gottfried, O'Doherty and Dolan, 2003, Rolls, McCabe and Redoute, 2008, Grabenhorst et al., 2010), and OFC reward responses appear to be modulated by their predictability, consistent with the presence of a prediction error signal in OFC (Berns et al., 2001, O'Doherty et al., 2003). The importance of dopaminergic input to OFC may explain value learning deficits seen when the VTA and OFC are disconnected in the rat (Takahashi et al., 2009). (However, this has also been interpreted as suggesting a role for OFC in signaling outcome expectancies to VTA, to allow the *computation* of prediction errors in VTA (Schoenbaum et al., 2009)).

A direct test of the role of the OFC in contingent learning was recently performed by Brodersen and colleagues, who used fMRI to compare a condition in which learning could be performed contingently with other conditions where it had to be performed non-contingently (Brodersen et al., in prep.). In the contingent condition, rewards were delivered with respect to the choice made on the current trial, as is normally the case in reinforcement learning or reversal tasks. In the non-contingent conditions, however, reward was delivered either with respect to a choice made one or two trials back in time, or the recent average of choices made by the subject. Reward contingency could be further probed by observing which outcomes effected a change in subjects' behaviour on the subsequent trial, and which outcomes did not. The results indicated that lateral OFC (and ventral striatum) were recruited on trials where reward learning was

contingent with the most recent choice, but not on trials where reward learning was non-contingent.

Although the OFC is recruited in response to rewarding outcomes, it is also the case that it can be activated by punishment or omission of reward (O'Doherty et al., 2001). A meta-analysis of fMRI studies has suggested that a lateral portion of OFC tends to be recruited in contrasts looking for signals reflecting punishment, and a more medial portion tend to be recruited when looking for reward signals (Kringelbach and Rolls, 2004). A similar claim has traditionally been made of the anterior cingulate cortex, which shows an event-related potential that is sensitive to whether an error is made during the execution of a movement (Gehring et al., 1995, Debener et al., 2005) or when punishing feedback is delivered in a gambling task (Miltner, Braun and Coles, 1997). This ERP is sometimes referred to as the 'error related negativity'.

It is often the case, however, that errors in such paradigms are far more infrequent than correct responses, and so are more informative for guiding future behaviour. Walton and colleagues devised a paradigm in which reward and error feedback were *equally* informative for determining which of several stimulus-response contingencies was currently to be used (Walton, Devlin and Rushworth, 2004). Lateral OFC and ACC were found to respond when new information was received about the appropriate contingency, irrespective of whether it was based on a reward or an error (figure 7). However, the OFC activity was only seen when feedback was framed in the context of information about a *stimulus* selected by the experimenter, whereas the ACC activity was restricted to trials in which information was with respect to an *action freely chosen by the subject*, in line with the hypothesis that OFC and ACC are more concerned with learning information about stimuli and actions, respectively. Subsequent investigations demonstrated that the error-related negativity is also equally sensitive to reward and punishment when these two are equated in terms of the influence they have on behaviour (Oliveira, McDonald and Goodman, 2007). These results are thus in line with the hypothesis that OFC and ACC are critically implemented in the learning of stimulus and action values, irrespective of whether this learning is driven by reward or punishment. The response in these regions will be scaled by the degree to which these values should be updated at the time of feedback delivery.

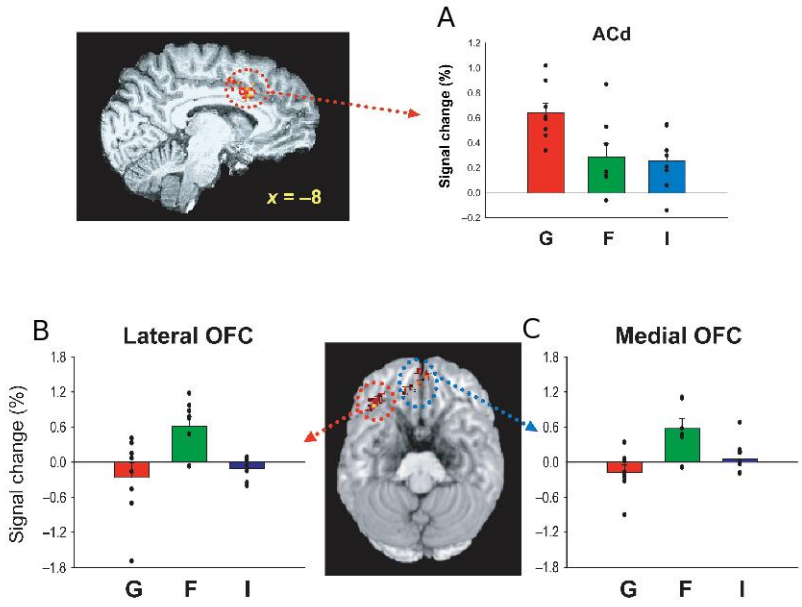

Figure 7. Responses of cingulate and orbitofrontal cortex during outcome monitoring is dependent upon whether stimulus-response learning is guided by internally generated actions (G), actions fixed by the experimenter (F), or a set-switching control condition (I). (A) Dorsal anterior cingulate cortex is most active on trials where the subject observes the outcome of an internally generated action. (B)/(C) Lateral OFC and medial OFC are most active on trials where the subject observes the outcome of an action tied to a particular cue, determined by the experimenter. Adapted from (Walton, Devlin and Rushworth, 2004).

The role of the ACC in the reward-guided adjustment of behaviour is corroborated by evidence from single-unit recording. Early studies showed that cells in the cingulate motor area (the most rostral portion of ACCs) responded to reward when it signalled a change in the action that was required to obtain future rewards (Shima and Tanji, 1998). In addition, error-related cells, whose activity is modulated by the degree of expected reward on a trial, were also found to be modulated by an external signal instructing the monkey to switch actions (Amiez, Joseph and Procyk, 2005). These signals are remniscent of a prediction error, in that they encode stimuli (either rewards or instructions) that drive future changes in behaviour. Indeed, direct evidence for prediction error encoding in ACC was provided by Matsumoto and colleagues, who showed that as monkeys progressed through a series of trials in which the value of an action was learnt, ACC neurons encoded the prediction error on action values (Matsumoto et al., 2007). Intriguingly, in this and two other studies, the distribution of reward- and error-related units was approximately equal (Ito et al., 2003, Seo and Lee, 2007), suggesting a heterogeneity of responses in ACC. This is in contrast to dopamine which is traditionally thought of as uniformly encoding rewarding stimuli positively

(Ungless, Magill and Bolam, 2004) (but see Matsumoto and Hikosaka, 2009). Recent evidence has shown that ACC neurons can encode counterfactual prediction errors on unchosen options (Hayden, Pearson and Platt, 2009) and also unsigned prediction errors signaling the degree to which behaviour should be updated in future (Hayden et al., 2011).

1.2.2.4 Summary

Converging evidence from causal manipulations (lesions, pharmacology) and physiological recordings (fMRI, electrophysiology) has implicated orbitofrontal and anterior cingulate cortex as being critical not only in signaling reward expectations during choice, but also in learning reward values via reinforcement. This role is likely stongly influenced by dopaminergic input that reflects reward prediction errors, relayed either directly via mesocortical projections, or indirectly via the striatum. It appears that OFC is more closely related to the learning of stimulus values, and ACC is more closely related to the learning of action values. It is important to bear in mind that these may not be the only differences between the two structures; for instance, ACC may also be important in generating exploratory actions or in integrating the costs and benefits associated with taking a particular action (Rushworth et al., 2007a). In any case, our understanding of the role of both dopaminergic and cortical activity in reward learning is greatly enhanced by the use of formal models describing strategies with which the value of stimuli or actions can be learnt; we investigate these further in **chapter 2**.

1.3 Summary

Adopting a component process account of value-guided decision making allows us to consider the roles of neural structures in the initial valuation of stimuli or actions, the means by which these are compared, and the way in which they are learnt via reinforcement. Much has been discovered about the functional neuroanatomy of value-guided choice, and many cortical and subcortical structures are implicated in initial valuation and reinforcement learning. Nevertheless, there remain many open questions.

One of the starkest is with regards to *mechanism*: how is it that after valuation, options are compared and an action is selected? Are there appropriate, physiologically realistic mechanisms by which this comparison might take place? If so, which of the numerous brain structures encoding values also supports the comparison of these values? Might such mechanisms explain the heterogeneity of signals that are observed in such structures?

A second question is how values learnt using traditional reinforcement learning strategies are *integrated with other, more cognitive information*, such as that derived from social interactions. Is social information unique and distinct from reinforcement learning in terms of the mechanisms underlying its acquisition, or might similar strategies be used for both social and non-social information? Does this depend upon the same brain regions discussed in this chapter, or an entirely separate set of brain regions? How is social information combined with non-social information at the time of choice?

Both of these questions appeal to adopting a mechanistic, more formal account of the neural activity described in this chapter. Fortunately, formal quantitative models have recently become very popular (and successful) in describing neural activity during value-guided decision making. In the next chapter, we examine some of these models.

Chapter 2: Computational models of decision making and their neural correlates

A major aim of this thesis is to infer the computational processes performed by distinct cortical regions during value-guided choice. This inference is made possible through the use of mathematical modeling, to make precise predictions of the neural data that might be observed in a region supporting a particular process. This chapter reviews some key computational principles of value-guided choice, and what has already been discovered about their neural correlates.

A 'model' is something that attempts to describe the *essence* of some feature of the world. It reduces it to the components that are crucial to understanding this feature, whilst eschewing unnecessary detail. For instance, consider the number of times a fair coin lands heads in a series of one hundred throws. One approach to modeling this problem is to use Newtonian mechanics to fully derive the expected trajectories of the coin, and use this (with the distribution over possible starting positions and velocities) to derive whether it is likely to lands heads up on each throw. A more parsimonious model, however, ignores these details and uses a binomial distribution to derive the probability distribution for the expected number of heads across all throws. This simpler model retains the essence of the outcome of tossing a coin, whilst removing features of the world that are superfluous to its description.

With complex phenomena such as human cognition and brain function, parsimonious models become essential if we are to make any progress towards their understanding. These models can be, and have traditionally been, somewhat informal – such as the subdivision of working memory into separate component processes (Baddeley, 1992), or the hierarchical model of visual processing in cerebral cortex (Felleman and Van Essen, 1991). In many cases, however, psychological and neural processes can lend themselves to being modeled more formally, using mathematical models that provide quantitative predictions about measureable data. In the psychological literature, for example, such quantitative models have been used to predict the likelihood of correct responses and false alarms during signal detection (Green and Swets, 1966). In brain research, quantitative models have been used at many levels of description, from the generation of the action potential (Hodgkin and Huxley, 1952) to the dynamic interactions of multiple brain regions (Friston, Harrison and Penny, 2003).

Quantitative models are also the cornerstone of 'model-based' analyses of neural data collected during cognitive tasks. This approach has become particularly popular in the analysis of brain imaging data collected to study decision making and learning processes (O'Doherty, Hampton and Kim, 2007, Behrens, Hunt and Rushworth, 2009). Model-based analysis frequently proceeds by using a model that makes predictions about the expected subject behaviour (choices, reaction times) at each trial. The model may be constrained by adjusting free parameters to fit the observed behavioural data more accurately. Trial-by-trial fluctuations in internal features of the model are then used as regressors in the analysis of neural data collected during task performance (Figure 1). It is argued that brain regions correlating with these regressors are performing a computational role analogous to the function of these internal features of the model (O'Doherty, Hampton and Kim, 2007). Such an analysis bears the advantage that we gain some insight into the *implementation* of a cognitive process in a given brain region, as opposed to merely showing that it is involved in this process. It also provides a means of retaining strong experimental control over low-level perceptual and motor task demands whilst still testing the hypothesis of interest. Finally, by performing an experiment in which several internal parameters are simultaneously varied, it is possible to functionally segregate different brain regions as performing separate computational roles.

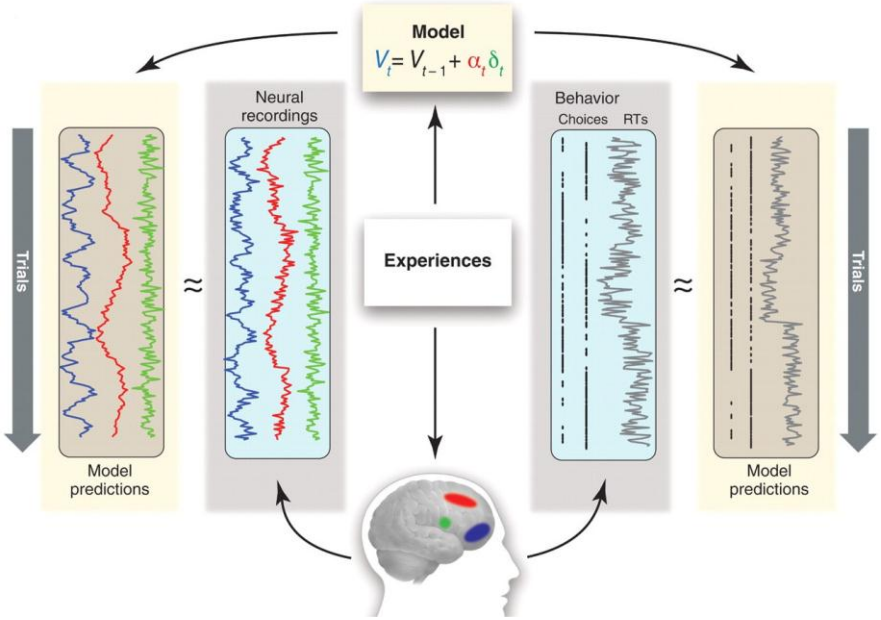

Figure 1. Schematic of an approach that combines mathematical models of behaviour with neural recordings. The model contains parameters that represent specific computations underlying behaviour. As the subject/model undergoes different experiences, these parameters will fluctuate. The fluctuation in these parameters is used to find neural correlates of the specific underlying computations. Separately, the same parameter fluctuations come together to predict changes in behavior. (adapted from Behrens, Hunt and Rushworth, 2009)

Quantitative models are useful in capturing several stages of value-based decision making. These include the initial valuation of different objects, the comparison of these objects to form a categorical decision, and the update of learnt values based on feedback after a decision has been made. The multidisciplinary nature of decision science is reflected in the broad spectrum of fields from which these quantitative models are drawn; they encompass psychology (Ratcliff, 1978, Bogacz et al., 2006), machine learning (Sutton and Barto, 1998), statistics (Wald, 1947), economics (Tversky and Kahneman, 1992) and neurobiology (Wang, 2002). In this chapter I will discuss some of the key models that have been used in describing the formation of a decision and learning by reinforcement, and how these models have been applied to the study of the brain.

2.1 Models of valuation

A defining feature of a value-guided choice is that it depends upon a quantity that is subjective and internal to the organism making the choice – namely the *subjective value* of the available options. As this subjective value is unobservable, it is necessary to devise ways of estimating it. This allows predictions to be made about how decision-makers should behave in future decisions. A large branch of economic theory is devoted to precisely this topic. Several observations suggest that human behaviour is irrational in that it deviates from optimal or 'normative' choice behaviour. Nonetheless, these irrationalities can be themselves captured in formal models. The deviations are typically also reflected in neural data collected during subjective valuation.

2.1.1 A brief history of value

A correspondence between Blaise Pascal and Pierre de Fermat in 1654 is often traced as being the root of modern theories of valuation. In it, Pascal proposed that when faced with different options of different reward magnitude and probability, a decision-maker should simply multiply magnitude by probability to obtain the *expected value* (now frequently referred to as the *Pascalian value*) of each option. The decision-maker should then select the option with the highest expected value.

Soon afterwards, Daniel Bernoulli highlighted a straightforward violation of this rule in human behaviour (Bernoulli, 1738). He gave the example of a lottery ticket with equal probability of 0 ducat gain or 20000 ducat gain – that is, a Pascalian value of 10000 ducats. He argued a poor man would happily sell this ticket at 9000 ducats, whereas a rich man would happily buy the ticket for 9000 ducats. The subjective value

of the ticket is thus different for the two men. The key ideas that explain this behaviour are that the values are *reference dependent* – that is, they depend upon the starting position of the decision-maker – and that they also show *diminishing marginal utility* – that is, the difference between 100 and 1100 ducats looms larger than the difference between 10000 and 11000. Diminishing marginal utility means that the expected utility function is *concave*, rather than linear, for increasing values.

In the early 20th century, economists began to combine these theories with *axiomatic* descriptions of choice behaviour (Samuelson, 1938, von Neumann and Morgenstern, 1944). The choice axioms reflected fundamental rules that should hold if subjects hold preferences between items, and those preferences remain stable. The key feature of such an axiomatic description is that starting from a very simple set of assumptions, economists could make predictions of consumer choices on decisions that subjects had not yet been presented with. The predictions depended upon ranking different options according to the *expected utility* associated with each option – typically the probability multiplied by Bernoulli's non-linear utility function – and then using the decision axioms to predict consumer behaviour. Such normative models are appealing in that they *generalise* to many decisions of different kinds.

Human behaviour, however, turns out not to be quite so neat. The Frenchman Mauice Allais became famous for approaching leading neoclassical economists at conferences, and asking them to make a series of choices in which they readily violated the axioms that they so strongly advocated. Allais' choice 'paradoxes' (Allais, 1953), along with other demonstrations of irrational choice behaviour e.g. (Ellsberg, 1961), became famous counter-examples for the expected utility theory of choice behaviour. The reaction to such counter-examples amongst economists was mixed. Some proposed that normative theories were still correct, but would not apply under 'exceptional circumstances' (Glimcher et al., 2008). The economists generally wanted to steer clear of a normative theory supplemented by a disarray of exceptions to the theory, however, as the exceptions largely depended upon the circumstances of the choice, and so did not *generalise* and were not *falsifiable*. Such concerns might be avoided, however, if it were possible to develop a theory that accounted for the behavioural anomalies whilst also being formally rigorous, and so could be falsified. Such a development lay in store in the 1980s, in the form of Prospect theory.

2.1.2 Prospect theory can explain observed deviations from normative behaviour

Working in the late 1970s and early 1980s, Daniel Kahneman and Amos Tversky developed a series of experiments which showed that the behavioural anomalies first

demonstrated by Allais and colleagues were much more widespread and pervasive than had previously been thought. They then developed a theoretical framework that encapsulated many of these anomalies, by extending expected utility theory in several directions (Kahneman and Tversky, 1979, Kahneman, Knetsch and Thaler, 1991, Tversky and Kahneman, 1992).

Firstly, they proposed that the value function was concave, as described by Bernoulli, but only for gains; for losses, it was steeper, and convex (figure 2). This explained two observations. It explained why human subjects tend to be risk-averse for gains, but risk-seeking for losses – that is, they would choose a certain option with a lower expected value than a risky win, but would risk a large loss to avoid a certain loss. It also explained why subjects tend to exhibit *loss aversion* – they require more compensation to give up an item than they do to purchase the same item. Importantly, these gains and losses are set relative to a *reference point*. For gamblers making a series of bets, for example, this reference point might be the *overall* winnings and losses for a given day (helping to explain why gamblers tend to chase riskier bets towards the end of a day (McGlothin, 1956)).

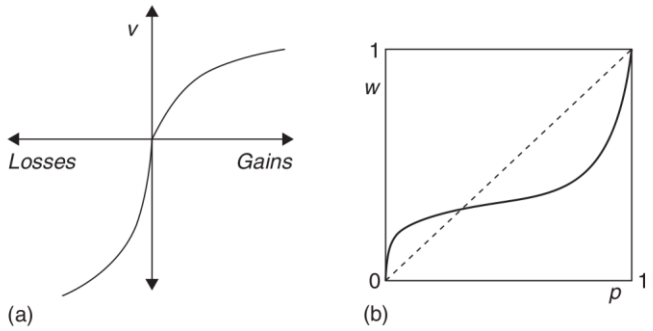

Figure 2. Prospect theory functions. (A) Value function, showing concave curvature for gains, convex curvature for losses, and steeper gradient for losses than gains. (B) S-shaped probability weighting function. Taken from Fox and Poldrack, 2008.

Secondly, they proposed that probability was not linearly related to the true probability, but instead showed a *non-linear* weighting function. The shape of that function is shown in figure 2. The S-shaped function means that subjects tend to *overweight* low probabilities, and *underweight* high probabilities, contributing to what Kahneman and Tversky referred to as a 'fourfold pattern' of risk attitudes (Kahneman and Tversky, 1979, Fox and Poldrack, 2008). The overweighting of low probabilities has often been invoked in explaining why people tend to buy lottery tickets and insurance premiums; although the Pascalian value of these items is less than their price, the overweighted probability of a large jackpot win or of your house burning down leads the decision-maker to buy these items (Camerer, 2000). Several parametric forms of the

non-linear weighting function exist, and there has been close examination of which of these forms best describes human behaviour (Gonzalez and Wu, 1999). However, recent evidence has suggested that the weighting function may be reversed when the probabilities are learnt from *experience* rather than from a *description* of the decision problem (Hertwig and Erev, 2009). This last point is discussed further in **chapter 4**.

Finally, Kahneman and Tversky proposed that human behaviour might be sensitive to framing and editing effects – that is, human behaviour violates *description invariance*. A simple set of rules was proposed for how certain decision problems might be framed and edited. These frequently have effects on the way probabilities are represented by the decision maker, and where the reference point is. For instance, subjects will choose differently in a choice between a sure gain of £20 or a 50-50 chance of gaining £50, than when they are given £50 and then asked to choose between a sure loss of £30 or a 50-50 chance of losing £50 – even though these two decisions are formally identical.

2.1.3 Value-related neural signals are typically subjective, rather than objective

Several recent studies have investigated how value signals in the brain are modulated during paradigms of decision under risk, and whether these modulations comply with the predictions of Prospect theory. Although there are some nuances to each of these studies, the broad picture is that value representations tend to closely match with the subjective values predicted by Prospect theory, rather than objective Pascalian values.

Tom and colleagues investigated BOLD fMRI signal in human subjects as they chose to reject or accept a mixed prospect with an equal probability of a gain or a loss (Tom et al., 2007). They varied the magnitude of wins and losses to identify in each subject the degree of loss aversion. The first important finding from this study is that losses and wins tended to affect the *same* network of brain structures in opposite directions, as opposed to one set of regions processing punishments and a separate network processing rewards (as has been proposed by other theories (Kringelbach and Rolls, 2004)). These regions included several of the brain regions discussed in **chapter 1**, such as ventromedial prefrontal cortex, striatum, and orbitofrontal cortex. Moreover, across subjects, the ratio of win/loss coding in these structures closely matched the win/loss ratio in choice behaviour – that is, behavioural loss aversion was closely correlated with 'neural' loss aversion. This suggests that activity in these structures is closely related to the *subjective* value of the decision.

Hsu and colleagues used a similar decision under risk paradigm in order to estimate the non-linearity of the probability weighting function in human volunteers (Hsu et al., 2009). They used fMRI to investigate the neural coding of the non-linear portion of the probability weighting function, and looked for whether it was encoded in the same regions that also correlated linearly with the presented probability. The representation of the non-linear component closely overlapped with the representation of the linear component. The overlap encompassed a distributed network of regions including a peak near the cingulate cortex, and several regions that receive dopaminergic input that the authors interpreted as striatal (but may have in fact had their peak in ventral anterior thalamus (Boorman and Sallet, 2009)). Irrespective of precisely which regions were recruited, the main conclusion is that neural activity generally tracked the subjective probability weighting function, rather than the true probability of reward.

Finally, De Martino and colleagues have shown that neural activity is sensitive to the way in which a decision is framed to subjects (De Martino et al., 2006). They asked participants to choose between a sure option and a risky gamble, but framed the decision either in terms of gaining money, or in terms of losing money that had already been given to the subject. Subject behaviour varied between the two conditions in accordance with framing predictions from Prospect theory, and neural activity in the amygdala was found to flip frames of reference depending upon how the decision was framed to the subjects.

So it can be seen that neural activity correlating with value closely tracks three key predictions of Prospect theory: the shape of the utility function (loss aversion), the curvature of the probability weighting function, and the effect of framing. Neural activity therefore closely matches with subjective, rather than objective, values during reward-guided choice.

2.1.4 Thesis work related to modeling of decision values

Most of the work in this thesis focuses on implementation of decision and learning processes, as discussed in the next two sections, rather than implementation of valuation processes. However, this section has highlighted the importance of accurately estimating subjective values from behaviour in paradigms of value-guided choice, in order to design regressors that accurately capture neural activity. This estimation of subjective values, and the design of appropriate experimental paradigms, is discussed in **chapter 4**.

2.2 Models of evidence accumulation and selection

A 'decision' is a process of considering alternatives in order to form a categorical commitment to one of these alternatives. The process of consideration typically requires some accumulation of evidence for each alternative. In a value-guided decision, this evidence relates to the internal value of each option to the decision maker, based on its associated costs, benefits and possible punishments. Several mathematical models can provide optimal accounts of evidence accumulation and decision formation (Bogacz et al., 2006)[1]. These models are closely related, and have each been used to describe neural data collected during decision tasks. Frequently, the tasks used have been in the perceptual domain, in which the evidence can be kept under tight experimental control. They have often been related to single-unit recordings in the saccade selective lateral intraparietal cortex (LIP), which, as discussed in **chapter 1**, also shows sensitivity to the value associated with making a particular saccade. The degree to which these models can also describe neural activity recorded during value-guided choice, and in brain regions other than LIP, is only beginning to be investigated (Lee and Wang, 2008).

2.2.1 A decision between competing hypotheses can be made using a likelihood ratio test

Consider the decision between two alternative hypotheses (H_1, H_2) to explain a series of independent observations $Y=y_{1\ldots n}$. If we know the conditional probability of each observation under each alternative a, we can write down the probability of the entire series Y as:

$$p(Y \mid H_a) = \prod_{i=1}^{n} p(y_i \mid H_a)$$

Equation 1

We can then compare the two hypotheses by calculating a *likelihood ratio* (LR) (Neyman and Pearson, 1933):

[1] Optimality is here defined as producing a decision of specified accuracy in the shortest possible time, or producing a decision of maximum possible accuracy in a specified period of time.

$$LR\left(\frac{H_1}{H_2}\right) = \frac{p(Y \mid H_1)}{p(Y \mid H_2)} = \prod_{i=1}^{n} \frac{p(y_i \mid H_1)}{p(y_i \mid H_2)}$$

Equation 2

If this ratio is greater than 1, the observations support H_1, and if it is less than 1, the observations support H_2. By taking the logarithm of both sides, we transform the product on the right-hand side of the equation into a sum, in which each piece of evidence is sometimes denoted as the 'weight of evidence' (WOE) in favour of H_1 over H_2:

$$\log LR\left(\frac{H_1}{H_2}\right) = \log\left(\frac{p(Y \mid H_1)}{p(Y \mid H_2)}\right) = \sum_{i=1}^{n} \log\left(\frac{p(y_i \mid H_1)}{p(y_i \mid H_2)}\right) = \sum_{i=1}^{n} WOE\left(\frac{H_1}{H_2}\right)$$

Equation 3

This logarithmic formulation linearises the accumulation of evidence. It also means that in cases where pieces of evidence are presented sequentially (i.e. *n* is increasing), and a decision must be made as to when to terminate evidence accumulation, we can simply add each new WOE to the logLR until sufficient evidence favours one hypothesis over the other. This is termed the 'sequential probability ratio test' (SPRT) (Wald, 1947). Accumulation of evidence continues until the logLR exceeds some fixed threshold, *Z*:

$$\left|\log LR\left(\frac{H_1}{H_2}\right)\right| > Z$$

Equation 4

By adjusting *Z*, decisions of a particular accuracy can be obtained. The SPRT is optimal in the sense of requiring the minimum possible number of observations to achieve a certain accuracy (Bogacz et al., 2006). It was made famous for its role in deciphering coded messages during World War II; the units of the log likelihood ratio were named 'bans' after the town of Banbury in Oxfordshire, in which sheets of paper were printed for accumulation of evidence favouring one hypothesised setting of the Enigma machine versus another (Good, 1979, Hodges, 1992).

The SPRT is also useful in considering evidence accumulation in neural systems during decision making (Gold and Shadlen, 2002). Such sequential hypothesis tests may be widespread in the nervous system, from the decision of considering what stimuli are present in the environment, to the decision of considering what action to take next.

Perhaps the most formal test of neuronal representations of the logLR to date was performed by Yang and Shadlen (Yang and Shadlen, 2007). In this experiment, monkeys were trained that certain shapes provided a certain WOE that a saccade to a coloured dot would yield a juice reward (Figure 3). Four shapes were presented sequentially, and monkeys were found behaviourally to integrate across the information provided by all shapes to form their decision. During evidence accumulation, neuronal activity in direction-selective cells in LIP reflected the cumulative evidence that a saccade in that direction would yield reward (Figure 3). LIP activity scaled linearly with the bans of evidence in favour of this saccade over the alternative (Figure 3).

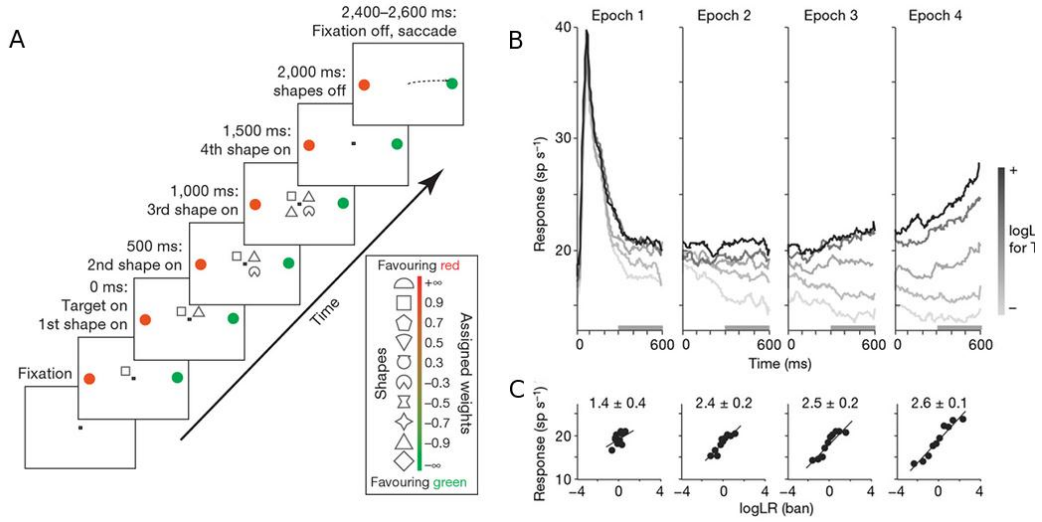

Figure 3. Neural recordings in LIP reflect log likelihood of responses towards receptive fields. A: Monkeys were shown a sequence of shapes, each with assigned weights favouring a saccade towards green or red dots. B: Population firing rates in LIP at each epoch, sorted by the logLR of a saccade into the receptive field. C: Firing rates scaled linearly with the bans of evidence in favour of a saccade at each shape presentation.

It remains unclear whether this probabilistic inference is performed by the LIP neurons themselves, or whether it is achieved elsewhere in the brain and subsequently relayed to LIP. Magnetoencephalography data from central and parietal sensors also reflects accumulation of evidence in human subjects performing a similar decision task (de Lange, Jensen and Dehaene, 2010). However, the signal is inversely related to the LLR, and so perhaps represents a measure of response uncertainty during evidence accumulation.

2.2.2 The drift diffusion model provides a continuous extension of the sequential probability ratio test

In many situations, evidence will not consist of a sequence of discrete observations, but will instead be a continuous process, such as when viewing a noisy visual stimulus (for example, trying to read a roadsign when driving in heavy rain). The SPRT can be extended into a continuous process by assuming that the logLR is captured by a decision variable, *x*, whose value is determined by a differential equation:

$$dx = Adt + cdW$$
$$x(0) = 0$$

Equation 5

where *A* is the drift rate of movement of the decision variable (i.e. the logLR) and *cdW* represents Gaussian white noise sampled at every timestep, with mean 0 and variance c^2dt. A decision is made once the variable *x* exceeds some pre-specified threshold. This is the *drift diffusion model* (DDM) (Ratcliff, 1978). Some example trajectories of the *x* through time are shown in (Figure 4).

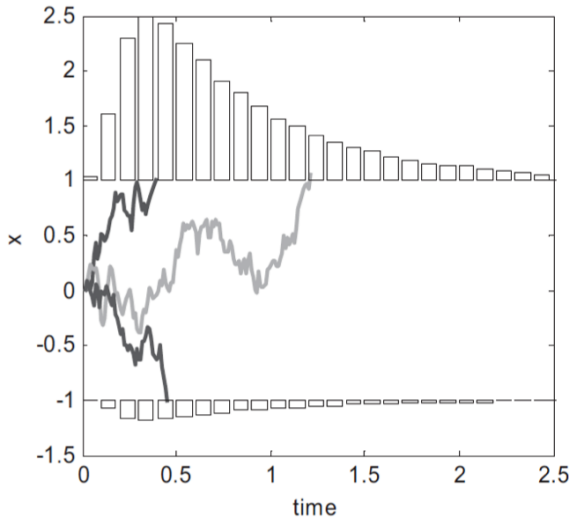

Figure 4. Example trajectories of the drift diffusion model. Three example trajectories are shown; one fast and one slow trajectory reaching the 'correct' answer (upper bound), and one intermediate trajectory reaching the 'wrong' answer (lower bound). Above and below are the histograms of reaction times for correct and error trials, respectively. Adapted from (Bogacz et al., 2006).

As the DDM is a continuous implementation of the SPRT, it is again optimal in the sense of yielding the shortest possible reaction time for a certain decision accuracy, or the greatest decision accuracy for a fixed reaction time. It has proven particularly

popular in experimental psychology as it captures several features of subject behaviour during decision tasks (Carpenter and Williams, 1995, Ratcliff, Van Zandt and McKoon, 1999). These include appropriate error rates and reaction time distributions, such as the skewed distribution towards longer RTs, and a speed-accuracy tradeoff that can be obtained by varying the decision threshold. In some cases an extended version of the DDM may be used, with inter-trial variability in A and $x(0)$ accounting for some further behavioural features of the data (Ratcliff and Rouder, 1998). Further continuous evidence accumulation models have been discussed, including those in which the x is repelled from or attracted to the point x=0 (Busemeyer and Townsend, 1993), models in which evidence for the two alternatives is accumulated separately in an accumulator race (Vickers, 1970), or in which two accumulators mutually inhibit one another as they gather evidence (Usher and McClelland, 2001). However, certain parameterisations and simplifications of all these suboptimal models (with the exception of the race model) reduce them to the optimal DDM (Bogacz et al., 2006).

2.2.3 Neural activity similar to a DDM decision variable can be observed during perceptual decision tasks

In addition to accounting for features of subject behaviour during decision tasks, the DDM also provides a prediction of what a 'decision signal' should look like. A signal that closely resembles the decision variable x through time may reflect a process of gradually integrating the evidence for a particular decision alternative. Such signals can be isolated in saccade-selective regions of parietal and frontal cortex. In a simple forced-choice saccade task, Hanes and Schall (Hanes and Schall, 1996) showed that trial-to-trial variability in response times was reflected in the rate at which neuronal firing increased in saccade-selective frontal eye field neurons. This trial-to-trial variability was therefore similar to trial-to-trial variability in the drift rate A in the DDM. Importantly, it was found that a decision was terminated at a fixed threshold common to all trials, as opposed to an alternative hypothesis that variability was driven by a variable threshold with fixed accumulation rate (Nazir and Jacobs, 1991).

These findings were extended in a series of experiments by Shadlen and colleagues (Shadlen and Newsome, 1996, 2001, Roitman and Shadlen, 2002, Huk and Shadlen, 2005, Churchland, Kiani and Shadlen, 2008, Kiani, Hanks and Shadlen, 2008, Churchland et al., 2011) using a task in which the coherent direction of motion has to be detected from within a stereogram of randomly moving dots. By varying the coherence of motion, the evidence for one direction can be parametrically manipulated from trial to trial. The effect of this manipulation on behaviour is well captured by varying the drift

rate *A* in the DDM; response latencies and error rates both decrease as *A* increases. When recording from cells in the saccade-selective region LIP, neuronal activity ramps up at a rate proportional to *A* until a decision threshold is reached (Figure 5). Once the decision threshold is reached, no further evidence is accumulated from the stimulus if the monkey is constrained to respond later in the trial (Kiani, Hanks and Shadlen, 2008), or a saccade is executed if the monkey is free to respond at any time (Roitman and Shadlen, 2002) – both consistent with integration to a fixed bound. Recent evidence has also highlighted that the cross-trial variance (as well as the mean) of LIP spike counts matches well with a DDM (Churchland et al., 2011). Similar DDM-like activity has also been found in frontal brain regions that are saccade selective (Kim and Shadlen, 1999). Microstimulation of both the frontal eye fields (Gold and Shadlen, 2000) and LIP (Hanks, Ditterich and Shadlen, 2006) biases decisions in favour of the stimulated direction and away from the unstimulated direction.

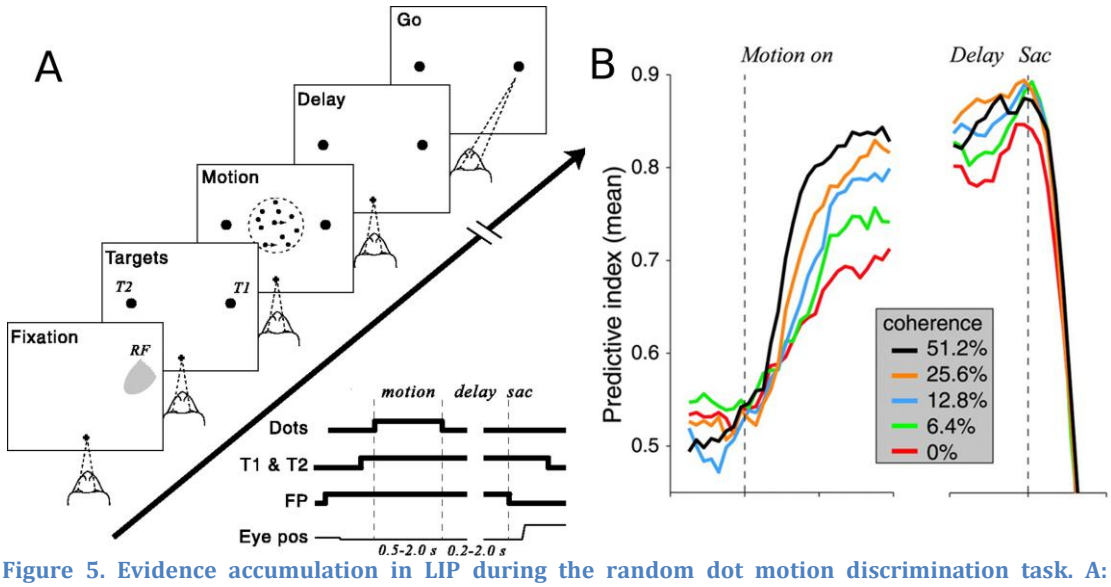

Figure 5. Evidence accumulation in LIP during the random dot motion discrimination task. A: Experimental timeline. Monkey is presented with two targets, views motion pulse, and then makes saccade in direction of inferred motion after delay period. B: Neural activity in LIP ramps at a rate dependent upon the coherence of motion in the stereogram, but reaches a common threshold prior to decision.

DDMs have also been used to derive predictions of decision-related activity in human fMRI data collected during perceptual decision tasks. Heekeren and colleagues (Heekeren et al., 2004) used a face-house discrimination paradigm, and searched for regions that were more active on easy trials than difficult trials, as they argued that this would reflect faster diffusion of a decision variable. They then searched within these regions for those that correlated with the absolute difference in activity between face-

selective and house-selective voxels in ventral temporal cortex, suggestive of an integration of the difference in inputs between these regions. They found an area of the superior frontal sulcus that satisfied both of these conditions, and argued that this region might fulfil the criteria necessary for a decision-making signal. Forstmann and colleagues (Forstmann et al., 2008) fit a DDM to subjects performing a random dot stereogram task during fMRI under conditions of varying time pressure. They found that in the pre-supplementary motor area and the dorsal striatum, individual differences in the rate of accumulation covaried with individual differences in a (fast – slow) contrast of brain activity during the task. They argued that these regions might implement variable setting of the reaction time during perceptual choice, either by varying the rate of evidence accumulation, or the threshold that the decision variable must reach.

However, the predictions of the fMRI signal from the DDM model are somewhat unclear. First, if a region contains some neurons that correlate positively with the decision variable for a particular response, and others negatively (as is the case in LIP/FEF, and presumably other regions), what is the prediction of a signal that reflects both populations simultaneously? Second, the predictions will vary depending upon whether activity is assumed to stay high after a bound is reached (and so be greater in easier trials) (Heekeren et al., 2004) or return to baseline after the bound is reached (and so be greater in more difficult trials) (Basten et al., 2010). Finally, many of the key predictions of the DDM relate to how the decision variable evolves through time, and the slow haemodynamic response means that fMRI is limited in how well it can disentangle these predictions. It may be that a time-resolved technique, such as magnetoencephalography, could provide a more rigorous test of some of the predictions made of the data.

2.2.4 Biophysically realistic models can be used to implement a drift diffusion-like model

The drift diffusion model, along with other continuous time models of decision making, was originally devised to make predictions of behavioural rather than neural data. Whilst several features of neural data seem to match with DDM predictions, it is not immediately obvious how such a model might be implemented in a neural circuit. In the random dot stereogram task, for example, the drift rate A is typically assumed to reflect a subtraction of motion away from a neuron's receptive field from motion towards a receptive field (Shadlen and Newsome, 2001); it is unclear how this subtraction mechanism might be performed neuronally. This has led Wang and colleagues to investigate whether the integration of evidence predicted by a DDM can be

observed in a biophysically realistic model of integrate-and-fire neurons (Wang, 2002, Lo and Wang, 2006, Wong and Wang, 2006, Wong et al., 2007, Furman and Wang, 2008, Wang, 2008). They find that a recurrent neural network endowed with N-methyl D-aspartate (NMDA) receptors to allow for slow integration of evidence can indeed perform probabilistic decision making based on noisy continuous inputs.

The biophysical models that they have studied exhibit 'attractor' dynamics – namely, once the network reaches a certain state (e.g. one population of neurons are high firing), it can stay in this state even in the absence of sensory input. Such networks were originally developed to capture the persistent and stable selective activity seen in dorsolateral prefrontal cortex during spatial working memory tasks (Goldman-Rakic, 1992, Compte et al., 2000). They depend upon NMDA receptors with long time constants to allow for stability of the attractor state, and GABAergic competition between selective pools to ensure only one pool of neurons reach the persistent high firing state, at the expense of other neuronal pools.

It was quickly realized that such competitive pools might also be able to realize the competition between selective cells that is witnessed during the random dot stereogram task (Wang, 2002). In this context, each pool of neurons in the network is used to model LIP neurons selective for a particular saccadic direction (Figure 6). By presenting inputs that reflect the instantaneous degree of motion in each direction (presumed to be projections from motion-sensitive neurons in MT), the network slowly integrates the difference in these two inputs to form a categorical choice, with one LIP pool ending up in a high firing (selected) state, and the other LIP pool ending up in a low firing (non-selected) state. Not only do reaction times and error rates derived from the model match closely with observed behavioural data, but also several features of LIP activity are reflected in model firing rates. These include the predicted timecourse of LIP activity on error trials, a mechanism for random selection ('coin tossing') on trials of 0% coherence, and sustained activity (as in the working memory paradigm) of the selected option after removal of the stimulus. This last point is particularly telling, as saccade selectivity in a working-memory guided saccade task is typically used to select the cells for investigation in the random dot stereogram task (Shadlen and Newsome, 2001).

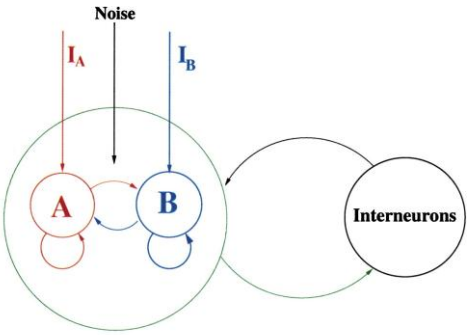

Figure 6. Schematic of biophysical model of decision making. Two pools of neurons, A and B, represent pools selective for different saccades in LIP. They receive inputs I reflecting the degree of coherent motion in that direction, as well as noise inputs. The attractor dynamics of the network are realised through recurrent self-excitation and mutual inhibition via a shared pool of inhibitory interneurons. (adapted from Wang, 2002)

Subsequent work investigated precisely what mechanisms in the circuit model supported the decision process (Wong and Wang, 2006). The network model was reduced to a simplified 'mean field approximation' which contained only 11 dynamical variables, and then an even simpler two-variable model which represented the firing rates of only the selective populations. These reduced models still exhibited all the main features of the original network model, but allowed for a more detailed investigation of variations in network parameters on its behaviour. It was possible to derive the *nullclines* of the neural populations – that is, where their temporal derivative of their activity is zero – under various input (motion stimuluation) parameters. This makes it straightforward to extract the attractor points of the network, as they are the locations where the nullclines for the two populations intersect (Figure 7). It was also established that NMDA receptors were essential for the slow integration of evidence in the model, and that increasing the AMPA:NMDA ratio led to a regime in which the network 'latches on' to decisions rapidly, and makes more mistakes. Subsequent work showed that introducing a motion pulse to the network matched well with predictions from LIP recordings (Wong et al., 2007) and that a version of the model could be developed that extended to multiple options (Furman and Wang, 2008). Finally, recent investigations have highlighted similarity of LIP recordings to the biophysical model's predictions of between-trial variance in spike counts (Churchland et al., 2011).

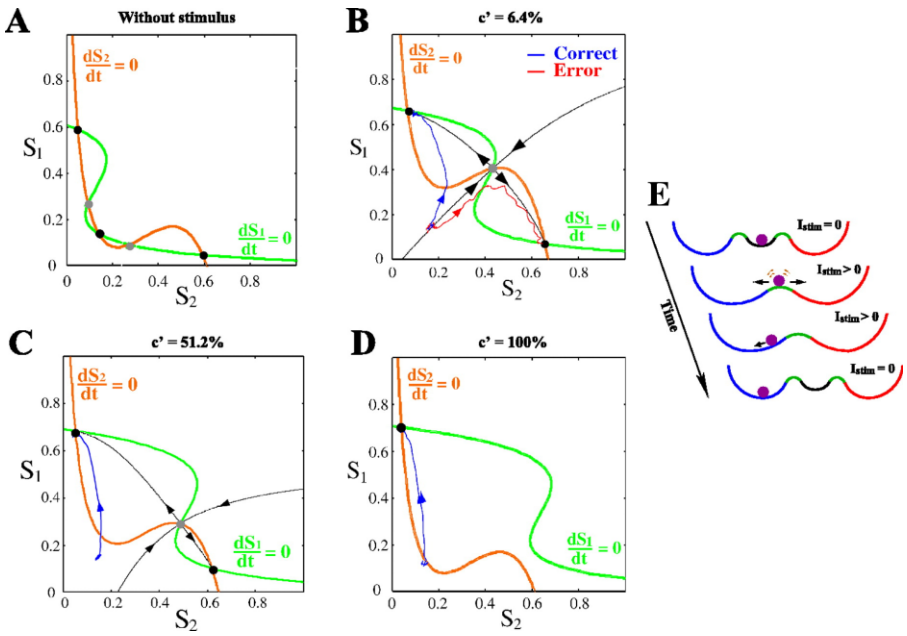

Figure 7. Behaviour of biophysical network model under different motion coherences. A-D: Phase plane representation of network model. Nullclines of populations 1 and 2 are shown in green and brown respectively; black dots are attracting fixed points (attractor states), grey dots are repulsive fixed points. A: Without stimulus, the network can either be in a low firing attractor state (e.g. pre-stimulus) or an attractor state where one population is high firing (e.g. post-decision delay period). B: With stimulus, fixed point with both populations high-firing is repulsive; stable attractor points are only where one population is high firing. Blue and red traces show example single trajectories for correct and error trials, respectively. C: Increasing coherence changes the energy landscape, making it increasingly difficult to reach the 'error' attractor state. D: 100% coherence changes energy landscape such that there is no attractor for error trials. E: A 1-dimensional schematic of the energy lanscape, showing the movement of the neural populations (represented by a ball) before and after stimulus onset. (adapted from Wong and Wang, 2006)

2.2.5 Biophysical networks can be extended into reward-guided decision making

Whilst the drift diffusion and biophysical models provide a good account of neural activity during perceptual decision tasks, it is not immediately clear how this modeling should translate into value-guided choices. As discussed in **chapter 1**, several studies have investigated activity in LIP during value-guided decisions, showing that activity correlates with the subjective desirability of making a saccade to a spatial location (Platt and Glimcher, 1999, Dorris and Glimcher, 2004, Sugrue, Corrado and Newsome, 2004, Seo, Barraclough and Lee, 2009). In these studies, however, all the information needed for choice is available immediately to the animal, and does not necessarily need to be integrated through time. Do similar computational mechanisms apply to perceptual and value-guided choice, or might they be achieved using distinct strategies?

As in the perceptual domain, a clue arises from behavioural data collected during such tasks. Perhaps the simplest example of this is the comparison of two numbers, to

determine which is the largest (Sigman and Dehaene, 2005). Here, subject reaction times decrease with increasing difference between the two numbers, and this effect is well captured using a drift diffusion model. A DDM is also appropriate in capturing reaction times in a value-guided decision task involving comparison of two food rewards (Krajbich, Armel and Rangel, 2010, Milosavljevic et al., 2010). The models provide an account of why value-based decisions tend to be probabilistic rather than deterministic in nature, and describe error rates under conditions in which the value difference between options are varied. The underlying assumption in these models is that although noise is not intrinsic to the stimulus, numerical and value representations in the brain will be intrinsically noisy, and so integration through time is still required to allow for successful choice behaviour.

Direct evidence for integration in neural activity during value-guided choice is still rather limited, however. Soltani and Wang have adapted the biophysical model discussed in section 2.2.4 to account for subject behaviour in tasks when the probability of reward varies across trials, using a Hebbian learning mechanism (Soltani and Wang, 2006). This model accounts for behavioural data, and some features of LIP data, collected during a dynamic foraging task (Sugrue, Corrado and Newsome, 2004). One key similarity between model and data is the evolution of value-related signals in LIP through time as a decision is made (Figure 8). But it is unclear whether further key predictions of neural activity from this biophysical model holds during value-guided choice. It is equally unclear whether the predictions of these models apply only to LIP, or also to the numerous other cortical regions in which value signals are found (reviewed in **chapter 1**). Modelling of activity in these regions may provide a way of disambiguating competing hypotheses of their function – such as whether value-related activity observed in ventromedial prefrontal cortex is reflective of initial valuation of an object (Kable and Glimcher, 2009), or of the comparison of values of different objects (Boorman et al., 2009, Noonan et al., 2010). The work presented in this thesis takes predictions from a biophysical model of decision making, and tests them in neural data collected during value-guided choice.

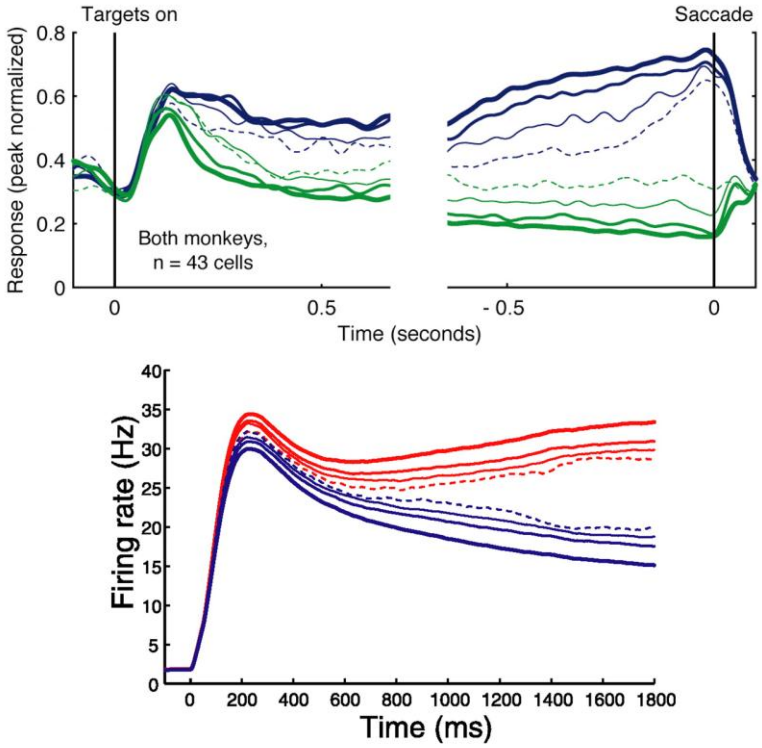

Figure 8. Timecourse of LIP activity during value-guided choice. Top panel: time course of activity related to the local fractional income (subjective value) for choices into (blue) and out of (green) the cell's receptive field, timelocked to stimulus (left) and saccade (right). Trials are subdivided according to the local fractional income of the chosen target. Note the increasing effect of local fractional value during the emergence of the decision. Bottom panel: The effect is recapitulated in a biophysical network model of LIP activity. (adapted from Sugrue et al., 2004 and Soltani et al., 2006)

2.2.6 Thesis work relating to biophysical modeling of value-guided decision making

In **chapter 5** of this thesis I present an analysis of a biophysical model of value-guided decision making (Wang, 2002, Wong and Wang, 2006) that makes novel predictions of the local field potential that may be observed in a region performing comparison during value-guided choice. For reasons discussed in **chapter 3**, the local field potential is assumed to reflect the summed post-synaptic potentials at excitatory pyramidal cells in the network. In **chapter 5**, these predictions are tested in magnetoencephalography data collected during a value-guided decision task. Crucially, the network model is found to predict signals whose *content* is unrelated to the *function* of the network. I outline more clearly the philosophical distinction between functional and content representations in neural activity in **chapter 7**. Finally, the model also makes predictions of behavioural data for the tasks. The behavioural findings from the study are presented in **chapter 4**.

2.3 Models of learning via reinforcement

The valuation and decision models described in the previous sections assume that estimates of the value of a given stimulus or action are *known* prior to the decision. By contrast, in the real world these values must typically be learnt, via previous experience of the reinforcers contingent upon a certain action, or presentation of a certain stimulus. This is a similar distinction to that which is sometimes drawn in economics, between decisions made under *risk* and those made under *uncertainty* (Knight, 1921). In the former case, the precise probability distribution of possible outcomes is known to the decision-maker; in the latter case, there is uncertainty about this probability distribution (and this uncertainty can normally be reduced via learning). Studies of how organisms learn and adapt associations between specific stimuli, actions and rewards formed the basis of much behavioural research in the 20th century. From the mid-20th century onwards, this research was bolstered by attempts to model this learning quantitatively. Recent findings in neurophysiology and neuroimaging have highlighted similarities between these quantitative models and neural recordings.

2.3.1 Theories of operant and classical conditioning depend upon an index of the surprise *associated with reward delivery*

In **chapter 1**, we reviewed how behaviourist psychologists at the turn of the 20th century began to explain learning in terms of the framework of operant (stimulus-response) and classical (stimulus-reinforcer) conditioning. A key development in the mid 20th century was the introduction of more formal mathematical models of the acquisition of stimulus-response and stimulus-reinforcer contingencies. The key idea established within these models was that *surprising* events were critical to driving learning. This was first described by Estes, Bush and Mosteller, who used differential equations to characterise changes in responses through time until an asymptote was reached (Estes, 1950, Bush and Mosteller, 1951, Bower, 1994). Their approach was later extended by Rescorla and Wagner to account for changes in the association strength between CS and US through time (Rescorla and Wagner, 1972).

Rescorla and Wagner proposed that for a stimulus *a* presented at trial *t*, the strength of a CS-US association, $V_{a(t)}$, would be updated using the following equations:

$$\Delta V_{a(t)} = \alpha\beta(\lambda - \sum_{s} V_{s(t)})$$

$$V_{a(t+1)} = V_{a(t)} + \Delta V_{a(t)}$$

Equation 6

In these equations, λ is the outcome (positive or negative reinforcement) on a given trial. α and β are learning constants associated with the stimuli and the reinforcer respectively (constrained such that $0<\alpha,\beta\leq1$). s are *all* stimuli presented on a trial (allowing the theory to account for phenomena involving multiple stimuli, such as 'blocking', in which a previously learnt contingency blocks the learning of an additional contingency (Kamin, 1969)).

From equation 6 we can see that the change in associative strength is driven by the difference between the *expectation of value* after stimulus presentation, $\sum V_{s(t)}$, and the *outcome* on the associated trial, λ . This difference is termed the *prediction error*, and is a measure of how *surprising* the outcome was – if the expectation and outcome are the same, there is no surprise when the outcome is delivered, and no new learning occurs. The $\alpha\beta$ term, sometimes amalgamated into a single value, is termed the *learning rate*, and determines the rate at which associations change through time. Subsequent theories attempted to account for how this term might change through time, between stages in the experiment where the organism is highly uncertain in its estimate of V_a and those where learning has reached an asymptote (Pearce and Hall, 1980). This is discussed further in section 2.3.3.

2.3.2 TD models extend the Rescorla-Wagner model to continuous time and predict dopaminergic firing

A limitation of the Rescorla-Wagner model is that it assumes events occur in discrete trials, and does not provide any account of the role of *temporal contingency* in learning – for example, that an association should only be formed if a stimulus or response *precedes* a reinforcer in time, and that intra-trial stimulus dependencies can affect learning. The real world is clearly not divided up into discrete experimental trials; this becomes even more tangible when trying to use associative learning strategies to design *artificial* agents capable of learning via interaction with their environment. This problem led researchers in machine learning to develop models appropriate to continuous interaction with the environment (Sutton and Barto, 1990). One model that has been particularly influential in both machine learning and neuroscience is *temporal difference* (TD) learning. The goal of a TD model is to build a prediction of the discounted sum of expected future rewards - the current *state value* - based on all current and previously presented stimuli. The full details of this approach are beyond the scope of this chapter (Sutton and Barto, 1998), but the key factor that drives learning of state values is the *temporal difference* prediction error, $\delta(t)$:

$$\delta(t) = r(t+1) + V(t+1) - V(t)$$

Equation 7

V(t) is the state value at time *t*, *V(t+1)* is the state value at time *t+1* (sometimes this is scaled by an exponential discount factor, γ) and *r(t)* is the reward delivered at time *t*.

The critical feature of this prediction error is that it depends both upon *reward delivered* and also *transitions into more or less valuable states than were predicted*. Consider, for example, a CS that always predicts reward delivery 1.5s later. Once this association is fully learnt, the unexpected appearance of the CS elicits a positive prediction error, as *V(t+1)* is greater than *V(t)*. The subsequent appearance of the US, however, does not generate any prediction error, as the delivery of reward *r(t+1)* is cancelled out by the transition from a valuable state *V(t)* into a state *V(t+1)* which is no longer predictive of reward. TD models are still dependent upon a measure of surprise to drive learning, but crucially, it makes predictions of the timing as well the delivery of reward, and these inform when a prediction error should be seen.

The TD model has been of particular interest to neuroscientists as it provides a parsimonious description of firing rates of dopaminergic neurons in the ventral tegmental area (VTA) and substantia nigra pars compacta during classical conditioning experiments. As discussed in **chapter 1,** recordings from these regions revealed that dopaminergic neurons fired not only in response to rewarding stimuli such as food and water, but also to stimuli predictive of reward delivery (Romo and Schultz, 1990, Schultz, Apicella and Ljungberg, 1993). Responses to rewarding stimuli were found to be critically dependent upon whether they were *un*predicted (Mirenowicz and Schultz, 1994). This led Schultz and colleagues to hypothesise that dopamine neuron activity reflected a TD prediction error (Schultz, Dayan and Montague, 1997), an account that explained both these phenomena (Figure 9). Subsequent, more formal tests confirmed that dopaminergic activity closely reflected several properties of TD learning accounts (Hollerman and Schultz, 1998, Waelti, Dickinson and Schultz, 2001).

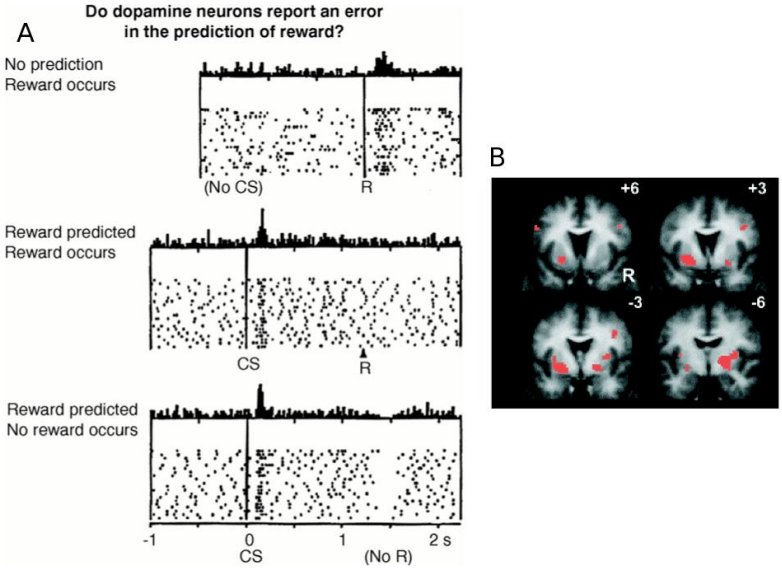

Figure 9. Reward prediction errors in the dopamine system. A: Single unit recording from the ventral tegmental area of the macaque monkey during classical conditioning experiment. Top: In the absence of a predictive CS, the neuron fires in response to US, or reward (R), delivery. Middle: After learning, a reward prediction error occurs at the time of the CS, but not reward delivery. Bottom: If a reward is unexpectedly omitted, a positive prediction error occurs at CS, and a negative prediction error at the precise time when reward was expected. B: Reward prediction error in a functional MRI study activates the ventral striatum. (adapted from Schultz et al., 1997; O'Doherty et al., 2003).

Subsequent research has isolated correlates of reward prediction errors in human VTA, using functional MRI (D'Ardenne et al., 2008). However, fMRI activity reflecting prediction errors is more commonly found in the ventral striatum (Figure 9) (McClure, Berns and Montague, 2003, O'Doherty et al., 2003, Seymour et al., 2004, Pessiglione et al., 2006). This is often explained in light of the fact the ventral striatum receives dense dopaminergic input from the VTA, and that the BOLD fMRI signal is more closely tied to synaptic input than spiking output (Logothetis et al., 2001). Formal tests of TD models in fMRI data, related to the timing of reward delivery, are also beginning to emerge (Klein et al., under review). Intriguingly, these studies have highlighted differences in the signals that can be isolated in VTA and ventral striatum, suggesting that there are additional influences on ventral striatal fMRI signals beyond that related to dopamine.

The reward prediction error has also been the basis of explanations of the event related potentials observed over frontal medial sensors associated with feedback delivery (Holroyd and Coles, 2002, Cohen, Elger and Ranganath, 2007). This 'feedback-related negativity' is often found to be localized to the anterior cingulate cortex (Debener et al., 2005). As discussed in **chapter 1**, the ACC is closely associated with

roles in action-outcome learning and valuation, and neuronal correlates of reward prediction errors are also sometimes found in ACC (Matsumoto et al., 2007).

2.3.3 Bayesian accounts of reinforcement learning can account for variability in learning rate through the experiment

A factor that has received less attention when searching for neural correlates of reinforcement learning models is the *learning rate* (termed $\alpha\beta$ in the Rescorla-Wagner theory of associative learning). It has been customary to fit the learning rate to behaviour (Daw et al., 2006, Rutledge et al., 2009), or to compare findings with different learning rates, and report fMRI activations that show the strongest results (O'Doherty et al., 2003, Hare et al., 2008). These analyses typically assume the learning rate to be *constant* through an experiment. Recent accounts have emphasized that the learning rate should in fact vary through time, with learning being high when the organism is uncertain about its predictions of the world, and reducing when certainty has increased (Yu and Dayan, 2005, Courville, Daw and Touretzky, 2006, Behrens et al., 2007, Preuschoff and Bossaerts, 2007, Krugel et al., 2009). These approaches are similar in spirit to those proposed by Pearce and Hall (Pearce and Hall, 1980), but with an emphasis on which neural structures support the ability to adapt the learning rate through time.

One study (Behrens et al., 2007) proposed a Bayesian account of reinforcement learning. The key distinction between this approach and the classical approach is that it attempts to track a *distribution* over the expected reinforcement associated with a stimulus, rather than a point estimate (as in the Rescorla-Wagner model). This *prior* distribution generates a prediction of the expected reinforcement to be delivered at each trial. Observing an outcome allows us to infer the *likelihood* of observing this outcome under our prior expectations. We can then update the prior using Bayes' rule, to generate a *posterior* distribution:

$$p(\theta \mid y) \propto p(y \mid \theta) p(\theta)$$

Equation 8

where $p(\theta)$ is the prior distribution over parameter θ, $p(y|\theta)$ is the probability distribution of observing data y given θ, and $p(\theta|y)$ is the inferred posterior distribution. By iterating this rule over multiple observations (with the posterior becoming the prior for the next iteration), Bayes' rule infers the *optimal* estimate of θ given the data observed. This Bayesian approach has been commonly applied in studies of sensory

perception (Weiss, Simoncelli and Adelson, 2002) and motor adaptation (Kording and Wolpert, 2004), but has only recently been extended to reward-guided reinforcement (Yu and Dayan, 2005, Courville, Daw and Touretzky, 2006, Behrens et al., 2007).

In the study by Behrens and colleagues (2007), the parameter being modeled at each trial *i* was the reward probability r_i associated with a particular choice. They constructed a Bayes' optimal model for tracking the probability distribution over r_i in a drifting environment, assuming that r_i would change from trial to trial using values drawn from a beta distribution:

$$p(r_{i+1} \mid r_i) \sim \beta(r_i, v)$$

Equation 9

(Here, the beta distribution has been reparameterised from its usual form, such that r_i represents its mean, and v represents its variance). v determines the width of the prior distribution over r (Figure 10), and so determines the *rate* at which r_i can change from trial to trial. It is equivalent to setting the learning rate in non-Bayesian accounts of reinforcement learning. If v is determined to be large, the environment is deemed to be *volatile*, allowing for r_{i+1} to cover a broad range of possible values (and thus change rapidly from trial to trial). If v is small, the environment is *stable*, meaning r_{i+1} is restricted to a limited range of values (and thus can only change slowly).

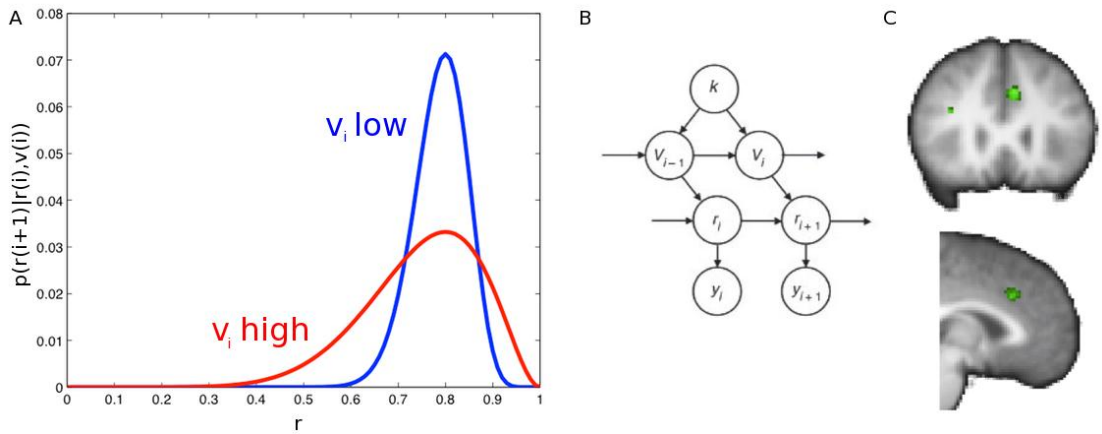

Figure 10. A Bayesian reinforcment learning model allows for volatility-based adjustments in the learning rate. A: The probability distirbution of r(i+1), conditional both upon r(i), here with mean 0.75, and v(i), here shown with high and low values. When v(i) is high, the width of the prior distribution on r is far higher, and so r can rapidly move to another value; the opposite is true when v(i) is low. B: Graphical desciption of the probability-tracking problem. Arrows indivate direction of influence. Data, y is observed under probability r. r can change from trial i to trial i+1, and the rate of change is determined by v. v can also change from trial to trial, determined by control parameter k. C: Trial-by-trial fluctuations in v are found to correlate with activity in anterior cingulate sulcus at the time of reward feedback.

Crucially, it was also assumed in the model that *v* could change from trial to trial. This depended upon the local *volatility* of the reward environment, with the trial-to-trial variations being drawn from a Gaussian distribution:

$$p(v_{i+1} \mid v_i, k) \sim N(v_i, k)$$

Equation 10

where *k* is a control parameter determining the rate of change of *v (*Figure 10*)*.

It is possible to use Bayes' theorem to invert this generative model, in order to build trial-by-trial estimates of the joint distribution of r_i and v_i through time[2]. Behrens and colleagues investigated where in the brain exhibited neural correlates of the modal value of the distribution over v_i in a reward-guided learning task. They found that, at the time when the outcome of the trial was revealed, the anterior cingulate sulcus (ACCs) correlated with this value (Figure 10). Moreover, this region correlated across subjects with individual subjects' learning rates during the task. Activity in ACCs therefore reflected the value of new pieces of information delivered to the subjects in updating their expecations of future reward. This result may provide a computational account of the deficit seen in the study of Kennerley and colleagues, discussed in **chapter 1**. During action-reward association learning in that study, macaque monkeys with lesions to the ACCs were found to have a particular deficit in the weight that should be attributed to new pieces of information provided by reward feedback (Kennerley et al., 2006).

A similar proposal has also been made of the role of the *variance* of reward prediction errors in setting the learning rate across trials (Preuschoff and Bossaerts, 2007); this is of particular interest as the dopaminergic reward prediction error signal is scaled by its variance in single unit recordings (Tobler, Fiorillo and Schultz, 2005). An explicit representation of variance is found in the anterior insula during a gambling task (Preuschoff, Quartz and Bossaerts, 2008), part of a network that is commonly found to be co-active with ACCs in fMRI data.

[2] A limitation of this approach is that it requires the brain to keep track of a full probability distribution over possible values. This appears unlikely at first, but has recently been found to be the case in neural populations recorded during a reward-guided reinforcement learning task. 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..

2.3.4 Theories of learning by reinforcement can also be extended to non reward-guided domains

Error- or surprise-driven signals have also been thought to play a role in learning outside the domain of reward-guided reinforcement. Prediction error signals have been suggested as a general mechanism for learning about statistical regularities in the environment (Friston, 2009, Rushworth, Mars and Summerfield, 2009), and have been observed in both motor and perceptual learning. Perhaps the earliest suggestion of error-based learning in the nervous system can be attributed to Marr (Marr, 1969), who proposed that activation of the climbing fibre input to the cerebellum would drive plasticity in the cerebellar circuit underlying action execution. Later work indicated that complex spikes in cerebellar Purkinje cells, reflective of input arising from climbing fibres, signaled prediction error-like information about the discrepancy of reaching movements in visually-guided arm movements (Kitazawa, Kimura and Yin, 1998). In the perceptual domain, Rao and Ballard have proposed a hierarchical predictive coding model in which higher level units signaled predictions of the incoming sensory input, and lower level units signaled discrepancies (prediction errors) between this prediction and sensory stimulation (Rao and Ballard, 1999). This hierarchical model accounts for certain receptive field effects in early sensory neurons. It has also been extended to a version with multiple hierarchical levels (Friston, 2005) and yielded novel predictions of responses to surprising perceptual events, even if these events are behaviourally irrelevant, which have then been tested in ERP and fMRI data (Summerfield et al., 2006, Garrido et al., 2007, den Ouden et al., 2009).

It is unclear, however, whether these computational principles might also be extended to higher cognitive demands, such as learning about the intentions of another individual during social interactions. Social learning is certainly an important brain function, and may play an important role in shaping value-guided decisions; this is especially the case for humans, in whom the number and complexity of social interactions is unique (Dunbar, 1993). Two literatures have emphasized the role of different brain systems in social inference (Figure 11). The first has highlighted the role of the motor system in simulating the intentions of another individual (Gallese, Keysers and Rizzolatti, 2004), placing particular importance on the discovery of 'mirror neurons' which are active when making particular actions but also when observing others' actions made with a similar intention (Rizzolatti et al., 1996). Computational accounts of mirror neurons have highlighted the possible computational similarity between motor control and intentional inference (Wolpert, Doya and Kawato, 2003). A second literature has highlighted a circumscribed set of regions that are active when having to infer the

beliefs of another individual, such as at the temporoparietal junction and dorsomedial prefrontal cortex; such regions are sometimes claimed to have evolved as functional specializations for social cognition (Fletcher et al., 1995, Saxe, 2006). These regions are distinct from those found when searching for mirror neuron-like signals in human fMRI data(Ramnani and Miall, 2004, Saxe, 2005). Accounts of activity in these regions derived from theories of reinforcement learning have begun to emerge during the past few years (Hampton, Bossaerts and O'Doherty, 2008, Yoshida et al., 2010). In this thesis, I present work which uses a reinforcement learning model to capture the inferred intention of another individual, and then uses fMRI to identify regions correlating with model parameters during value-guided choice.

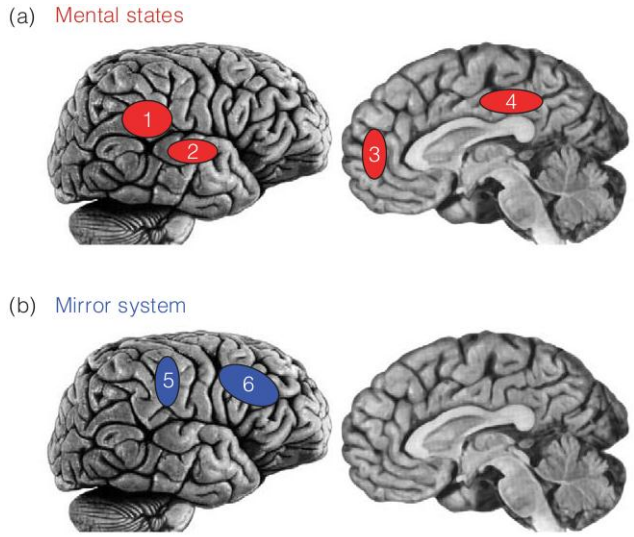

Figure 11. Two proposed systems for intentional inference in the human brain. A: Attribution of mental states using theory of mind is typically found to activiate a network including temporoparietal junction (1), superior temporal sulcus (2), dorsomedial prefrontal cortex (3), and posterior cingulate cortex (4). B: Intentional inference via action observation (or 'simulation'), presumed equivalent of mirror neurons, is found to activate a network including right inferior parietal cortex (5) and inferior frontal gyrus (6). (from Saxe, 2005).

2.3.5 Thesis work related to learning by reinforcement

In **chapter 6** of this thesis I use data collected from an fMRI study of social interaction to present one of the first accounts of activity in theory of mind regions in the theoretical framework of reinforcement learning (Behrens et al., 2008)[3]. I demonstrate that activity in the dorsomedial prefrontal cortex and temporoparietal junction signal a prediction error on the intentions of a confederate, and highlight a novel distinction between gyral and sulcal portions of anterior cingulate cortex, for

[3] An equal contribution to this work was made by the first two authors.

setting the value of social and non-social information respectively. This study highlights that social and reward-based information might be learnt using similar computational strategies, but in distinct *frames of reference* using separate neural substrates; I elucidate this argument in **chapter 7**. First, however, I introduce the task and behavioural modeling in **chapter 4**.

2.4 Summary

Computational modeling allows for precise predictions to be derived of subject behaviour and neural activity during cognitive tasks. Biophysical models have been used to describe single unit data recorded from lateral intraparietal cortex in perceptual decision tasks, but the extent to which these models can describe activity in value-guided choice and in other cortical regions is unclear. Reinforcement learning models are frequently used to describe the learning of stimulus and action values, and even basic sensory and motor learning, but it is unknown whether these models can capture higher cognitive functions such as social cognition. The work in this thesis attempts to resolve some of these issues. In the next chapter, I introduce some of the methodology currently available to study neural correlates of these models in the human brain.

Chapter 3: Non-invasive methods for investigating physiological brain activity in human subjects

Non-invasive measurements of physiological brain activity provide a window onto the functional neuroanatomy of healthy human subjects. Several techniques were developed during the course of the 20th century that allow for such measurements to be made. One study in this thesis primarily use magnetoencephalography (MEG), a technique affording high temporal resolution and adequate spatial resolution, to measure the temporal dynamics of different brain regions during value-guided choice, and compare these dynamics to a computational model of decision making. A second study in this thesis uses functional MRI (fMRI), a technique with high spatial resolution, to investigate the distinct computational roles played by separate anatomical regions during social inference and value-guided choice. This chapter describes basic principles of MEG and fMRI recordings, and methodological considerations specific to the work described in this thesis.

"If a frog is so held in the fingers by one leg that the hook fastened in the spinal cord touches a silver plate and if the other leg falls down freely on the same plate, the muscles are immediately contracted at the instant that this leg makes contact. Thereupon the leg is raised, but soon, however, it becomes relaxed of its own accord and again falls down on the plate. As soon as contact is made, the leg is again lifted for the same reason and thus it continues alternately to be raised and lowered so that to the great astonishment and pleasure of the observer, the leg seems to function like an electric pendulum"

Luigi Galvani, 1791 (trans M.B. Foley, 1953, *Commentary on the effects of electricity on muscular motion*)

Luigi Galvani's discovery that he could turn a frog's leg into a pendulum using static electricity not only provided 'astonishment and pleasure'; his descriptions of 'animal electricity' gave rise to our modern understanding of the role of electrical activity in neural communication.

Galvani's experiments depended upon electrical stimulation of the nervous system. Historically, interference techniques such as stimulation have preceded successful attempts to record the physiological functioning of brain activity. Whilst 'Galvanism' raced ahead through the 19th century with hyperbolical claims of its success (Finger, 1994), it was not until 1875 that Joseph Paton first demonstrated the weak electrical currents that could be induced in the brain by shining light through a rabbit's eye (Cohen of Birkenhead, 1959). Paton's discovery might itself have been forgotten were it not cited in the more widely known work by Hans Berger (1929), who first demonstrated that electrical activity could be recorded from the human brain. Berger was motivated by his quest to discover the *psychiche Energie* (mental energy) that could transmit thoughts from person to person, prompted by an accident during his service in

the German cavalry. His research program here was to fail, but he instead succeeded in detecting the 'alpha' oscillation (~10Hz) that can be measured at rest from an awake human (and becomes stronger when the eyes are closed). This 'electroencephalogram' (EEG), which gained widespread acceptance following subsequent replication by Adrian and Matthews (1934), was the first documented measurement of physiological human brain activity. It was to form the basis of much of the human cognitive neuroscience research conducted in the 20th century.

The EEG is complemented by its close relative, the 'magnetoencephalogram' (MEG). Since the work of Michael Faraday and James Clark Maxwell in the 19th century, it was known that an electrical current induces a magnetic field in the surrounding environment. It should, therefore, be possible to measure the magnetic field associated with the EEG waveforms that Berger had first described. MEG measurements were to prove much more challenging, however, as they required exquisite sensitivity of the measuring device and also suppression of the much stronger magnetic fields naturally encountered in the environment. Such problems were eventually solved using magnetically shielded rooms and highly sensitive magnetometers. The first reports of MEG (again focusing on the robust alpha oscillation) emerged in the late 1960s (Cohen, 1968), and measurements became dramatically more sensitive following the discovery of the superconducting quantum interference device or SQUID (Cohen, 1972). It would later become apparent, when attempting to localize the sources underlying activity measurable at the scalp, that MEG bore several advantages over EEG – due to the magnetic field being less smeared by the resistivity of the skull as it propagates out towards detectors at the scalp (Hamalainen, 1993).

In contrast with EEG and MEG, 'neuroimaging' techniques such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) are relative newcomers in the cognitive neuroscientist's armoury (Phelps and Mazziotta, 1985; Ogawa, 1990; Belliveau, 1991). Both techniques are sensitive to the metabolic consequences of neural activity. PET depends upon injection of a radioactive isotope incorporated into a biological molecule, whose radioactive decay is then measured using detectors placed outside the subject's body. By selecting a molecule that is metabolized when a brain region is active, such as Flurodeoxyglucose, it is possible to measure the activity of the brain that elicits metabolism of that particular molecule. This is a slow process, occurring on the timescale of seconds to minutes. PET became popular in the mid-1970s and 1980s, but many of its applications were soon eclipsed by the development of fMRI, which bore several technical advantages.

fMRI depends upon the haemodynamic response to neural activity, which is an increase in the relative concentration of oxyhaemoglobin to deoxyhaemoglobin caused by an increase in local blood flow. This blood-oxygenation level dependent (BOLD) MRI contrast was first described at the beginning of the 1990s (Ogawa, 1990), with the first demonstration of activation of visual cortex arriving a year later (Belliveau, 1991). The haemodynamic response is specific to the area in which there is increased neural activity, but it takes several seconds to occur. Nevertheless, fMRI has improved temporal (and spatial) resolution relative to PET, and does not require the injection of radioactive isotopes into the experimental subject. Many researchers in cognitive neuroscience therefore shifted to using fMRI as the tool of choice for investigating neural activity in the human brain.

The techniques of M/EEG and fMRI can be seen to be somewhat *complementary* to one another. MEG and EEG are directly sensitive to the electrical activity of the brain, and so can measure changes at the millisecond timescale – the timescale at which most perceptual, motor and cognitive events of interest occur. However, these two techniques are only sensitive to more superficial regions of the brain; and moreover, determining the underlying sources that generated the activity observed at the surface is an ill-defined problem, as we will see below. This means that the *spatial* resolution of M/EEG is somewhat limited. By comparison, fMRI provides us with an excellent spatial resolution, and does so for both superficial and deep structures of the brain. However, it is limited by the sluggishness of the haemodynamic response, which provides a fundamental limit on its temporal resolution. It is therefore important to choose the technique that is most appropriate to the question asked.

In this thesis I adopt MEG for one of the studies (**chapters 4/5**), to investigate predictions derived from biophysical models of reward-guided decision making. These models make fine-grained predictions of the *temporal dynamics* of cortical activity, and so require a technique with high temporal resolution. In one study (**chapters 4/6**), however, I use fMRI to ask a question that is primarily *anatomical* in nature – whether different *subregions* of anterior cingulate cortex are involved in the same computational process, but in distinct frames of reference. Spatial resolution is therefore of utmost importance, but the computational predictions are derived from a reinforcement learning model and do not require a technique with high temporal resolution.

In this chapter, I review the principles of MEG, with particular reference to methodological considerations relevant to this thesis. This is done in quite some detail, as there was considerable need for methodological development and understanding to obtain the results shown in **chapter 5**. I also briefly review the measurement and

processing of BOLD fMRI timeseries, focusing on some of the key points for the study presented in **chapter 6**. As more standard fMRI acquisition and processing pipelines were used for this analysis, the section on fMRI is more limited in scope.

3.1 Magnetoencephalography

3.1.1 What are we measuring with MEG?

3.1.1.1 Passive and active conductances are generated at the neuronal membrane

There are two principle means of communication within the nervous system: chemical and electrical. Both means of communication have a common underlying mechanism, namely the opening and closing of ion channels in the neuronal membrane. Opening these ion channels alters membrane permeability to selective ionic species, and if the ions have an electrochemical gradient across the membrane, causes a change in the local membrane potential. Several ions have electrochemical gradients across the cell membrane. Sodium, chloride and calcium all have a lower intracellular concentration than extracellular, whereas potassium has a higher intracellular concentration than extracellular. This chemical concentration gradient for each ion is balanced by an electrical gradient, and there is a membrane potential for each ion at which the two gradients match and no net movement of charge will occur across the membrane (the 'reversal' potential). The value of this reversal potential relative to the membrane 'resting' potential (around -65mV) will determine whether increases in ionic permeability will 'depolarise' or 'hyperpolarise' the membrane. The reversal potentials for Na^+ (around +40mV) and Ca^{2+} (around 0mV) mean that opening channels selective for these ions depolarizes the membrane; by contrast, opening ion channels selective for Cl^- (around -70mV) and K^+ (around -110mV) will hyperpolarize the membrane.

Chemical transmission involves local release of neurotransmitters that act on *ligand-gated* ion channels to generate small electrical *post-synaptic* potentials across the membrane. The neurotransmitter glutamate is the principle excitatory neurotransmitter in the brain; it acts on AMPA (Na^+-permeable) and NMDA (Na^+/Ca^{2+}-permeable) receptors) to generate excitatory post-synaptic potentials (EPSPs). GABA is the principle inhibitory neurotransmitter, and acts on Cl^--permeable receptors to generate inhibitory post-synaptic potentials (IPSPs). EPSPs and IPSPs primarily occur on the cell's dendrites, where the majority of a neuron's synaptic input is found, and the currents that are generated sum as they propagate towards the neuron's soma. Compared to electrical transmission, they are relatively slow in nature, with PSP durations between a few milliseconds and several hundred milliseconds.

The fast electrical transmission that underlies the *action* potential also modifies the membrane's permeability to these ions, via the opening and closing of *voltage-gated* ion channels – primarily Na^+-permeable (VGNaCs) and K^+-permeable (VGKCs) channels. The action potential initiates when the membrane potential reaches a sufficient threshold to cause rapid opening of VGNaCs, producing a rapid (~0.5ms) depolarization of the membrane. This is followed by inactivation of the VGNaCs, and opening of VGKCs, which repolarizes the membrane. The action potential is thought to initiate in a segment of the axon close to the cell soma, and propagates rapidly along the neuron's long, thin axon, eliciting the release of neurotransmitter at axonal terminals (Figure 1).

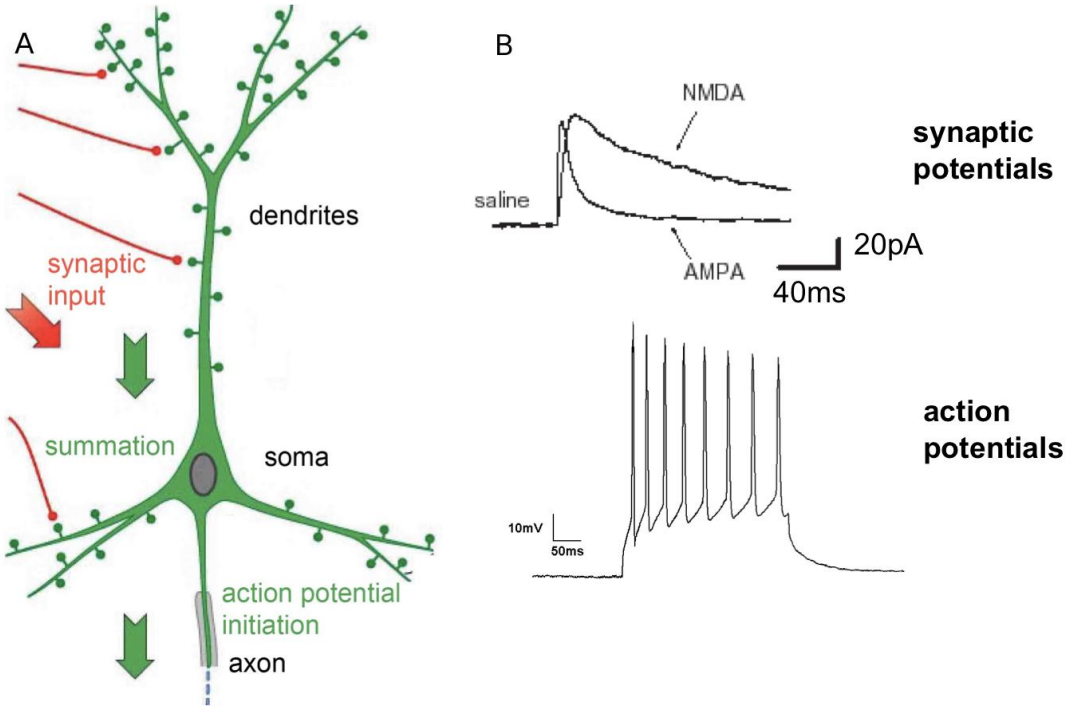

Figure 1. Structure and electrical responses of a simplified pyramidal neuron. A. The neuron receives synaptic input on its dendritic tree, apical (near the top of the picture) and basal (near the bottom of the picture). Excitatory inputs sum as they propogate towards the soma, and initiate an action potential in the initial segment of the axon. B. Excitatory postsynaptic potentials (EPSPs, top) are small in amplitude and vary in duration depending upon the receptor type. Action potentials (APs, bottom) are 'all-or-nothing' spikes of short duration; the action potentials shown here were recorded intacellularly and generated by artificially injecting current through the recording electrode.

3.1.1.2 Electrical potentials generate magnetic fields; dipoles are detectable at distance

The transmembrane current flows that are generated at the level of the neuronal membrane underlie the magnetic fields detectable at the scalp. Each patch of the cell membrane acts as a small current source (outward current) or sink (inward current), and generates an associated magnetic field. The relationship between the electrical currents and the magnetic field is given by Maxwell's equations. Because the fastest

neuronal electrical event (action potential) occurs at a frequency less than 1kHz, we can adopt a 'quasistatic' approximation of these equations (Hämäläinen et al., 1993) that assumes the magnetic field at time *t* depends only upon the electrical potential at time *t* – i.e. that the propagation of magnetic fields to the scalp from electrical currents in the brain is instantaneous. The relationship is also linear, meaning that a weighted sum of two currents will produce the same weighted sum of their corresponding magnetic fields (Kutas and Dale, 1997). The quasistatic relationship between current **J(r')** at a location **r'** and the magnetic field **B(r)** at another location **r** is given by the Biot-Savart law:

$$\mathbf{B(r)} = \frac{\mu_0}{4\pi} \int \mathbf{J(r')} \times \frac{r - r'}{\|r - r'\|^3} dv'$$

Equation 1

where dv' is the differential element of volume and μ_0 denotes the permitivity of free space, a fundamental constant (Hämäläinen et al., 1993, Baillet, Mosher and Leahy, 2001). Notably, the current **J(r')** depends upon both *primary* current flows (relating directly to activity at the neuronal membrane) and *volume* current flows (relating to the effects of electrical potentials on the movement of ionic species in the extracellular space). We can calculate the effects of volume current flows if we accurately model the conductivity profile and geometry of different tissues within the head, as discussed below in section 3.1.3.

Importantly, the magnetic field decays with increasing distance from the current source, and rate of the decay depends upon the form that this current source takes (Hämäläinen et al., 1993). If $\mathbf{r_d}$ is the distance between current source **J(r')** and the magnetic field at location **B(r)**, then the field decays proportional to $1/\mathbf{r_d}$ if the current source is a monopole, $1/\mathbf{r_d^2}$ if the current source is a dipole, $1/\mathbf{r_d^3}$ for a quadropole, and so on. This means that the magnetic field associated with quadropolar current sources (and further higher order terms) decay rapidly as we move away from them, and will be far less detectable at a distance from the source (e.g. at the scalp). The monopolar term from primary neuronal currents will also be negligible, as the total amount of current entering the neuron must equal the total amount of current leaving the neuron (Kutas and Dale, 1997). Thus, the primary current sources and sinks whose magnetic fields are detectable at distance will be dipolar. The sources can be modeled by placing an

'equivalent current dipole' at the location of interest when estimating the resultant magnetic field from cortical activity in a particular location in a forward model[1].

3.1.1.3 Action potentials are primarily quadrupolar, and high frequency

We can then consider whether action potentials or synaptic potentials are likely to generate magnetic fields that will be detectable at the scalp. The action potential is well modeled as two oppositely oriented current dipoles a short distance apart - one corresponding to the leading edge of the action potential, where VGNaCs open and depolarize the membrane, and the second corresponding to its trailing edge, where VGNaCs inactivate and VGKCs open and hyperpolarize the membrane. In an unmyelinated cortical axon, with conduction velocity of 1m/s, the distance between these two dipoles is approximately 1mm (Hämäläinen et al., 1993). As the two dipoles are close and of opposite orientations, they will form a current quadropole at a distance (Plonsey, 1977, Milstein and Koch, 2008). As argued above, the contribution of a quadrupole relative to dipolar sources will be relatively weak when measured at the scalp[2].

Two further considerations also suggest that action potentials may provide little contribution to the MEG signal measured at the scalp. Firstly, the action potential is of very brief duration (~1ms to depolarize and repolarize the membrane), and so very high synchrony would be needed across different cortical cells to generate a sufficiently large gross current to be observed at the surface. The fast, transient nature of the action potential might also suggest that the generated signals would occupy very high frequencies (upwards of 300 Hz for a single action potential), far higher than those typically examined in most MEG studies. Secondly, unlike the dendritic arbors of cortical pyramidal neurons, axons tend not to be closely aligned with one another (except once they enter long white matter pathways), and so might not form the 'open field' arrangement (discussed below, in 3.1.1.4) necessary to generate a macroscopic electric potential.

In spite of this, recent modeling studies (Murakami and Okada, 2006) have suggested that cortical action potentials may generate sufficiently large dipolar fields to be detectable at distance. This finding is perhaps because these modeling studies have used neuronal models that contain dendritic active conductances (Mainen and

[1] But see Jerbi K, Mosher JC, Baillet S, Leahy RM (2002) On MEG forward modelling using multipolar expansions. Phys Med Biol 47:523-555., for a discussion of the potential contribution of multipolar sources.

[2] By contrast, in fast conducting peripheral nerves, the distance between the leading and trailing edges of the action potential generated in the median nerve can be sufficient to generate a dipolar source whose magnetic field is detectable outside the arm (Hari R et al., 1988).

Sejnowski, 1996), somewhat slower than the fast axonal active conductances underlying action potentials. Moreover, there are some circumstances under which high (<1ms) synchronization between action potentials can be seen in recordings from multiple cortical neurons, such as during epileptic seizures (Bragin et al., 2002) or somatosensory stimulation (Barth, 2003). It may be possible that action potentials underlie the magnetic fields observed at very high frequencies (>100Hz) at the scalp (Curio, 2000, Baker, Curio and Lemon, 2003) corresponding to these events.

However, these very high frequency responses are not the focus of this thesis; the predictions from the biophysical model (see **chapters 2 and 5**) are predominantly low frequency (2-10 Hz) responses, and we consider responses in these frequency bands typically studied in EEG/LFP recordings. Action potentials will likely make minimal contribution to the signal recorded at these frequencies.

3.1.1.4 Synaptic potentials may generate dipoles, depending upon dendritic orientation

In contrast with action potentials, postsynaptic potentials at the dendrites are far longer in duration. For example, the decay time constant of glutamatergic receptors ranges from approximately 2ms for an AMPA receptor up to approximately 100ms for an NMDA receptor (Spruston, Jonas and Sakmann, 1995). Moreover, integration of many postsynaptic potentials is needed to generate a single action potential. This temporal filtering of synaptic inputs might mean that physiologically there would be synchronous synaptic input arriving on the dendrites of multiple cells simultaneously. Within a single dendrite, the flow of current can be well approximated by a single dipole, with a current source at one point in the dendrite and a current sink at another location[3]. Certain orientations of dendrites and synaptic input will mean that, within a patch of brain tissue, the current sources and sinks are asymmetrically distributed and give rise to an 'open field' configuration (Lorente de No, 1947) – one that elicits a macroscopically observable equivalent current dipole. An 'open field' configuration is to be contrasted with a 'closed field' configuration, in which the microscopic dendritic dipoles cancel one another out; this can be seen, for example, in radially symmetric current sources or in randomly oriented dendritic trees (Figure 2).

[3] This is strictly true only if the synaptic input is near either the somatic or distal ends of the dendrite. Dendritic input near the middle of the dendritic tree may create a source at input and two sinks - one at the somatic end of the dendrite and a second at the distal end of the dendrite – which will cancel one another to generate no macroscopically observable dipole (Mitzdorf, 1985).

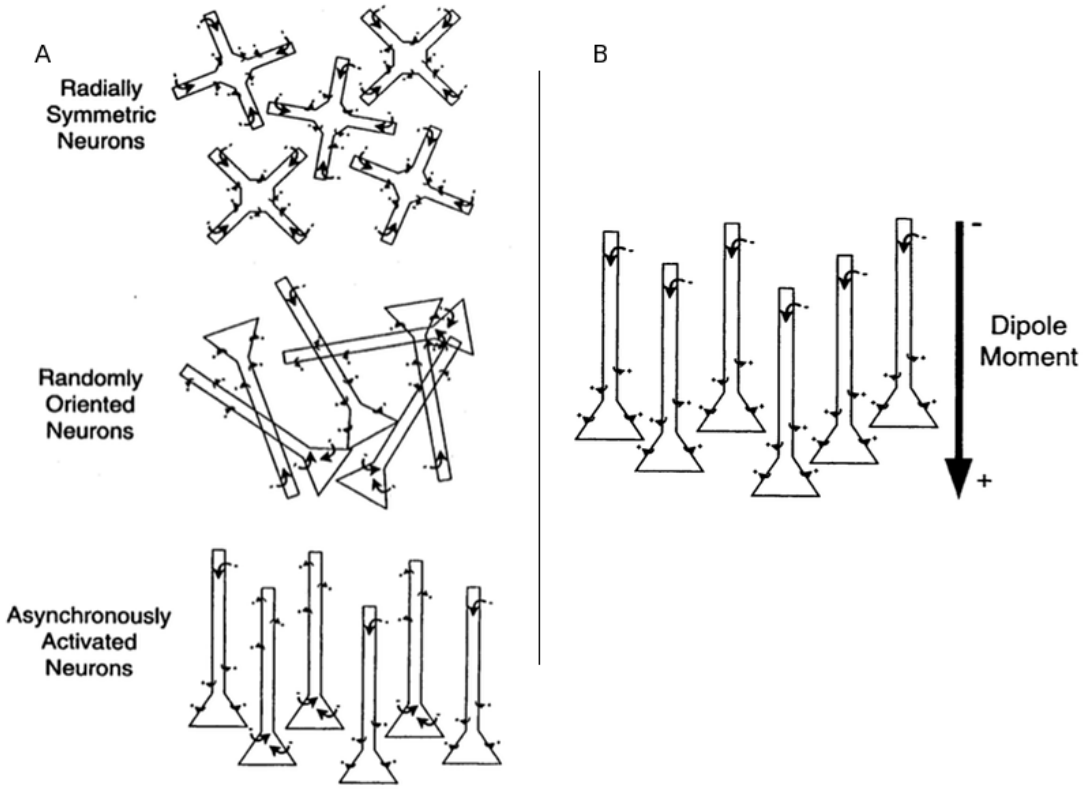

Figure 2. Open and closed fields are generated by different dendritic structures. A. Examples of closed fields. Radially symmetric dendrites (as might be found, for example, in striatal medium spiny neurons), randomly oriented neurons or neurons that are activated asynchronously all show cancelling of the microscopic dipoles that are generated in the dendritic tree, and therefore no macroscopically observable current dipole. B. Example of an open field. Aligned cortical pyramidal neurons receiving synchronous excitatory input to their apical dendrites have a current sink at the top of the dendritic tree and a current source near the soma, yielding a macroscopically observable current dipole. Adapted from (Kutas and Dale, 1997).

Importantly, the arrangement of synaptic input to the apical dendrites of pyramidal neurons in the cerebral cortex is an 'open field' configuration, and capable of generating an externally observable dipole (Mitzdorf, 1985). Cortical pyramidal neurons are aligned with one another, with their apical dendrites from somata in cortical layers III and V extending up towards the cortical surface. Distinct domains of the dendritic tree receive distinct synaptic inputs; inhibitory GABAergic input is principally located proximal to the soma, whereas excitatory input is distributed throughout the dendritic tree (Spruston, 2008). The perisomatic IPSPs are thought to have a far weaker capacitative loss current associated with them the than dendritic EPSPs (Mitzdorf, 1985), and so their contribution to externally observable field potentials (and associated magnetic fields) is far smaller. This has been confirmed by experimental studies (Mitzdorf and Singer, 1979).

The relationship between excitatory synaptic input to pyramidal dendrites and the resultant equivalent current dipole is not straightforward. Different dendritic inputs

arising from distinct origins (thalamic inputs, cortico-cortical pyramidal connections, interneurons) target the dendritic arbors at specific cortical layers. The arrangement of current sources and sinks will vary depending upon where input is delivered on the dendritic tree; a schematic for how synaptic inputs to different cortical layers might result in different scalp EEG deflections (and so different magnetic fields, also) is shown below (Figure 3).

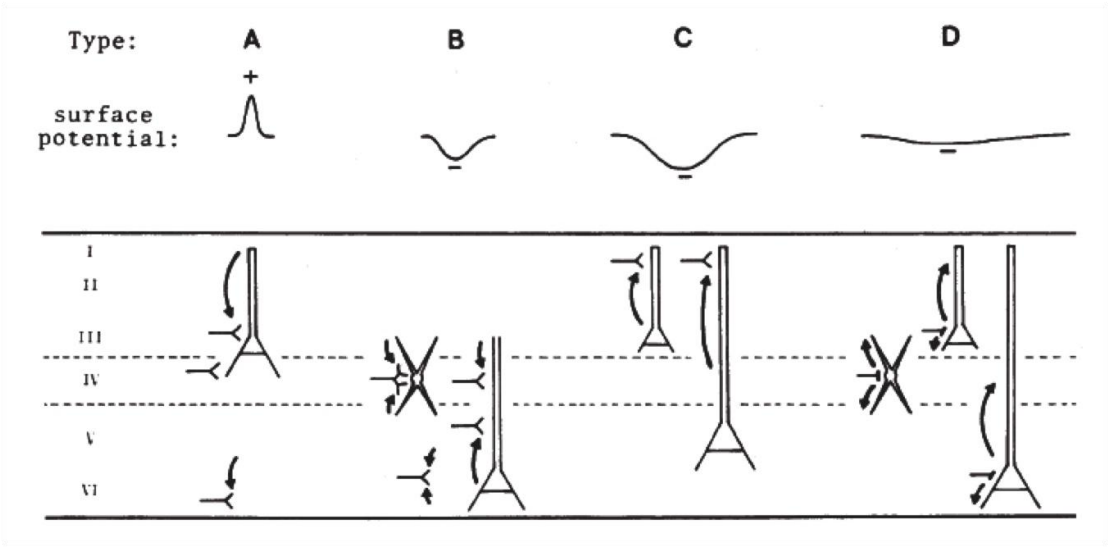

Figure 3. Different excitatory and inihibitory synaptic input to cerebral cortical layers yield different EEG surface potentials. Examples are given for EEG measurements, but a similar interpretation can be drawn for the corresponding MEG signal. A. Somatic excitation of supragranular (layer III) pyramidal cells yields a somatic sink and a source above, producing a positive going EEG surface deflection. B. Excitation at the apical dendrites of layer VI pyramidal cells and at layer IV spiny stellate cells yields a negative going EEG surface deflection, but the field is closed and so the deflection may be relatively small. C. Excitation at the apical dendites of layer III and layer V pyramidal neurons yields a dendritic sink and a source at the soma, producing a negative going EEG surface deflection. D. Cortical inhibition generates a relatively weak surface EEG potential as the net membrane currents that flow during inhibition are small. Adapted from (Mitzdorf, 1985).

The principal conclusion from these results is that we should take the MEG signal to primarily reflect synchronous excitatory input onto cortical pyramidal cells. We should not, however, draw strong conclusions on whether this input has increased or decreased in magnitude based upon whether the change in the cortical dipole is positive or negative in sign.

3.1.2 How are magnetic fields measured at the scalp?

3.1.2.1 The magnetic field generated by the brain is extremely weak, and requires sensitive magnetometers for detection

The amplitude of the magnetic fields generated by the brain is far smaller than other magnetic fields present in the environment. The amplitude of a typical evoked potential (~100 fT), for example, is approximately a billionth of the size of the earth's magnetic field (~30 μT) (Figure 4). Noise cancellation techniques are required to allow for the detection of the MEG signal. Typical environmental noise sources include moving magnetic objects such as cars, trains, metallic doors and people, and electrical devices such as computers and power lines. During MEG recordings, external magnetic noise is minimized by conducting the recordings in a magnetically shielded room, and by minimizing environmental noise sources within this room (for example by placing the display projector outside the room, and by using response pads with no metallic moving parts). However, special consideration must also be made of physiological noise sources originating from the experimental subject - in particular those associated with magnetic fields generated by muscle contraction, such as during saccadic eye movements, eyeblinks, or from the heartbeat. The reduction and elimination of these artifacts is discussed in section 3.1.5.

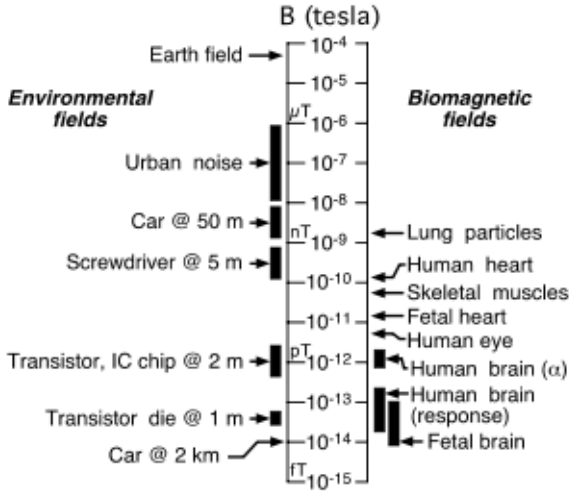

Figure 4. Environmental and biomagnetic noise sources, and comparison to magnetic field generated by human brain. Adapted from (Vrba, 2002).

The superconducting quantum interference device (SQUID) sensor, invented in the mid-1960s, underlies the detection of the very weak magnetic fields generated by the brain. The dc SQUID, typical of most MEG sensor arrays, is a ring of superconducting

material (usually niobium) in a low temperature environment (usually liquid helium) interrupted by two Josephson junctions (Figure 5). When external magnetic flux is applied, the Josephson junctions cause current in the SQUID sensor to oscillate, at a frequency dependent upon the level of magnetic flux (Vrba and Robinson, 2001, Vrba, 2002). These oscillations can then be picked up in an external circuit and amplified electronically. The magnetic flux is applied to the SQUID sensor through a superconducting flux transformer (or 'pickup coil'), which serves to maximize the sensor's sensitivity to the magnetic field.

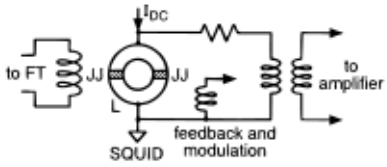

Figure 5. Diagram of a direct current SQUID. Magnetic flux is applied from a flux transformer, which produces oscillations within the SQUID that can be detected by an external circuit and amplified. FT = flux transformer; JJ = Josephson junction; L = SQUID inductor; IDC = externally applied direct current. Adapted from (Vrba and Robinson, 2001).

3.1.1.2 The measured magnetic field and sensitivity depends upon the type of flux transformer used for field detection

There are several different types of flux transformer, sensitive to different components of the magnetic field. The simplest is a single loop of wire, or magnetometer, which measures the magnetic field in an orientation perpendicular to the loop. By arranging two loops together, a first-order gradiometer is made, which is insensitive to homogenous magnetic fields, but instead sensitive to the spatial gradient of the magnetic field. The directional sensitivity of the planar gradiometer varies depending upon the orientation of the two loops; first-order planar gradiometers are maximally sensitive to the gradient of the magnetic field along one particular direction. They are therefore frequently fabricated in a thin film 'double-D' arrangement, with two superimposed planar gradiometers sensitive to magnetic fields in different orientations (Knuutila et al., 1993). The two planar gradiometers in such an arrangement yield orthogonal information about the spatial gradient of the magnetic field. Further, higher-order gradiometers can also be generated by combining first-order gradiometers, and can be helpful for dealing with environmental noise (as the spatial gradient of magnetic fields generated at a distance will be very low) (Vrba and Robinson, 2001). However, higher-order gradiometers are not used in the studies in this thesis, and are not discussed further here.

Importantly, the sensitivity patterns of the different flux transformers also mean that they are differentially sensitive to current dipoles in certain locations and orientations. The sensitivity of a magnetometer and a planar gradiometer (the 'lead field' of the sensor) is shown below (Figure 6) (Malmivuo and Plonsey, 1995). It can be seen from this figure that the planar gradiometer is most sensitive to sources located directly underneath the sensor, and its sensitivity decreases rapidly with increasing distance from the sensor. The magnetometer, by contrast, has a zero sensitivity line directly beneath the sensor, and is instead sensitive to sources located at some distance. The relative strength of the measured magnetic field also decays less rapidly with increasing distance from the sensor than for the planar gradiometer. Thus the planar gradiometer will be maximally sensitive to superficial sources in the brain, and activity at a planar gradiometer will give a reasonable suggestion that the source is located underneath that sensor. The magnetometer will be more sensitive to deeper sources, but it will be very difficult to infer (without source reconstruction techniques, discussed below) whereabouts the source of the magnetic field is located.

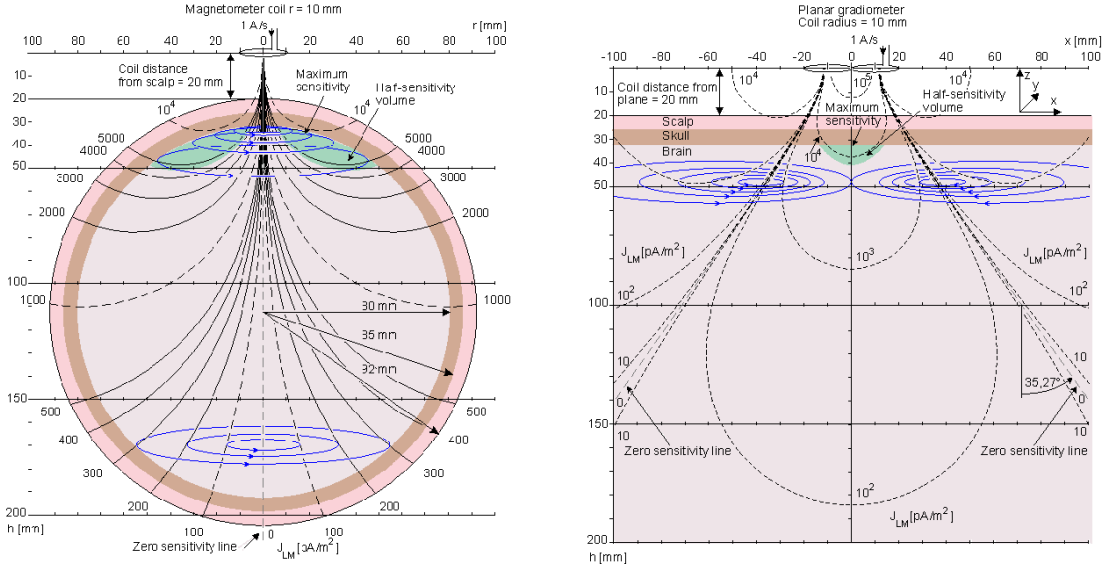

Figure 6. Lead fields of a single magnetometer and a single planar gradiometer. The planar gradiometer (right) has maximal sensitivity to sources directly underneat the sensor, but its sensitivity decays rapidly with increasing distance. The magnetomter (left) has maximual sensitivity off-centre, but is more sensitive to deeper source locations. Adapted from (Malmivuo and Plonsey, 1995).

3.1.1.3. SQUID sensors are typically arranged in a helmet containing several hundred sensors

Early MEG studies used sensor arrays containing only one SQUID sensor, repeatedly moving the sensor over the cortex as subjects repeated the same experiment many times, in order to obtain a map of the evoked magnetic field. Such studies were time consuming and error-prone. Between 1972 and 1994, when systems containing a single channel or a few channels were prevalent, it was estimated that less than 1000 SQUIDs were manufactured in total worldwide (Wiskwo, 1995). From 1992 onwards, however, sensor arrays with many sensors began to be introduced, and during the 1990s helmet-shaped sensor arrays containing over a hundred SQUID sensors became increasingly common. By 2001, nearly 10,000 SQUID sensors had been installed in around 60 MEG helmet systems worldwide (Vrba and Robinson, 2001), and the number of such installations has grown to approximately 160 MEG scanners worldwide in 2011 (source: http://www.elekta.com, accessed 15/04/2011). The studies in this thesis use an Elekta Neuromag system containing 306 sensors in a helmet arrangement, with a 'triple sensor' containing one magnetometer and two orthogonal planar gradiometers at each of 102 locations around the skull.

3.1.3 Forward modeling of MEG data

3.1.3.1 A forward model predicts the expected magnetic field from an equivalent current dipole

Estimating the underlying neural activity that has generated a pattern of responses at the MEG sensors is a two-fold problem. Firstly, an accurate 'forward model' must be built that estimates the predicted scalp distribution that would be produced if a dipole of a given orientation and magnitude were placed at a particular location in the brain. Secondly, this forward model needs to be 'inverted', in order to reconstruct the underlying neural activity that generates the observed data (discussed in section 3.1.4).

The Biot-Savart law, derived from Maxwell's equations and described in section 3.1.1.4, can be used to derive the predicted magnetic field at a given location if the underlying currents generating the field are known. However, there are several currents that need to be estimated, beyond that associated with the primary current dipole, which are likely to impact upon the magnetic field. The volume (return currents) will generate associated magnetic fields, and these will also vary at boundaries between areas of different conductivity, such as the surfaces between the brain and skull, between skull and scalp, and between scalp and air. Once both the primary current and

the potential distribution on all these surfaces is known, we can calculate an estimate of the predicted magnetic field (Baillet, Mosher and Leahy, 2001). In the case of a model consisting of concentric spheres for each of these surfaces, the calculation can be derived analytically; however, more accurate solutions ('boundary element method', or BEM, solutions) can be derived numerically if an accurate representation of the geometry of each of the surfaces is obtained. Further accuracy can be obtained by deriving solutions ('finite element method', or FEM solutions) that capture inhomogeneities in the conductivity of different tissues, generated by factors such as white matter anisotropy and the presence of sinuses in the skull (e.g. (Wolters et al., 2006)).

3.1.3.2 Surface extraction and registration of sensor locations to surfaces is needed

Accurate forward modeling therefore requires estimation of the location of different surfaces (between brain, skull, scalp and skin) that form the boundaries of areas with distinct conductivity, and also the geometry of the cortical surface (if the inverse solution is to use dipoles constrained to this surface). This problem can be solved using accurate surface extraction tools now available to automatically recover surface information from a T1-weighted anatomical MR scan (Dale, Fischl and Sereno, 1999). This process is computationally extensive, however, and can instead be reasonably solved using non-linear registration of a individual's MR scan to a canonical template brain, which is then inverted to derive individual subject surface boundaries (Mattout, Henson and Friston, 2007). This approach has been shown to have approximately equal model evidence to the individually-derived surfaces when using a Bayesian inversion routine for MEG data (Henson et al., 2009).

The location of these surfaces must also be accurately registered to the appropriate location relative to the MEG sensors. Two registrations must take place. Firstly, the position of fidicual points (nasion, left and right pre-auricular points, which can be reliably identified on an MR scan) and the scalp must be identified relative to coils fixed to the subject's head that will generate a signal detectable by the MEG scanner. The location of these 'head position indicator' (HPI) coils will be accurately measured by the scanner by emitting pulses (at a high frequency that is unlikely to contain neural activity of interest), to determine the location of the fiducial points and the scalp surface relative to the MEG sensors. Several systems for this head position registration exist, and the studies in this thesis use a Polhemus Isotrak II system (Colchester, VT). Secondly, the digitized fiducial points and scalp surface must be registered to the surfaces (skin, scalp, skull, cortex) derived from the anatomical MR

scan. This typically uses an interatively reweighted least squares approach to minimize the distance between the digitized locations and the MR-derived surfaces (Mattout, Henson and Friston, 2007). Combining the information from these two registrations, we can calculate the location of the individual subject's surfaces relative to the MEG sensors, and an accurate forward model can be derived (Figure 7).

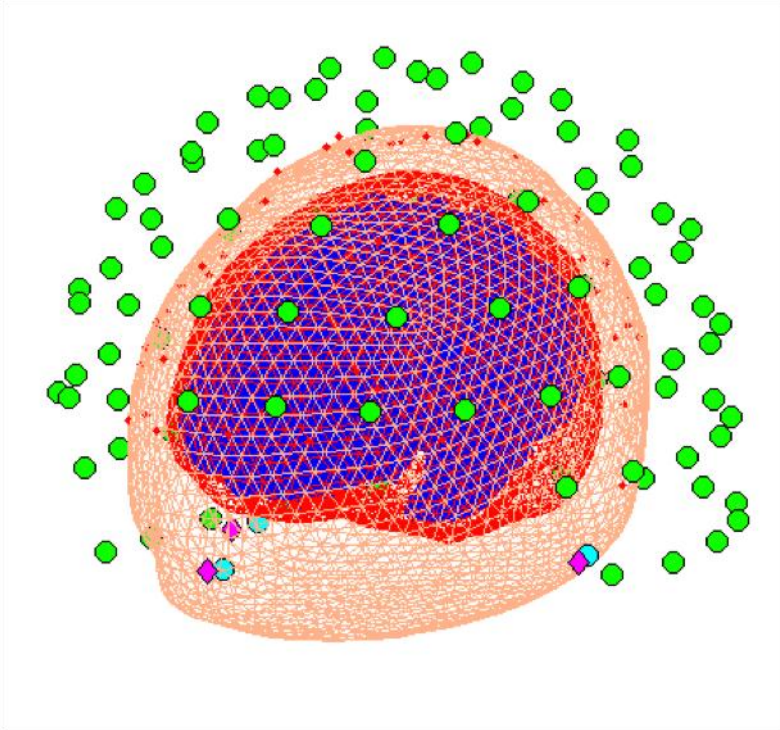

Figure 7. Forward modeling requires accurate registration between individual subject surfaces and sensor locations. The figure shows the registration from an individual subject used in the studies described in this thesis. The individual surfaces, derived from a non-linear registration to a canonical cortical template, are shown; in blue is the cortical surface, in red the skull, and in pink the skin. The three pink diamonds denote the location of the nasion, left and right pre-auricular points on the canonical surface, and the blue circles denote these locations as digitized when collecting MEG data. The small red dots denote the location of EEG sensors on the subject's scalp, and the large green dots denote the location of MEG sensors in the helmet surrounding the subject's head.

3.1.3.3 Predicted fields and sensitivity vary depending upon dipole location and orientation, and sensor type

From the forward model, an estimate can be built of the sensitivity profile of MEG sensors to current dipoles placed in different locations. In a purely spherical forward model, it can be shown that dipoles radially pointing out towards the spherical surface generate no externally observable magnetic field (Sarvas, 1987). This is less strictly true when accurately modeling the geometry of the head surfaces, but it is

nevertheless the case that MEG will be primarily insensitive to those sources radial to the surface of the head, and much more sensitive to those tangential to it. Thus, sources within cortical sulci will be more prominent in MEG recordings, as these are likely to be tangential sources, but those on the banks of cortical gyri or at the depth of the cortical sulci will be less visible, as these are likely to be radial. (EEG, by contrast, will provide complementary information, as it is sensitive to radial sources). Furthermore, cortical locations that are further from the sensors will also generate weaker magnetic fields. It is possible to construct a map (assuming constant signal to noise ratio from each region) to gain an impression of the relative sensitivity of MEG to different regions of cortex (Figure 8) (Hillebrand and Barnes, 2002).

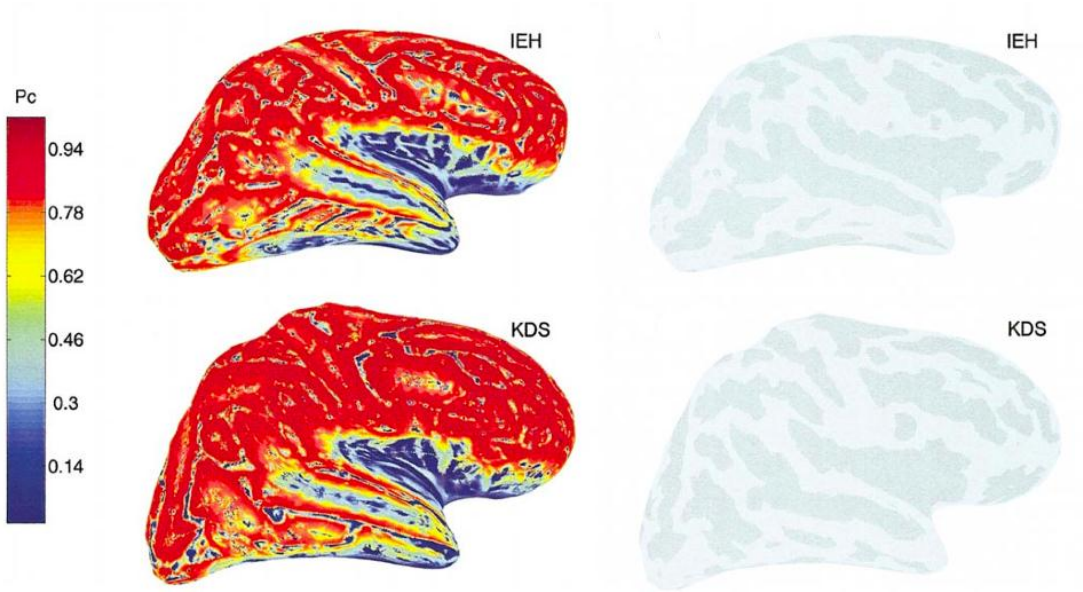

Figure 8. MEG sensitivity to cortical sources varies with cortical orientation and depth. The colour scale reflects the probability of detecting a source of fixed signal to noise ratio, perpendicular to the cortical surface, in each of the different locations, for two subjects. The map is a flattened cortical surface; the surface can be seen on the right, with gyri in light grey and sulci in dark grey. It can be seen that sensitivity is lower in deeper cortical structures (e.g. inferior temporal lobes) and also lower on the crests of gyri and the depths of sulci. Adapted from (Hillebrand and Barnes, 2002).

Once the forward model has been derived, it is also possible to generate simulated data for dipoles of different orientations, locations and magnitudes, to visualize what the corresponding scalp topography should be for a given cortical source, and how this should differ across different flux transformers (and EEG sensors, if also used). This also allows us to get a qualitative feel for how distinct cortical sources in different locations will look at sensor level, and also which sensor types will provide the most discriminative information when distinguishing one cortical region from another. Three examples are shown below (Figure 9) simulated from a single subject used in the

experiments presented in **chapter 5** of this thesis, in locations that have previously been of importance in studies of reward-guided decision making – the ventromedial prefrontal cortex (VMPFC, MNI coordinates x=0mm, y=38mm, z=-16mm), the anterior cingulate sulcus (ACCs, MNI x=0mm, y=24mm, z=32mm) and the posterior cingulate cortex (PCC, MNI x=0mm,y=-46mm,z=32mm). In each case the dipole is of orientation (-1,0,0) – it points horizontally from right to left with no contribution in the vertical (z) or coronal (y) directions. One feature that can be qualitatively seen from this figure is the difficulty in distinguishing sources from the ACCs and VMPFC using EEG sensors, but the relative ease with which these sources can be distinguished with MEG sensors.

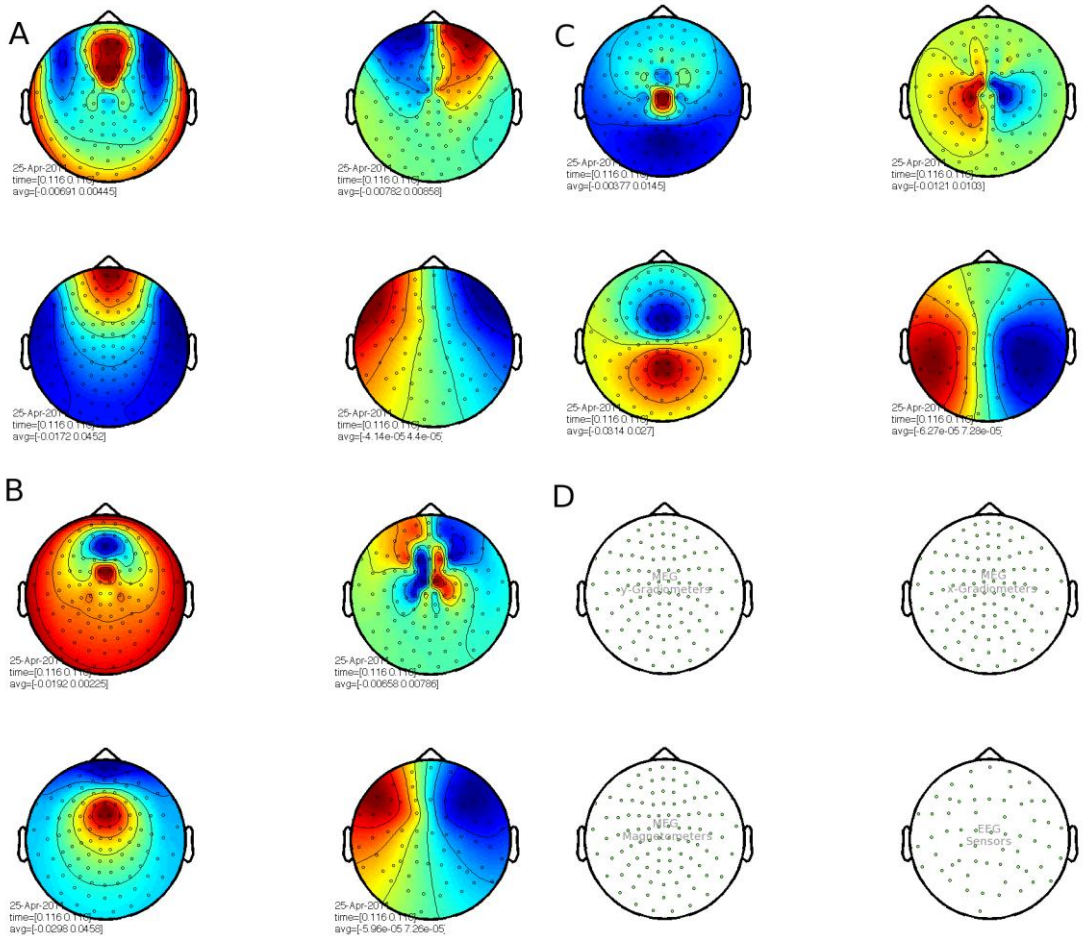

Figure 9. Simulated scalp topographies for dipoles placed in ventromedial prefrontal cortex (A), anterior cingulate cortex (B) and posterior cingulate cortex (C). Simulated dipoles are pointing from right hemisphere to left, with no contribution in y or z directions; see text for MNI co-ordinates. As depicted in D, the top left and right section of each figure show MEG planar gradiometer topographies (sensitive to gradients in y- and x-orientation respectively), and the bottom left and right sections show MEG magnetometers and EEG sensors respectively. Note similarity between EEG scalp topography for VMPFC (A) and ACCs (B), two structures that are often implicated in reward-guided decision making – but relative difference between scalp topography at MEG sensors for these two structures.

3.1.4 Source reconstruction and the inverse problem

3.1.4.1 The inverse problem, unlike the forward problem, is ill-posed

The 'forward problem' described in the previous section is *well-posed* – once the forward model has been constructed, there is one unique solution for each current dipole. When performing a MEG experiment, however, we have the opposite challenge – determining the current sources that generated the data, given the magnetic fields measured at the sensors. Determining activity in 'source space' from these 'sensor space' MEG recordings is an *ill-posed* problem. There are an infinite number of possible source combinations that can be constructed to generate the observations (von Helmholtz, 1853, Baillet, Mosher and Leahy, 2001).

The inverse problem can (in a somewhat simplified framework) be described in terms of the linear contribution of a set of dipoles at locations and orientations of interest to the MEG data:

$$\mathbf{M} = \mathbf{A}\mathbf{S}^T + \mathbf{E}$$

Equation 2

where **M** is a matrix of observed data (dimensions nSensors * nTimepoints), **A** is the lead field matrix for the locations of interest (nSensors * nSources), and $\mathbf{S}^T$ is the estimated activity at each of the sources (nSources * nTimepoints). **E** is an error matrix of residuals. We record **M**, estimate **A** from the forward model, and have to model the source activity $\mathbf{S}^T$, typically with the aim of minimising the residuals **E** (Darvas et al., 2004). Of course, the number of source locations may vary across different implementations of the solution. More subtle approaches may also allow for dipoles to rotate in space (as well as simply change in magnitude) during the course of a response, for instance by having three orthogonal dipoles in each spatial location. Given that the number of sources may vary, and may exceed the number of sensors, it can easily be shown that the problem is ill-posed.

In order to solve this problem, we therefore have to impose some biologically plausible constraints on the form that the sources **A** and their activity **S** can take, in order to distinguish between the likelihood of the possible source reconstructions. There are three primary ways in which this can be done. First, we can assume only one or a very small number of dipoles are active ('equivalent current dipole' (ECD) approaches). Second, we can place some constraints on the distribution of activations over the cortical sheet (distributed approaches). Third, we can use spatial filtering techniques to minimise the contribution of other brain regions to the reconstructed

signal at a location of interest (beamformer approaches). Each of these approaches will now be examined in turn.

3.1.4.2 Equivalent current dipole approaches make the inverse problem well-posed by imposing a constraint on the number of dipoles

If we constrain our solution so that we only allow one equivalent current dipole (ECD), then the number of observations now dramatically outweighs the number of sources (i.e. the number of free parameters) in our source model, and so the inverse problem is no longer ill-posed. The aim is now to select the source location (and associated lead-field matrix **A**) and associated timecourse $\mathbf{S}^T$ that minimizes

$$\left\|\mathbf{M}-\mathbf{A}\mathbf{S}^T\right\|_2^2$$

Equation 3

and provides a least-squares fit to the observed data. The model has five free parameters for the dipole – three for spatial location (x, y, and z coordinates) and two for orientation (elevation, azimuth) – and a further free parameter for the dipole's magnitude. The cost function is non-convex (i.e. multiple local minima will exist), meaning that minimization is non-trivial. However, these problems can be overcome by initializing the search in a large number of possible locations or using sophisticated optimization techniques (Scherg, 1990, Darvas et al., 2004).

The problem remains well-posed if the number of free parameters in the model is less than the number of dimensions in the sensor data. ECD approaches can therefore typically be extended to capture the signal of up to a few dipoles simultaneously. Traditionally this might have proceeded by eye until no more dipoles were needed to reasonably match the observed scalp topography, but recently Bayesian approaches have been developed that infer the optimal number of dipoles from the data in a probabilistic fashion (Kiebel et al., 2008). The ECD approach works very well in cases where responses are well described by one (or a few) dipoles at particular brain locations. This is particularly the case when capturing early sensory responses (Hillyard, Teder-Salejarvi and Munte, 1998) or epileptic activity (Barkley and Baumgartner, 2003) but becomes less useful when attempting to describe widespread cortical activity, as might be evoked in a cognitive paradigm.

3.1.4.3 Distributed source imaging approaches make the inverse problem well-posed by imposed constraints on the spatial distribution of responses

Imaging ('distributed') approaches assume that the lead field matrix **A** is fixed to a particular number of sources distributed throughout the brain. These sources are normally generated by tessellating the cortical surface and placing dipoles across the cortical mesh (orthogonal to the mesh at each location in space). The number of sources (typically several thousand) vastly outweighs the number of sensors, and so the problem is ill-posed until some regularization is applied to possible solutions of the dipole timecourses, $\mathbf{S}^T$. Typically this regularization imposes some form of smoothness or sparseness on the solution, allowing the problem to be solved by constraining the spatial distribution of source activity.

The regularization can be imposed by adding an additional term to the cost function that is minimized during estimation of the dipolar responses:

$$\left\|\mathbf{M}-\mathbf{A}\mathbf{S}^T\right\|_2^2 + tr\left(\mathbf{S}\mathbf{C}_s^{-1}\mathbf{S}^T\right)$$

Equation 4

The form imposed by the regularization depends upon the structure of the covariance matrix. If $\mathbf{C}_s$ is an identity matrix scaled by a noise covariance matrix, $\lambda^{-1}\mathbf{I}$, then this is the 'Tikhonov regularized', or minimum norm, solution to the inverse problem (Tikhonov and Arsenin, 1977, Hämäläinen et al., 1993). By punishing solutions where the trace of the source covariance ($\mathbf{S}\mathbf{S}^T$) is high, at each timepoint any solution with a 'peaky' response (high variance across source amplitudes) will be punished, whereas solutions with a smooth response will be favoured. However, this also means that solutions with signal in superficial locations are favoured over solutions with deep sources, as a larger magnitude response (and associated increased variance) is required at depth to produce the same magnetic field strength (see section 3.1.3.3). Solutions for this have been proposed that increase the *a priori* variances attributable to deeper sources over superficial ones (Lin et al., 2006). Other regularizations include low-resolution electromagnetic tomography (LORETA), which uses a Laplacian operator to define $\mathbf{C}_s$ (Pascual-Marqui, Michel and Lehmann, 1994).

It is also possible to formulate the inverse problem in a Bayesian framework, in which priors are placed on hyperparameters to constrain the amplitude of responses allowable across small patches of cortex. Hierarchical Bayesian inference allows for inference on these hyperparameters from the observed data. This produces a data-

driven selection of a distributed or sparse solution, by squashing the hyperparameters to zero in locations where there is no evidence for signal in the data (Friston et al., 2008, Wipf and Nagarajan, 2009).

3.1.4.4 Beamformer approaches do not attempt to generatively model the data, but construct an optimal spatial filter at a location of interest

An alternative approach to source reconstruction is to use a beamformer analysis. Unlike ECD and imaging approaches, the beamformer does not attempt to generatively model the observed data and minimize the difference between the model and the observations. Instead, it adopts the assumption that no two macroscopic sources are correlated with one another (Hillebrand and Barnes, 2005). This principle can be used to design a spatial filter for each location in the brain that aims to achieve unity passband at the location of interest and zero contribution from all other locations. The two ingredients needed in designing the spatial filter are the covariance matrix of the data, $\mathbf{C}_b$ (nSensors * nSensors), and the lead field matrix **A** for the location of interest:

$$\mathbf{S}_\Theta = \mathbf{W}_\Theta^T \mathbf{M} = (\mathbf{A}_\Theta^T \mathbf{C}_b^{-1} \mathbf{A})^{-1} \mathbf{A}_\Theta^T \mathbf{C}_b^{-1} \mathbf{M}$$

Equation 5

W is a nSensors*nSources weights matrix for location(s) of interest θ (Hillebrand and Barnes, 2005), and acts as a linear combination of the underlying sensor data (Figure 10). By estimating the weights matrix for all locations throughout the brain, it is possible to reconstruct a whole-brain image. The beamformer approach was originally developed for applications in radar signal processing (van Veen and Buckley, 1988), and modified later for applications to M/EEG signal processing (van Veen et al., 1997). Several formulations of the beamformer have been proposed in subsequent years, but it has been shown that each formulation has the same underlying estimation of the weights matrix (Huang et al., 2004).

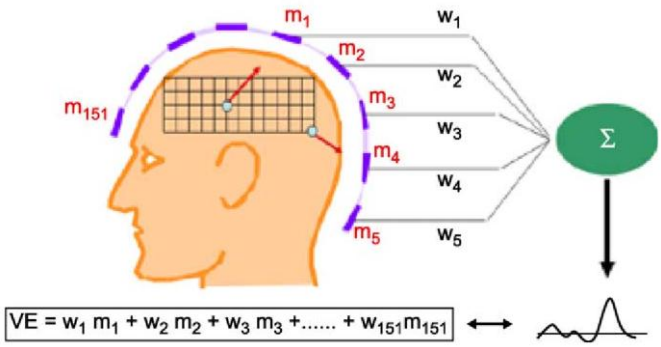

Figure 10. Beamformers reconstruct source activity via a linear combination (weighted sum) of sensor data. An MEG beamformer takes a series of measurements (m1,2,3…n) and computes a sum weighted by the corresponding sensor weights (w1,2,3…n) to spatially filter signals to the location of interest. Reprinted from (Hillebrand and Barnes, 2005).

The central challenge in constructing a beamformer is therefore accurate estimation of the data covariance matrix $\mathbf{C}_b$. It is typically estimated from a temporal window of data of finite length (Hillebrand and Barnes, 2005), but this can be provide a poor estimate if relatively short windows are used, as may be desirable when focusing on responses within a limited time-frequency window (Dalal et al., 2008). This limitation can be countered by applying regularization to the data covariance matrix, also referred to as diagonal loading. Recent approaches have attempted to infer the optimal degree of regularization from the data using a Bayesian approach (Woolrich et al., in press). It is also desirable to estimate the data covariance matrix only within the frequencies that are to be examined in the beamformed responses, as this provides optimal spatial filtering for the frequencies of interest. This can be achieved by temporally filtering the data prior to computation of the covariance matrix (Dalal et al., 2008).

Two caveats of the beamformer approach are that it has difficulties resolving distant sources that are highly correlated with one another, and that noise sensitivity typically increases with depth. The first limitation was noted when first applying the beamformer to M/EEG signal analysis, where it was shown that two distant sources with high correlation will produce a beamformer image with a source in between the two locations (van Veen et al., 1997). Importantly, this is no longer the case when there is only partial correlation between the sources, and simulation studies have shown that beamforming may be successful even in the case of high correlation if the period of high correlation is only transient (Hadjipapas et al., 2005). The second limitation can be overcome by applying noise normalisation to the resultant beamformer images; several approaches to this normalization have been proposed (Huang et al., 2004).

3.1.4.5. Summary

The inverse problem of MEG is ill-posed, but solutions can be found using appropriate biologically plausible constraints. The constraints used should depend upon the question asked. ECD approaches work well when trying to model activity in one or a few locations in the brain. Distributed and beamformer approaches are more useful when trying to capture widespread activation in a cognitive paradigm. Both approaches have certain conditions under which they will produce spurious results, and these limitations should be borne in mind when analyzing the data.

3.1.5 Basic principles of MEG analysis

3.1.5.1 Rejection and correction techniques are needed to remove artefactual data from MEG recordings

Although procedures such as performing recordings in a magnetically shielded room and minimizing sources of environmental noise will reduce the contribution of external artefacts to MEG recordings, it is inevitable that some artefactual signals will remain in the data. There are two principle ways of dealing with these signals - rejection and correction.

Rejection techniques are the most straightfoward, and typically the most reliable. They depend upon detecting trials where it is estimated that an artefact has occurred, and removing these trials from subsequent analysis. The key to reliable rejection is successful detection of artefactual signals, with both a low false negative *and* false positive rate, to retain as much artefact-free data as possible for subsequent processing. For ocular artefacts the most reliable approach involves placing an electrooculogram (EOG) electrode above and below the eye (and sometimes to the left and right of the eyes, to measure horizontal deflections). There is a natural electrical gradient within the eyeball (negative at the back, positive at the front), whose conduction is modulated when the eyelids move across the eyes; during an eyeblink, this causes an electrical deflection, with opposite polarity above and below the eyes (Luck, 2005). A simple peak-to-peak voltage threshold can be used to isolate eyeblink artefacts from the EOG, although the threshold may vary across subjects, so inspection of the EOG channel by eye is also helpful to confirm accurate detection. Similar criteria can also be built up for other artefacts, such as unwanted saccades or instrumental artefacts, to reject these from the data. One caveat is to make sure that brain responses (e.g. strong evoked components) are not incorrectly classified as artefactual, and that artefactual components are not more common on one experimental condition than another. Such

effects can sometimes be quite subtle, and go unnoticed in several years of published research (Yuval-Greenberg et al., 2008).

An alternative to rejection is to 'correct' artefactual segments of data, by attempting to isolate and remove the artefactual component whilst retaining the neural activity of interest. Such methods bear the advantage that maximal quantities of data are retained for subsequent signal processing, but the limitation that artefact correction is imperfect – the remaining signal may retain some of the artefactual component, or have some true neural activity removed. A popular approach to artefact correction is the use of independent component analysis (Delorme, Sejnowski and Makeig, 2007), which attempts to separate out independent sources contributing to the measured signal, using non-gaussianity as a measure of their independence. Often independent components can be isolated that match closely with the expected scalp topography (and component timecourse) of a particular artefact, and can simply be subtracted out of the data. In this thesis, a variant of this approach (using PCA of an 'idealised' artefact) was used and is discussed in section 3.1.6. Another approach to artefact correction is the use of spatial filtering techniques, to remove signals that are estimated as having originated from outside the head, based on the quasistatic approximation of Maxwell's equations. This Maxwell filtering (or 'MaxFilter') approach uses spherical harmonics to derive a basis set that describes signals that are likely to have been generated from within a sphere centred around the head, and signals that are likely to have been generated from outside this sphere (Figure 11) (Taulu and Simola, 2006). By removing the contributions of signals generated outside the sphere, effective artefact suppression can be obtained, including suppression of the nearby cardiac artefact.

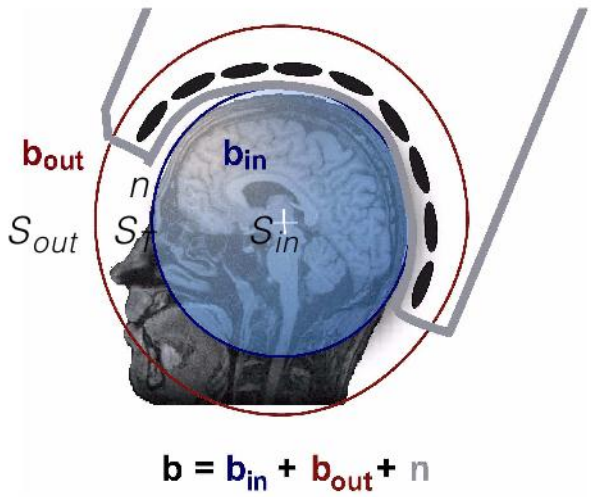

Figure 11. Schematic of regions affected by Maxwell filtering. Maxwell filtering ('signal space separation') uses properties of electromagnetic fields in order to estimate the contributions to the

magnetic field, b, of signals originating inside the sensor array, b(in), and signals originating outside this array, b(out). A further spatio-temporal extension of the approach can also estimate the contributions of artefactual signals generated very close to the sensors, n. Once these have been estimated, b(out) is retained; b(out) (and n, if estimated) are discarded. From MaxFilter user's guide, version 2.0.

3.1.5.2 Trial-to-trial variability in responses can be reduced using event-related averaging

The earliest EEG recordings of Hans Berger and MEG recordings of David Cohen focused on phenomena that could be seen in the raw data - such as the alpha rhythm present when the eyes are closed (Berger, 1929, Cohen, 1972). However, little can be seen relating to cognitive processing in the raw M/EEG signal, as the neural signals are swamped by noise – this includes both measurement noise, discussed above, and also unwanted trial-to-trial variability in cognitive state. The simplest approach to reducing this noise is to average across multiple repetitions of the same stimulus, with the assumption that the signal will be constant across repetitions but the noise will cancel, to generate an 'event-related potential' (ERP – or 'event-related field' (ERF) for MEG). The study of ERPs and the effects of cognitive manipulations has been the cornerstone of most EEG research for the past 50 years, and has given rise to a vast lexicon of isolated 'components' related to different cognitive processes (see (Kutas and Dale, 1997, Luck, 2005) for reviews). A component is sometimes assumed to refer to a particular peak in the ERP traces, but such a definition is limited as there are examples in which the same 'component' may generate peaks of different latency or even polarity in EEG recordings. The favoured definition of many EEG researchers is therefore with respect to a particular computational function being performed by a particular neuroanatomical module (Luck, 2005).

A key limitation of event-related averaging is that it may be insensitive to processes that vary in latency, or are not 'phase-locked', across trials. An example of this can be seen below (Figure 12). This has prompted researchers to analyse the effects of a known component in single-trial data, and see how the latency is affected by cognitive processing – indeed, this can provide an interesting chronometric insight into how long it takes for a computational process to be performed (Kutas, McCarthy and Donchin, 1977). However, this typically requires the component to be visible at the single-trial level, and so can only realistically be used for components with high signal-to-noise ratio; techniques such as ICA can help to minimize noise when extracting single-trial data. An alternative method to dealing with latency variability is to adopt a time-frequency approach to event-related averaging, which will also be sensitive to components that are not phase-locked across trials.

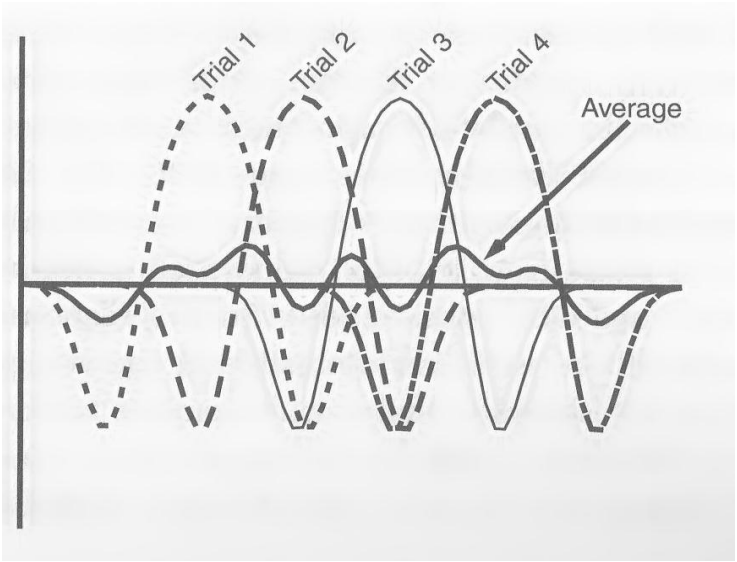

Figure 12. Latency variability can cause problems for event-related averages. Although the same component is present on four separate trials, the average of these trials produces an ERP that does not resemble the underlying component at all. Adapted from (Luck, 2005). Further examples are discussed in (Luck, 2005) and (Tallon-Baudry et al., 1997).

3.1.5.3 Time-frequency analysis allows for the detection of signals that are not phase-locked across trials

Fourier (Fourier, 1822) demonstrated that any function could be modeled exactly as an infinite trigonometric series – that is, a combination of sine waves of different frequencies. The Fourier transform allows us to convert a timeseries (such as MEG data) into a combination of sine waves of specific phase and frequency. Similarly, we can invert this transform to go back to the time domain from frequency domain. A frequency-domain representation of resting M/EEG data reveals that the power spectrum of the data obeys a power law or '*1/f*' distribution.

However, it is also possible to use frequency-domain representations to measure variation in this power spectrum as a function of time. This requires focusing our estimation of the spectrum on a short time window. The simplest approach to this is to divide our data into short windows and perform a short-time Fourier transform on each window. A trade-off must be drawn between how reliably power is estimated at each window, and the temporal resolution of the time-frequency decomposition; increasing the window size produces a better estimate of power (as it will include multiple oscillations) but will be influenced by points more distant in time. Moreover, the window required for reasonable estimate of a high frequency will be much smaller than that required for a low frequency.

An alternative approach is to use a 'wavelet' decomposition of the data, which attempts to measure the time-varying signal at one particular frequency of interest (reviewed in (Tallon-Baudry and Bertrand, 1999)). This is achieved by convolving the data with a 'wavelet', which is a windowed oscillatory function of a particular frequency. By repeating this process at multiple frequencies, a full time-frequency decomposition of the data can be obtained. Whilst a wavelet decomposition still suffers from the trade-off between window size (here 'wavelet factor') and temporal resolution, the wavelet adapts between different frequencies such that very high frequencies use only short time windows for their estimation, and lower frequencies are influenced by longer time windows. By performing a time-frequency decomposition of each trial, and averaging the changes in power through time, it is possible to detect non phase-locked responses that are missed by conventional ERP analyses.

3.1.6 MEG methodological considerations

In this section we consider several methodological choices that were made for studies in this thesis.

3.1.6.1. Adoption of a GLM approach to analysis of MEG data; orthogonalisation

The studies in this thesis investigate how neural activity covaries with several computational parameters of interest during value-guided choice. These computational parameters will frequently covary with each other, and so it is sometimes challenging to determine which parameter is encoded in the neural signal (Hunt, 2008). This is particularly the case if we adopt the traditional approach of examining the effect of one, or a few, variables on an evoked component of interest. We therefore adopted an approach that should allow us to examine which portion of the variance is explained by our parameter of interest, when controlling for the effects of all other variables. We can achieve this using a general linear model (GLM) to describe our data. The GLM approach was first introduced to functional MRI analysis in the 1990s (Friston et al., 1995). The principles of GLM-based analysis for M/EEG and statistical inference on the results, borrowed from the fMRI literature, have been proposed (Brookes et al., 2004, Kilner, Kiebel and Friston, 2005) but not yet widely adopted.

The GLM models a vector of observations, $\mathbf{y}$ (one observation per trial), as a linear combination of n regressors (or *explanatory variables* (EVs)), $\mathbf{x}_{1\ldots n}$, that vary from trial to trial:

$$\mathbf{y} = \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 \ldots + \beta_n \mathbf{x}_n + \varepsilon$$

Equation 6

where $\beta_{1\ldots n}$ are the *parameter estimates* for the regression and ε is a vector of *residuals*. We adopt a least squares solution – that is, we select the parameter estimates that minimize the sum of squares of the residuals.

If we have multiple observations per trial (e.g. multiple timepoints), we can write the GLM in matrix form as follows:

$$\mathbf{Y} = \mathbf{MX} + \mathbf{E}$$

Equation 7

where **Y** is a matrix containing the data (nTimepoints*nTrials), **M** is a matrix of unknown parameter estimates (nTimepoints*nEVs), **X** is a *'design matrix'* of explanatory variables (nEVs*nTrials) and **E** is an error matrix of residuals (nTimepoints*nTrials). The least squares solution to this equation is given by multiplying the data by the Moore-Penrose pseudoinverse of the design matrix:

$$\mathbf{M} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \mathbf{X}^+\mathbf{Y}$$

Equation 8

This yields an estimate of the effect size of each of our regressors of interest through time. Notably, the effect can be at any point in time, and does not have to be confined to the time of a particular peak in the M/EEG timecourse. Suitable design matrices can cater for the computation of more traditional averaging or difference waveforms, if desired. It is also worth noting that this general framework can be applied to data recorded at each sensor, or at each location in source space, or (by repeating over multiple frequency bands) for a time-frequency analysis.

Importantly, the parameter estimate controls for any shared variance between the regressors included in the design matrix; **M** is the effect size of each regressor in a subspace *orthogonal* to all other regressors. It is still important to try to minimize correlation between regressors of interest, however, as examining effects in this subspace is not the same as examining effects in the true space of the regressor, and can substantially affect the interpretation of the results. Although it is tempting to artificially orthogonalise one regressor with respect to another regressor, this procedure will not

affect the parameter estimates of the orthogonalized regressor, but instead assign any shared variance to the unorthogonalised regressors (Andrade et al., 1999, Hunt, 2008) – artificially removing the main advantage of the GLM approach. This is therefore a practice normally best avoided in functional imaging studies.

3.1.6.2. Choice of source reconstruction technique

As most of the phenomena of interest in this thesis are late cognitive phenomena related to choice (as opposed to early sensory phenomena), we first adopted a distributed approach to source reconstruction rather than a simple ECD approach. We used the multiple sparse priors (MSP) approach to distributed source reconstruction implemented in the SPM toolbox (Friston et al., 2008). Whilst this sometimes yielded successful reconstruction of early evoked potentials, we wanted to use the parameter effects of interest (computed in sensor space) to constain our inverse solution, as these were the effects that had previously been localized to brain regions of interest. (This is equivalent to computing a difference waveform in sensor space first, and then reconstructing the difference into source space (a 'localisation of differences'), rather than reconstructing the main effect and then computing the difference within source space (a 'difference of localizations') (Henson et al., 2007)). With this approach we frequently found that strong priors were placed in the medial inferior temporal lobes, and nowhere else in the brain (Figure 13) – a region that has not traditionally been strongly implicated in reward-guided decision making. Similar results have been obtained from passing noise through this source reconstruction routine (Rik Henson, personal communication), and so it appears that the MSP may not be sufficiently powerful to infer source locations based upon comparatively weak sensor-level effects.

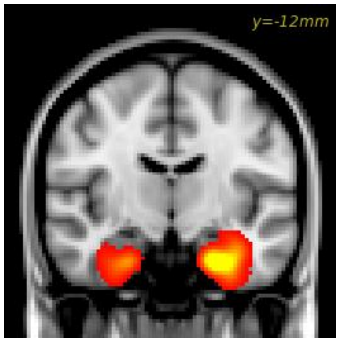

Figure 13. Artefactual placement of priors in the anterior temporal lobes by the multiple sparse priors (MSP) source reconstruction approach. Such effects may result from the relatively low signal-to-noise ratio in a value-based contrast used to constrain the source reconstruction. Similar results could be obtained by passing noise through the source reconstruction algorithm.

We therefore switched to a beamforming source reconstruction approach. This yielded successful reconstruction of distributed activity in anatomically plausible locations, and could successfully localize value-related MEG activity (see **chapter 5**). A drawback of this approach is that fusion of multiple modalities is not yet as well developed as for the MSP approach (Henson, Mouchlianitis and Friston, 2009); thus, although in all our experiments we recorded simultaneous EEG to improve localization of deep sources, we were unable to use this data in our source reconstructions. Approaches to beamformer M/EEG fusion are currently being investigated (Woolrich et al., in preparation).

3.1.6.3. Methods for rejecting and suppressing physiological artifacts

As noted in section 3.1.2.1, ocular artefacts are a major source of noise in M/EEG recordings, and the sources of these artefacts are within the sphere that is retained using the MaxFilter technique described in section 3.1.5.1, so we need to consider alternative methods for their elimination. This is particularly important in studies of reward-guided decision making, as the eyes are very close to two important structures associated with reward and reinforcement, the orbitofrontal and ventromedial prefrontal cortex. Alarmingly, we found that when MEG data was timelocked to saccade or eyeblink onset, the resultant scalp topography could often be well described by a single ECD placed in ventromedial prefrontal cortex.

One method of removing ocular artefacts from M/EEG data is to design a spatial topography that describes the artefact well, and can be regressed out of the data whilst leaving brain sources intact. This is similar to what is done when ICA is used for correction. ICA can be problematic, however as it requires a large amount of data (and computer power) to estimate the independent components present in the data, can be unreliable (with artefacts split between multiple components), and requires manual identification of artefactual components.

We therefore adopted an approach in which we attempted to construct an idealized spatial topography based on *known* artefactual components, and regress this spatial topography out of the entire dataset. We first detect artefacts using some criterion (e.g. eyeblinks using vertical EOG), and average these to create a 'template artefact' timecourse. This is then submitted to principal component analysis (PCA), which identifies spatial topographies that explain maximal variance within the template. Principal components with artefact-like spatial topographies and timecourses are retained as spatial confound regressors. The contribution of each of the spatial confound regressors to each timepoint in the *raw* data is then estimated and regressed out of the

data. This approach has the advantage that the sensor topography is based upon *detected* artefacts, but removes the artefact from *all* the data – i.e. it does not depend upon reliable detection of every artefactual component. The approach is similar to that advocated by (Berg and Scherg, 1994), without the inclusion of brain sources as co-regressors of no interest. We found it to be a powerful and reliable way of reducing ocular artefacts in sensor data (Figure 14).

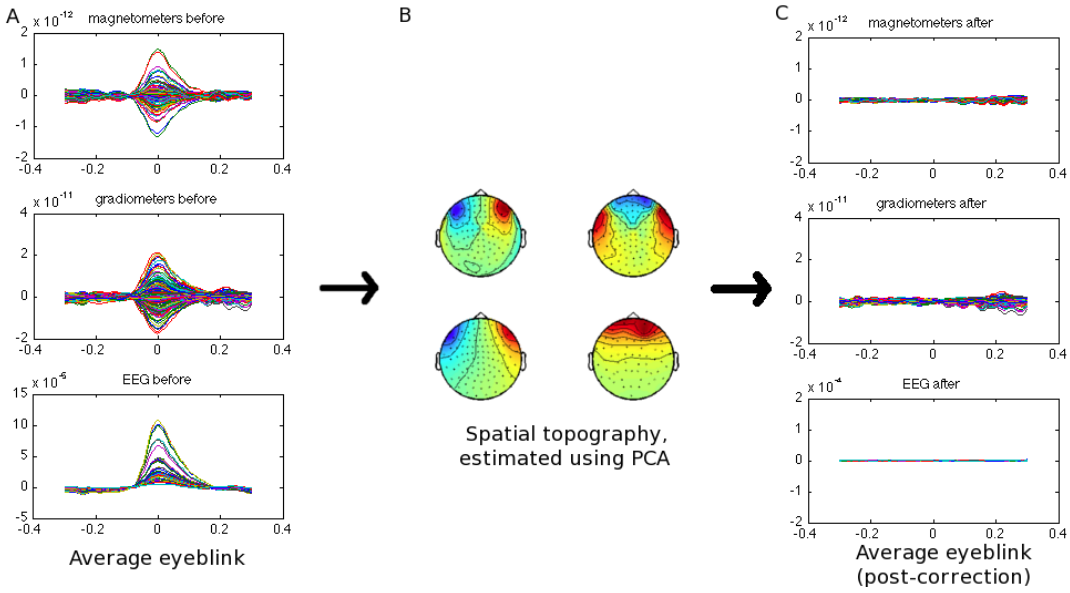

Figure 14. Eyeblink correction procedure developed as part of this thesis. A: Eyeblinks are detected in the EOG sensor, and used to generate a template 'average' eyeblink. B: The average eyeblink is submitted to principle components analysis (PCA), and eyeblink spatial topographies are regressed out of the *raw* data. Importantly, this means that the method does not require every eyeblink to be reliably detected in the EOG channel. C: Same data as in A, after PCA-based correction. The eyeblink is no longer visible (despite this data being timelocked to blinks in the EOG channel).

3.2 Functional magnetic resonance imaging

The second neuroimaging technique used in this thesis is blood oxygen level dependent (BOLD) functional magnetic resonance imaging (fMRI). As far less work went into the development of appropriate methodology for fMRI acquisition and analysis for this thesis, and we instead used relatively standard acquisition protocols and analysis pipelines, I will not discuss fMRI acquisition and analysis in the same level of detail as for MEG. There are many good reviews available of these techniques (Jezzard, Smith and Matthews, 2003, Smith, 2004, Smith et al., 2004). However, I briefly review some of the key points of fMRI acquisition, the basis of the BOLD signal, and fMRI preprocessing and analysis. In addition, I provide some methodological considerations that are specific to the studies in this thesis.

3.2.1 What are we measuring with fMRI, and how is it measured?

Placing the human head inside a strong static magnetic field, such as a 1.5-7T electromagnet, will cause the magnetic moment of hydrogen nuclei found in water to align with this magnetic field (referred to as the 'B0' field). The axis along which the magnetic moments align is the '*z*-axis', and at a macroscopic scale can be represented by a vector that aligns with this axis. Hydrogen protons process around the *z*-axis at a particular frequency (much like a spinning top), termed the Larmor frequency. The Larmor frequency for hydrogen protons is different to that found in other nucleons (~128MHz in a 3T magnet). Hydrogen protons are abundant in water, which is present in different concentrations in different tissues of the body.

By transiently applying radio frequency pulses to the protons at the Larmor frequency for hydrogen, the nuclei are 'flipped' out of the *z*-direction into an *x*-*y* plane. Oscillations around this X-Y plane generate an electromagnetic signal, which can then be detected in a receiver coil placed around the head. The key feature determining the strength of the measured signal is the rate of decay back towards the *z*-axis, along which the protons process when no RF pulse has been applied. Two constants determine the rate of this decay: T1 determines the rate at which protons regain magnetization in the *z*-direction (typically on the timescale of seconds), and T2 determines the rate at which magnetization is lost in the x-y plane (typically on the timescale of milliseconds). These are different for distinct tissue types (white matter, grey matter, cerebrospinal fluid), and can be used to generate images of different contrast, which make use of these different constants.

A further constant, T2*, also influences the rate of decay of transverse magnetization, but this is dependent upon local field inhomogeneities, causing dephasing of the spins of nearby protons. T2* is shorter than T2, and is sensitive to the ratio of oxy:deoxy-haemoglobin in tissue. Specifically, deoxyhaemoglobin is more paramagnetic than oxyhaemoglobin (Pauling and Coryell, 1936, Ogawa et al., 1990), and so the signal loss due to T2* is reduced when the oxy:deoxy-haemoglobin ratio decreases. This is important, as it allows us to measure a quantity that is dependent upon local blood flow. Blood flow is known to reflect a change in neural activity, as discussed below.

Information about a particular spatial location is obtained by adding 'gradient coils' (mT in strength) in *x*, *y* and *z* directions, that interact with the much stronger (1.5-7T) stable magnetic field. As these additional 'gradient coils' affect the Larmor frequency at specific spatial locations, this information can be used to gain spatial information about the obtained signal: that is, the RF pulse can be matched to the Larmor frequency

of a particular spatial location. Sampling across different spatial locations, using different RF frequency pulses, allows for sampling across the entire brain.

The image generated during this sampling process depends upon several factors. Firstly, it depends upon how long a gap is left between the initial flip application and the measurement of the electromagnetic signal (the 'echo time', or TE). Secondly, how long is left before another flip application occurs (the 'repetition time', or TR). Thirdly, the 'flip angle', i.e. the degree to which the protons are flipped out of alignment with the *z*-axis. Finally, the strength of the RF pulse during flip application (the 'B_1 field'). Different combinations of these parameters will create different images. The images have contrast sensitive to different tissue types, or to different features of the tissue. For example, collecting the signal immediately (low TE) but with a long delay between acquisitions (high TR) will give a measure of the proton density of the voxel of interest. More complex parameters can be used to produce an image whose contrast is sensitive to T2*, and thus is blood oxygen level dependent (BOLD).

The relationship between neural activity and the BOLD signal is also a complex one, as net activity in a region will not necessarily depend upon the firing rate of cells within a cortical microcircuit in a simple fashion. The key problem is that cortical microcircuits are known to include inhibitory as well as excitatory cells, and it is unclear which of these makes the strongest contribution to determine changes in blood flow (Logothetis, 2008). Nevertheless, in the classical model, an 'increase' in neural activity (which could be an increase in inhibitory, as well as excitatory, firing rates) elicits a spatially discrete increase in blood-flow which outweighs the increase in oxygen demand of a tissue. Thus, within the timecourse of 4-6 seconds of an increase in neural activity, there is an *increase* in the ratio of oxy:deoxy-haemoglobin. This produces a decrease in magnetic susceptibility (T2*-dependent signal) and so produces a positive BOLD response. This is captured during analysis of the data by convolving the predicted neural activity with a 'haemodynamic response function' (see figure 15).

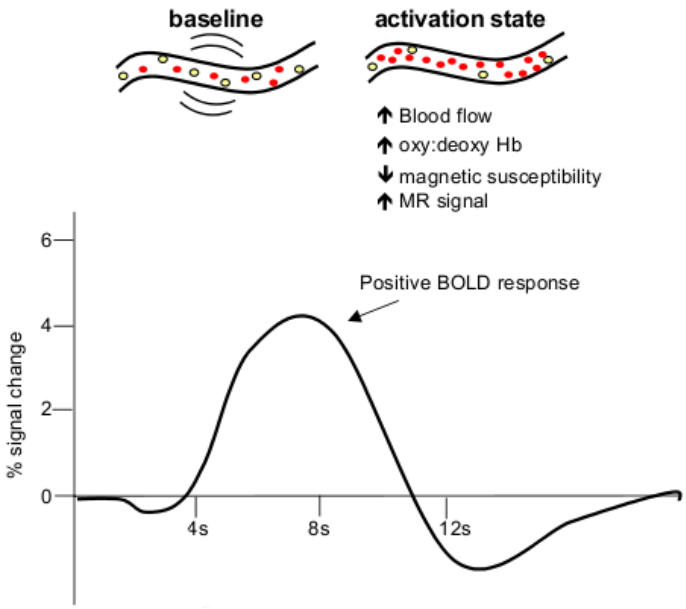

Figure 15. BOLD response to neural activity. In the baseline state (top left), the magnetic field is susceptible to distortion in the area surrounding the blood vessel. During activation (top right), there is an increase in blood flow that outweighs the demand for oxygen consumption, causing an increase in the ratio of oxy:deoxy-haemoglobin. This produces a decrease in the field distortion, and so an increase in MR signal. This signal change takes several seconds to occur after the initial activation (bottom graph). Adapted from (Johansen-Berg, unpublished doctoral dissertation, 2001)

3.2.2 Basic principles of fMRI preprocessing and analysis

In a typical fMRI experiment, several hundred volumes of T2*-weighted images are acquired at a sampling rate of around 0.33-1 Hz. The data produced are 4-dimensional – they have an *x*, *y*, and *z*-coordinate (3D 'voxels') and have been acquired at a particular timepoint in the experiment. Each slice of each image acquired is taken at a slightly different point in time. Typically, 'slice timing' correction is carried out (via temporal interpolation) to correct for this, to make subsequent statistical modeling more straightforward. It is then typical to correct for motion-induced artifacts by registering each volume to all other volumes in the timeseries. The data are then spatially smoothed to improve the signal:noise ratio. Finally, it is common to grand-mean normalize the data, to account for changes in intensity across subjects, and also to filter each voxel's timeseries prior to statistical modeling, to account for slow changes in intensity across time (Smith, 2004).

After preprocessing, statistical modeling of the fMRI timeseries is used to find which brain regions are statistically 'activated' (that is, where signal is increased) or 'deactivated') (where signal is decreased) during the course of the experiment. Although there are many techniques popularly used for analysis of fMRI timeseries, such as multivariate and model-free approaches, the most common approach remains a *mass-*

univariate approach. Here, the 1D timeseries at each voxel, *y*, is modeled independently using the general linear model:

$$y(t) = \beta_0 + \beta_1 x_1(t) + \beta_2 x_2(t) + \ldots + \beta_n x_n(t) + \varepsilon$$

Equation 9

where β_n is the parameter estimate for the explanatory variable *x*, which takes a different value at each point in time. Notice the similarity between this equation and equation 6; the general linear model is a very flexible means of timeseries analysis.

In fMRI, the traditional way of designing an experiment would be a subtraction between multiple blocks, in which a component process was varied. Each block would be captured in the design matrix *x* using a 1 during the block, and a 0 at all other times. By contrasting different blocks, a brain region involved in a component process might be isolated. However, such a technique relies upon the principle of 'pure insertion' – that is, that adding an extra component to the task will not affect the performance of tasks during the 'baseline' period (Friston et al., 1996). A more sophisticated approach is to perform 'event-related' analysis, and look for signals which vary across trials whilst still modeling the baseline of the event taking place, in a separate explanatory variable.

The parameter estimates from this first-level analysis are then submitted to a group-analysis, in which maps from multiple subjects are combined into a single statistical parametric map. This is the approach we adopt in the computational modeling analysis presented in **chapter 6**. Once parameter estimates for individual subjects have been obtained, it is typically desirable to make inference not about the response for an individual, but instead about the population as a whole. This can be done using multi-subject statistics. The approach adopted is exactly the same as in equation 9, but the parameter estimates $\beta_{1\ldots n}$ for each subject become the timeseries *y*, and the design matrix can include a group mean (i.e. β_0) as well as additional explanatory variables for cross-subject variation in first-level parameter estimates.

However, before any of this can proceed, there must be registration of the first-level statistical maps to a common template. This is typically done in two stages. Firstly, the T2*-weighted BOLD image is registered to a T1-weighted structural scan of the subject. Secondly, this T1 scan is registered to a 'standard' brain template, of which several exist. By performing these registrations accurately, it is possible to align statistical images from subjects of different head and brain shapes, and infer (using a common coordinate system) where in the brain activation has occurred. A standardized

coordinate system allows for alignment across subjects, and across studies, in order to allow for group level analysis using a second-level general linear model.

Once this second-level general linear model has been estimated across the population of subjects, it is typical to perform some statistical inference on the parameter estimates. Normally, when examining the population effect in a group of healthy participants, this will inferring whether the distribution of observed β_n values are significantly different from zero, using a 2-tailed T-test. The observed T-score (or corresponding Z-score) is calculated at every voxel in the brain, and a statistical threshold is used to obtain the probability of observing this value under the null hypothesis of no effect.

An important consideration here is that there are many thousands of voxels at which this statistic is computed, and so it is necessary to correct for the number of statistical tests performed; this is termed the 'multiple comparisons problem'. Instead of controlling for the *type I* error rate, which only takes into account one observation, we instead want to control for the *family-wise error (FWE) rate*, which takes into account the number of observations. Because there is inherent spatial smoothness in the data, there are several approaches that are more sensitive than the simplest approach of Bonferroni correcting for multiple comparisons. These include resel-based correction and cluster-based correction, and require some quite complex mathematics dependent upon Gaussian random field theory, as reviewed in (Worsley et al., 1996, Hayasaka and Nichols, 2003).

3.2.3 fMRI methodological considerations

3.2.3.1 Decorrelation of regressors; orthogonalisation

The general linear model tests for the effects of each of the explanatory variables on the acquired fMRI timeseries. Importantly, it does this in a subspace that is orthogonal to all the other explanatory variables included in the design matrix. This can lead to ambiguous results in cases where there is a high degree of correlation between regressors (Andrade et al., 1999). One possible solution is to orthogonalise the regressors with respect to one another, but this approach can lead to problems itself, as described above in section 3.1.6.1 (Hare et al., 2008, Hunt, 2008). The best approach is to initially design the experiment such that the explanatory variables (regressors or independent variables) are intrinsically decorrelated from one another. This becomes particularly important in studies using a computational model with several different features of the model that might be used in a general linear model-based analysis.

3.2.3.2 BOLD acquisition sequences for the orbitofrontal cortex

Because of local magnetic field inhomogeneities caused by the presence of air-water boundaries induced by sinuses in the head, certain brain regions are susceptible to artifactual 'signal dropout' during echo-planar fMRI. One area of particular susceptibility is in the orbitofrontal cortex (OFC), which, for reasons outlined in **chapter 1**, is of particular interest when studying reward-guided decision making. Several previous solutions have depended upon only collecting data from a region of interest encompassing the OFC, but Deichmann and colleagues presented a novel solution to this problem which allowed for whole-brain imaging whilst also avoiding susceptibility artifacts. This involved tilting the slice acquisition angle such that it was not along one of the standard axes used to acquire data in other studies, and adding an additional preparatory pulse to further suppress the artefactual dropout (Deichmann et al., 2003). We adopted this imaging protocol during our fMRI study, as it allows for whole-brain imaging whilst providing strong signal from OFC.

3.2.3.3 Temporal separability of decision- and feedback-related activation

As outlined in **chapter 1**, in this thesis I adopt a component process account of decision making that involves separate computations at the time of making a decision and at the time of receiving feedback about the decision. By adopting an event-related design we were able to separate out these two components of decision making and learning temporally. This required leaving a sufficiently long gap between the onset of the decision and feedback presentation, and jittering the length of information presentation onscreen. Full details of the timing of the event-related design are given in **chapter 4**.

Chapter 4: Estimating subjective values in paradigms of value-guided choice

*One of the primary aims of this thesis is to investigate which neural structures support value-guided choice in the human brain. To do this, I make use of the imaging techniques introduced in **chapter 3**, the mathematical models of decision and learning introduced in **chapter 2**, in order to characterize neural activity in the structures introduced in **chapter 1**. In this chapter, I introduce two new paradigms of value-guided choice appropriate for human subjects. I describe the subjective values used by participants in these tasks. These subjective values inform the predictions of the mathematical models and neural activity investigated in subsequent chapters.*

One of the most common means of investigating candidate models of decision making has been to use tasks involving a perceptual discrimination. Examples of perceptual discrimination tasks are widespread, from those in the visual domain such as the random dot stereogram task discussed in **chapter 2**, to examples in the somatosensory (Mountcastle, Steinmetz and Romo, 1990), auditory (Lemus, Hernandez and Romo, 2009) and olfactory (Uchida and Mainen, 2003) domains. Perhaps one of the main attractions of perceptual tasks is that the stimulus – and so the input to any model of the decision process – is placed under the control of the experimenter, and so can be accurately estimated and carefully manipulated.

Many real-world decisions, however, do not depend upon objective features of the environment, but instead upon some *subjective* quantity that is *internally* generated by the organism. Consider a decision between a healthy sandwich and a bag of jelly babies. A child who has been fed a large bag of sweets before this choice is more likely to choose the sandwich than a child who has not; he has become selectively satiated on the sweets. To the outside observer, however, these children appear identical to one another. The key difference, then, between these two children, lies in the *subjective value* of the sandwich and the jelly babies to each child. To interpret neural activity that is measured during value-guided choice tasks, we therefore need to estimate (as accurately as we can) the subjective values associated with each decision option presented in the task. By observing the choices made by subjects at each decision, we can reveal their preferences for one option over another, and use this information to constrain models of the subjective value of each option.

Recent investigations have highlighted that neural activity recorded during value-guided choice typically reflects the subjective, rather than the objective, value of

decision options. As discussed in **chapter 1**, neural activity in the lateral intraparietal cortex (LIP) reflects both the probability and magnitude of receiving a reward after making a saccade into the neuron's response field (Platt and Glimcher, 1999). However, it is possible to engineer a situation in which the probability of receiving reward varies independently of the subjective desirability of making a saccade. In this case, LIP firing rates track the subjective desirability rather than the objective reward probability (Dorris and Glimcher, 2004). As discussed in **chapter 2**, investigations using fMRI have also found that neural responses more closely reflect the subjective expected value of the presented options, rather than their objective value (Kable and Glimcher, 2007, Hsu et al., 2009).

It therefore becomes important to accurately estimate the subjective values of participants performing value-guided decisions, in order to derive regressors that are likely to best reflect neural activity. This chapter introduces the behavioural tasks used in subsequent chapters of this thesis, and describes some of the models that can be used to estimate the subjective value of different options. In the first section, I introduce a simple economic choice paradigm that is used in the MEG study in **chapter 5**; in the second section, I introduce a social learning paradigm that is used in the fMRI study in **chapter 6**.

4.1 Can Prospect theory successfully model behaviour during a choice task involving multiple trials, and learning from experience?

4.1.1 Introduction

A highly influential theory of how humans decide between options with different probabilities and magnitudes of reward is *expected utility* theory (von Neumann and Morgenstern, 1944). This proposes that subjects estimate the utility associated with a given reward magnitude, and that this relationship may be non-linear (concave). They then multiply this utility by the true probability of reward to obtain the subjective desirability of that option. The convex expected utility curve explains the diminishing marginal return associated with larger rewards – that is, the difference between a £0 and £1000 reward looms larger than the difference between a £50000 and £51000 reward. However, several simple demonstrations have shown that this theory falls short of explaining all economic choices (Allais, 1953, Kahneman, Knetsch and Thaler, 1991). These limitations led to the development of *Prospect theory*, in some ways an extension of expected utility theory, which proposes: (i) that the *probability weighting* function is also non-linear, with small probabilities in particular being overweighted; (ii) that the

utility function is steeper for losses than for gains; and (iii) that a 'gain' or 'loss' is measured relative to some reference point (Kahneman and Tversky, 1979, Tversky and Kahneman, 1992). These proposals explain many of the observed deviations from expected utility theory, and can also explain numerous 'real world' decisions in which human behaviour deviates from rational behaviour (Camerer, 2000).

However, Prospect theory was developed on the basis of single-shot responses made in the absence of any feedback. By contrast, neural recordings require multiple presentations of the same problem, as many trials are needed to obtain reliable estimates of neural activity. Subjects will also typically receive feedback on every trial. How well does Prospect theory extend to these circumstances? It appears that some of its predictions may be violated. One key prediction of Prospect theory, for instance, is that subjects *overweight* low probabilities of events occurring. By contrast, it has recently been shown that when the probabilities are learnt by *experience* (as opposed to being learnt from *description*), subjects *underweight* low probabilities (Barron and Erev, 2003, Hertwig et al., 2004, Hertwig and Erev, 2009). Explanations of this phenomenon often appeal to the statistical undersampling of low probability events (Fox and Hadar, 2006).

Importantly, this effect may vary depending upon the *kind* of experience that the subjects can gain (figure 1). In one study, subjects learnt the probabilities by exploring each option independently in the absence of any choice (Hertwig et al., 2004). A similar underweighting effect was also observed if the options were explored whilst *choosing* repeatedly, with 'partial' feedback delivered on the chosen option only (Barron and Erev, 2003). If, however, feedback was given about *all* options irrespective of which option was chosen (a 'full-feedback' choice paradigm), then decisions revert to being similar to a 'description'-like *over*weighting of low probabilities (Hertwig and Erev, 2009). Thus, the distinction between 'decisions from experience' and 'decisions from description' may depend upon the nature of feedback that subjects receive.

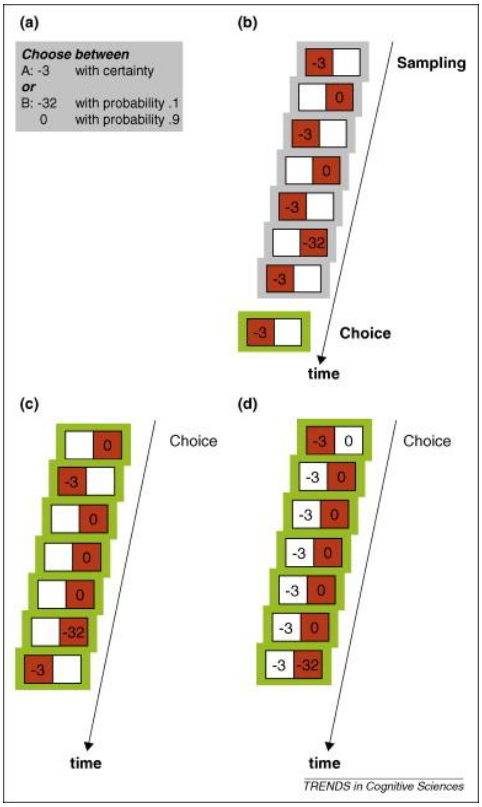

Figure 1. Distinct learning paradigms in 'decisions from experience'. (A) The overall expected value of the two prospects is identical on all three paradigms. (B) 'Sampling' paradigm – subjects are free to sample observations from each prospect, and then make a single choice is made after sampling. (C) 'Partial feedback' paradigm – subjects choose repeatedly, and only receive feedback on the chosen option. (D) 'Full feedback' paradigm – subjects choose repeatedly, and receive feedback on both chosen and unchosen options. From Hertwig and Erev, 2009

Perhaps most interesting is the situation in which *both* description and experience are available. This might be the case, say, for a doctor who decides both on the basis of summary statistics available in the literature and also on the basis of his own experience with clinical populations. Recent studies have investigated such situations and again highlighted the importance of the feedback delivered on influencing subject behaviour. Jessup and colleagues investigated decisions under risk where a description was always available, and compared conditions where feedback was either present or absent (but only on the chosen option) (Jessup, Bishara and Busemeyer, 2008). When feedback was absent, subjects persistently overweighted low probabilities, consistent with Prospect theory; but when the partial feedback was presented, subject behaviour moved closer to being normative. Newell and Rakow also found that feedback pushes subjects towards a more normative set of responses, and that this transition occurs gradually as more feedback is presented, but this was in a task in which the normative strategy was relatively transparent (Newell and Rakow, 2007). It is still unclear, therefore, whether Prospect theory is useful in describing choice behaviour where more difficult decisions are presented, and where full feedback is delivered. This

becomes important when trying to understand the subjective values used by subjects in many neuroimaging studies.

In this section, I investigate whether Prospect theory models are still useful under these conditions. The task used involves full feedback and description, with multiple trial repetitions. I also characterise more basic properties of choice behaviour in this paradigm that are important for subsequent analyses of the neural data. These include whether subjective value functions change through time, whether they are affected by *recent* experience, how subjects integrate across multiple dimensions in guiding their choices, and what effect value has on subject reaction times.

4.1.2 Methods

4.1.2.1 Experimental task

30 subjects repeatedly chose between two risky prospects to obtain monetary reward. Stimuli comprised a rectangular bar, whose width determined the amount of reward available, and a number presented underneath the bar, whose value determined the probability of receiving reward on that option (figure 2). The probabilities of winning on each option were independent; thus, on any given trial, both, neither or either option(s) might yield reward. Stimuli were drawn such that reward magnitude and probability were never identical across the two options; subjects therefore needed to integrate across stimulus dimensions to make optimal choices (see below). On some trials, however, both probability and magnitude were larger on one side than the other, a decision we classify as a 'no brainer' trial. By design, mean correlation between overall value and value difference (chosen-unchosen value) was kept at 0.31±0.08 (mean±s.d.), allowing them to explain largely separate portions of variance in behavioral and neural data.

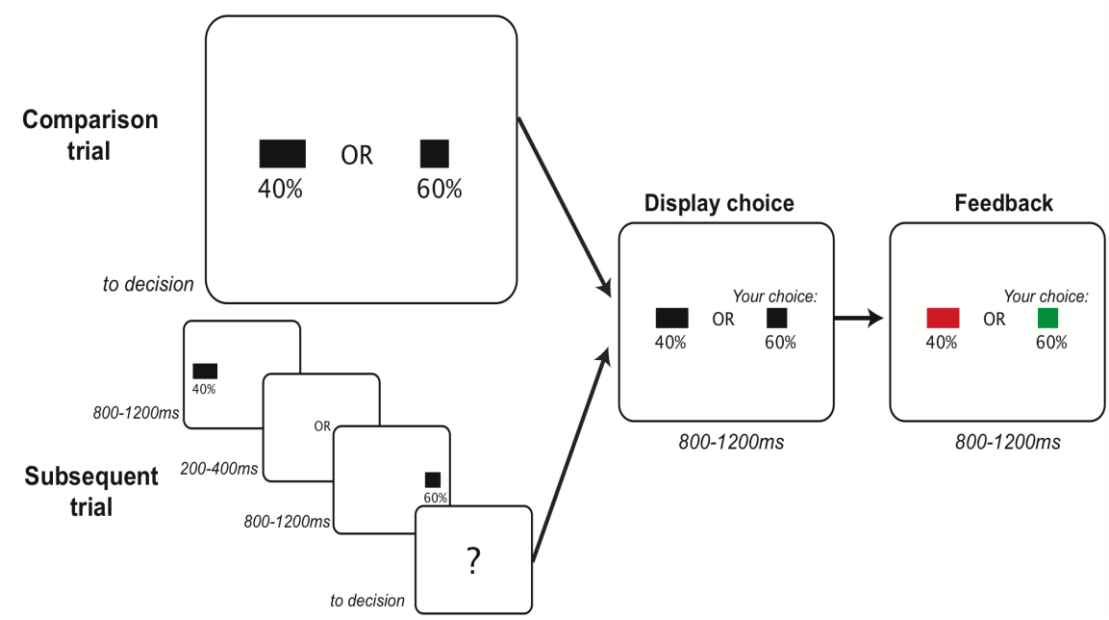

Figure 2. Experimental task (for experiment 1). Subjects chose between two options, each with different reward probability and magnitude, to receive monetary reward (money bar not shown). All analyses presented in this chapter relate to 'comparison trials', in which both options appeared simultaneously; subjects also completed interleaved 'subsequent' trials, in which options were presented sequentially rather than simultaneously.

Decisions were presented onscreen until a response was made. After selection, the chosen option was highlighted for 800-1200ms jittered, and outcomes were presented for 800-1200ms jittered. Feedback was presented on both chosen and unchosen options by turning a rewarded option green, and an unrewarded option red. Stimuli were then removed, and an ITI of 500-800ms was presented.

On choosing a rewarded option, a 'winnings bar' displayed at the bottom of the screen increased in magnitude in proportion to the width of the chosen option. When this winnings bar reached a gold target on the far right of the screen, £2 was added to subjects' earnings, and the winnings bar reset itself to its original size. Total typical earnings for the task ranged from £26 to £34.

We collected a secondary dataset with high-resolution eyetracking ('experiment 2', below) to exclude ocular artifacts as a possible confound to MEG signals recorded during the task. No major differences in task-related activity were seen, so MEG data was collapsed across the two experiments (see **chapter 5**). Minor task differences between the two experiments were as follows:

Experiment 1. 18 subjects participated in experiment 1 (age range 21-33, 10M, 8F). Each subject completed 324 trials. Rewards were drawn from the following set: [2,5,8,11,14,17] (width of reward bar in pixels). Probabilities were drawn from the following set: [10,25,40,60,75,90] (probability of reward in %). An additional 324 trials were interleaved in which options were presented subsequently rather than

simultaneously, but are not discussed in this chapter. Stimuli were presented on either side of a fixation point; subjects selected the left option with a left-thumb button press, and the right option with a right-thumb buttonpress. Trials were presented in 12 blocks of 54 trials (~5 minutes), yielding a total experiment time of ~1 hour.

Experiment 2. 12 subjects participated in experiment 2 (age range 21-35, 4M, 8F). Each subject completed 200 trials. Rewards were drawn from a uniform distribution from 1 to 18 (width of reward bar in pixels). Probabilities were drawn the following set: [10,15,20,25,30,35,40,45,50,55,60,65,70,75,80,85,90] (probability of reward in %). An additional 200 trials were interleaved in which 3 options were presented, but are not discussed in this chapter. To restrict saccades during the time of making the choice, each trial was preceded by a fixation point with the word 'Fixate!' presented immediately underneath, and subjects self-initiated each trial when ready using the response pad. After self-initiation, the word 'Fixate!' disappeared and a 500-800ms pre-stimulus period in which only a fixation cross remained was presented, prior to stimulus presentation. Eye movements during the trial were minimized by presenting stimuli immediately adjacent to the fixation point, and tracked using a high resolution eyetracker. Stimuli were randomly distributed around two of three possible response locations, which subjects selected using right index, middle and ring fingers. Trials were presented in 8 blocks of 50 trials, yielding a total experiment time of ~40 minutes.

In both experiments, stimuli were presented on a screen situated 1.5 meters away from the subject, inside the magnetically shielded room; stimuli were displayed via projector (refresh rate 60Hz) situated outside the room. Stimulus presentation and timing was controlled using Presentation software (Neurobehavioral Systems, Albany, CA).

All subjects provided informed consent in accordance with local ethical guidelines.

4.1.2.2 Behavioural analysis

Fitting of subjective value functions. Subjective utility and probability weighting functions were derived from Prospect Theory (Tversky and Kahneman, 1992), and were of the following form:

$$v(r_o) = r_o^{\alpha}$$

Equation 1

$$w(p_o) = \frac{{p_o}^{\gamma}}{({p_o}^{\gamma} + (1-p_o)^{\gamma})^{\frac{1}{\gamma}}}$$

Equation 2

where r_o and p_o are the reward magnitude and probability of gaining reward, respectively, on outcome o. The subjective expected value of outcome o was calculated as:

$$sEV_o = v(r_o) * w(p_o)$$

Equation 3

The probability of choosing each option was then calculated using a softmax choice rule (Sutton and Barto, 1998):

$$P(C = o) = \frac{e^{\frac{sEV_o}{\tau}}}{\sum_{i=1}^{n} e^{\frac{sEV_i}{\tau}}}$$

Equation 4

where n is the number of options (2 for this study) and τ is a temperature parameter that determines the stochasticity of action selection. Values of α, γ, and τ were fit by maximizing the likelihood of each subject's choices in the experiment, using non-linear fitting routines in MATLAB (The Mathworks, Natick, MA).

Analysis of choice data. We used logistic regression to investigate the influence of probability, magnitude and value of each option on the probability of choosing option 1. We normalised each variable before entry into the logistic regression (to ensure that parameter estimates were comparable across the different variables), and included a constant term to model any bias towards choosing one option over the other. Logistic regression is similar in spirit to linear regression, with the aim of performing a least squares fit to the following function:

$$p(C = o) = \frac{1}{1 + e^{-\beta \hat{x} V_T}}$$

Equation 5

where $\hat{\beta}$ are an array of parameter estimates associated with the array of independent variables $\vec{x}$. We performed a one-sample T-test across subjects for each parameter estimate, to infer which variables had a significant effect on choice behaviour.

Analysis of reaction time data. We examined the effects of value, trial number and 'no brainer' trials on each subject's reaction time data using multiple linear regression. We entered log(RT) as the dependent variable, as it has a distribution that is closer to normal than RT. In one analysis, we entered the following regressors as independent variables: (i) A constant (to model mean RT); (ii) a 'bias' term to capture speeding of RTs for left vs. right choices; (iii) the difference in subjective values between chosen and unchosen options (sEV_{chosen}-$sEV_{unchosen}$); (iv) the summed overall subjective value of both options (sEV_1+sEV_2); (v) a term to capture any linear change in reaction times as a function of performing the task (see figure 4C); (vi) a term to capture autocorrelation, containing the RT from the previous trial; (vii) a term to capture any additional bonus for the trial being a 'no brainer', containing a 1 wherever a 'no brainer' trial occurred and a 0 otherwise. We normalized regressors (iii) and (iv) before entry into the design matrix, to account for any differences in the relative scales of subjective value functions across subjects. In figure 4B, we plot the mean +/- s.e. across subjects of parameter estimates (iii), (iv) and (vii) from this regression, and test for statistical significance using a two-tailed one-sample T-test across subjects.

In a subsidiary analysis (figure 6) we included two additional regressors; (viii) the objective value difference and (ix) the objective overall value of each trial, both normalized. In this analysis, we orthogonalized regressor (iii) with respect to regressor (viii), and regressor (iv) with respect to regressor (ix). These orthogonalized regressors will then capture just the effects of subjective value that deviate from a straight line (Andrade et al., 1999, Hunt, 2008, Hsu et al., 2009) – with the linear component of value being assigned to the objective value regressors. This provides a further test of the non-linearity of the subjective value functions used in the experiment.

4.1.3 Results

4.1.3.1 Subjects integrate across probability and magnitude in guiding their decisions

Our first analysis aimed to investigate whether subjects integrated across probability and magnitude when selecting which option to choose. To test this, a logistic regression analysis of subject choices was performed, with subject choices (option 1 or option 2) as the dependent variable, and objective probability, magnitude and value (probability * magnitude) of each option as independent variables. A constant was also modeled, to capture any bias towards choices of one option over the other. The

parameter estimates from this first-level regression were then submitted to a group-level one sample T-test, to test for significant effects of each factor across the group. The results are presented in table 1 and figure 3. It can clearly be seen that the probability and magnitude both influenced subjects' choices, with probability having a slightly larger influence than magnitude. The interaction of these two factors (Pascalian value) also had a significant influence on subject choices. Thus, subjects appeared to be integrating across both probability and magnitude when deciding which option to select.

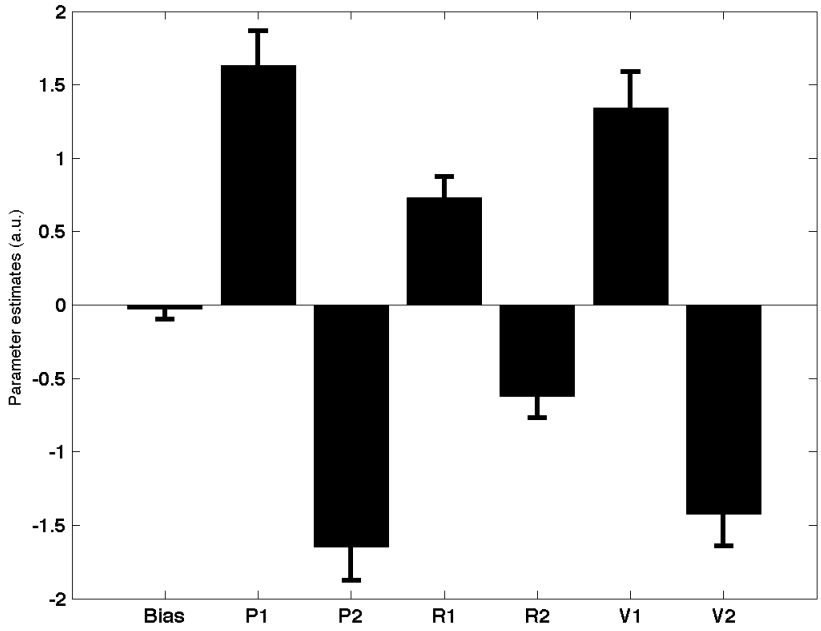

Figure 3. Logistic regression of choice behaviour. Mean +/- s.e. of parameter estimates associated with probability of option 1/2 (P1/2); reward magnitude of option 1/2 (R1/2); value of option 1/2 (V1/2).

Independent variable	T(29)
Choice bias (left>right)	-0.4263 *(n.s.)*
Probability(opt 1) (P1)	7.1569 *(***)*
Probability(opt 2) (P2)	-7.3434 *(***)*
Reward magnitude(opt 1) (R1)	5.3456 *(***)*
Reward magnitude(opt 2) (R2)	-4.3046 *(***)*
Pascalian value(opt 1) (V1)	5.5523 *(***)*
Pascalian value(opt 2) (V2)	-6.5689 *(***)*

Table 1. Group T-statisics for logistic regression on choices of right option. * denotes p<0.001**

4.1.3.2 Reaction time depends upon overall value, value difference, time through task and requirement for value integration ('no brainer' effect)

We then investigated what factors influence subject reaction times (RTs) (figure 4). We used a linear regression analysis to measure the influence of several variables on RTs. As expected, we found that trial difficulty – the *difference* in value between chosen and unchosen options – had an effect on RTs, with more difficult trials taking longer (blue bar, figure 4B; T(29)=-7.98, p<0.0005). Surprisingly, we also found that the *overall* value of a decision influenced RTs, with less valuable trials taking longer (green bar, figure 4B; T(29)=-2.36, p<0.05). We also included some trials in which both reward magnitude and probability were higher on one option than the other. There was an additional bonus in speed beyond that related to value for these 'no brainer' trials (brown bar, figure 4B; T(29)=-8.32, p<0.0005). Subjects were therefore faster on average on these trials than on those where probability and magnitude advocated opposing choices, and so needed to be translated into a 'common currency' in which the two stimulus features could be integrated. There was also a steady decrease in reaction time as subjects progressed through the task, suggesting subjects became less deliberative and more automated in their choices as they became familiar with the task (figure 4C).

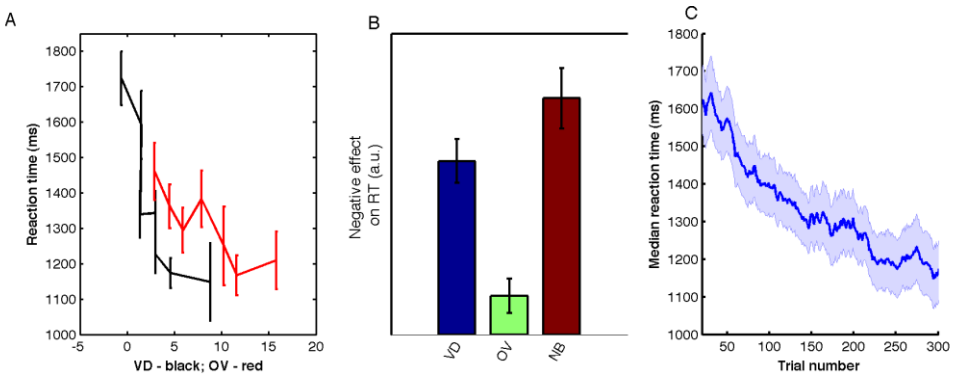

Figure 4. Effect of task variables on reaction times. (A) Reaction time (mean +/- s.e.) for an example subject, as a function of subjective value difference (black) and subjective overall value (red). (B) Effects of value difference (VD), overall value (OV) and 'no brainer trials' (NB) on subject reaction times (mean +/- s.e. across subjects), estimated using linear regression. Y-axis is flipped; positive values equate to a negative effect on RTs. (C) Running group mean +/- s.e. of reaction time (smoothed across 40 trials) as a function of trial number.

4.1.3.3 Subject behaviour is well described by Prospect theory

We next investigated the fits of Prospect theory models to our data, to see whether subjects tended to overweight or underweight low probabilities, and see whether Prospect theory provided a sensible fit to subject choices. We found that all but two of our subjects (28/30) had a probability weighting function which overweighted

low probabilities (i.e. $\gamma<1$; figure 5A), and all but one subject had a concave utility function (i.e. $\alpha<1$; figure 5B). Individual subjects' parameter fits are given in table 2. We also used Bayesian information criteria (BIC) to investigate whether a Prospect theory model provided a better explanation of subject choices than a simpler model, using objective rather than subjective values, and having only one free parameter (the temperature parameter τ). BIC favours models that provide a better fit to the data whilst penalizing models that have a higher number of free parameters (Pitt and Myung, 2002). In 25 out of 30 subjects, the subjective Prospect theory value function had a lower BIC (i.e. better fit) than the objective value function. This provided strong evidence that in our task, subject behaviour was well described by Prospect theory and subjects tended to overweight low probabilities.

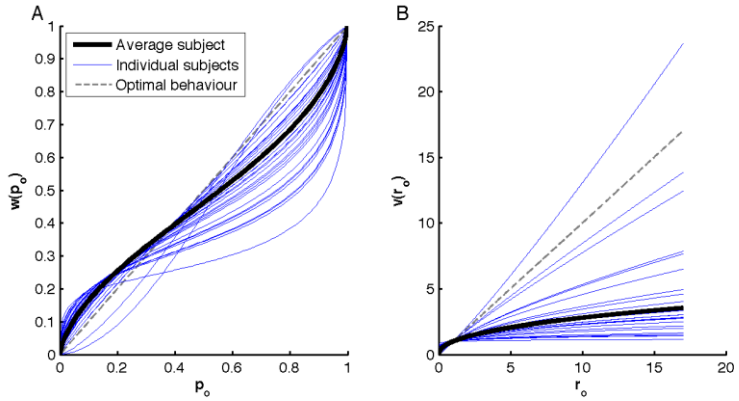

Figure 5. Prospect theory fits to subject behaviour. (A) Probability weighting function. (B) Expected utility function.

Subject	γ	α	τ	Log likelihood (subjective)	Log likelihood (objective)	BIC (subjective)	BIC (objective)
1	0.75	0.32	0.14	-72.9	-133.8	163.2	273.3
2	0.68	0.73	0.31	-68.4	-73.3	154.1	152.4
3	0.8	0.18	0.08	-45.5	-161.9	108.4	329.6
4	0.6	0.72	0.3	-75.1	-83.4	167.5	172.5
5	0.62	0.66	0.18	-52.8	-63.2	122.9	132.1
6	0.77	0.36	0.19	-90.6	-136.2	198.5	278.2
7	0.53	0.05	0.03	-36.8	-181.6	91	368.9
8	0.66	0.13	0.06	-43.5	-168.8	104.3	343.3
9	0.87	0.89	0.49	-61.9	-62.7	141.1	131.1
10	0.93	0.45	0.59	-156.2	-171.5	329.8	348.8
11	0.55	0.36	0.1	-72.9	-113.7	163.2	233.1
12	0.81	0.43	0.19	-74.9	-115.1	167.2	236
13	0.41	0.25	0.05	-80.1	-136.1	177.6	277.9
14	0.68	0.56	0.53	-144.1	-149.1	305.5	304
15	0.64	0.45	0.16	-74	-95.9	165.4	197.6
16	0.48	0.43	0.16	-110.1	-124.2	237.5	254.1
17	0.53	0.28	0.1	-93.4	-138.1	204.1	282
18	0.68	0.43	0.1	-44.5	-86.9	106.4	179.5
19	0.61	0.49	0.35	-94.6	-100	205	205.3
20	0.92	1.12	1.71	-66.9	-69.5	149.8	144.3
21	1.29	0.44	0.34	-53.7	-84.8	122.5	174.7
22	1.59	0.39	0.55	-95.6	-121.7	207.2	248.7
23	0.59	0.39	0.12	-63.3	-84.2	142.6	173.7
24	0.79	0.43	0.18	-58.2	-85.7	132.3	176.7
25	0.49	0.93	0.47	-54.7	-85.8	125.2	177
26	0.49	0.36	0.14	-87.7	-101.7	191.2	208.6
27	0.52	0.15	0.11	-96.2	-129.3	208.2	263.9
28	0.81	0.13	0.12	-70.3	-130.3	156.5	265.9
29	0.66	0.54	0.32	-80.3	-85.5	176.5	176.2
30	0.66	0.37	0.34	-104.4	-115.3	224.8	235.9

Table 2. Parameters γ, α and τ refer to probability weighting, utility and temperature parameters respectively in Prospect theory model; log likelihood is compared between this model and an 'objective value' model with only one free parameter (softmax temperature); Bayesian Information Criteria (BIC – lower numbers indicate better fit), which compare models whilst penalizing those with higher numbers of free parameters , showed a strong preference for the subjective model over the objective model (paired $T(58)=3.79$, $p<0.0005$).

4.1.3.4 A further test of Prospect theory: effect of subjective value on reaction times

We then performed a further test of whether the subjective value function explained behavioural data more successfully than the objective value function. We repeated the linear regression on subject RTs, but we included the subjective value functions *orthogonalised* with respect to the objective value functions. This orthogonalisation assigns any *shared* variance between subjective and objective value to the objective regressor, meaning that the effect size of objective value remains the same as before. For the subjective value regressor, it tests the portion of subjective value that deviates from a straight line – i.e. the portion that is *non-linear*. We found that the non-

linear portion of value difference (but not overall value) also had a significant negative effect on subject reaction times (figure 6).

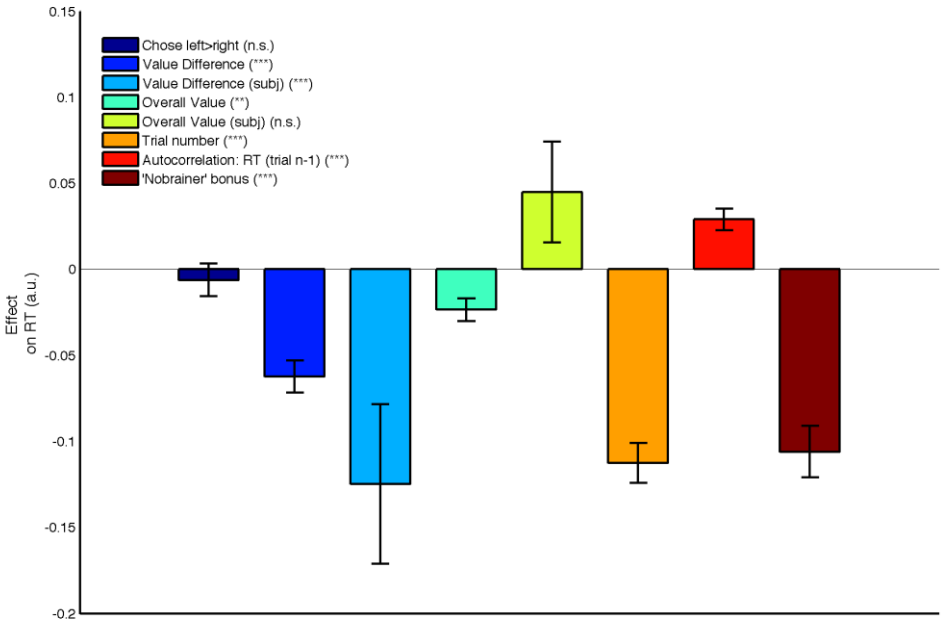

Figure 6. Results of a linear regression on reaction time, in which subjective values were orthogonalised with respect to objective values, allowing separate interrogation of the effects of the linear and non-linear components of the subjective value function on reaction time. There is an additional effect of the non-linear component of value difference (although no significant effect of the non-linear component of overall value). Additional co-regressors included whether subject chose left or right, the linear effect of trial number, a term accounting for autocorrelation (RT on previous trial), and the effect of 'no brainer' trials. Bars show mean +/- s.e. across subjects. ** denotes p<0.005; * denotes p<0.0005.**

4.1.3.5 Subject choices are consistent throughout the experiment

We next tested whether subjects changed their choice preferences as they progressed through the task. Perhaps the simplest way to test this is to ask what subjects *maximize* during the task – whether they choose the option with the highest reward probability, the highest reward magnitude, or the highest expected value, at each trial. Figure 7 shows the results of such an analysis, smoothed over a running window of 40 trials. Although there is some drift in each of the variables (with subjects becoming slightly more likely to choose the option with the highest probability at the end of the experiment than at the beginning (blue lines)), it can generally be seen that choice preferences are stable throughout the experiment. This was also reflected in Prospect theory parameters fitted to the first and second halves of the experiment (table 3); there are small changes in the parameters between the first and second half of the experiment, but none of these changes are statistically significant, especially when

compared to the dramatic change in subject reaction times between the first and second halves. Thus it can be seen that subject choice behaviour stays relatively stable throughout the experiment.

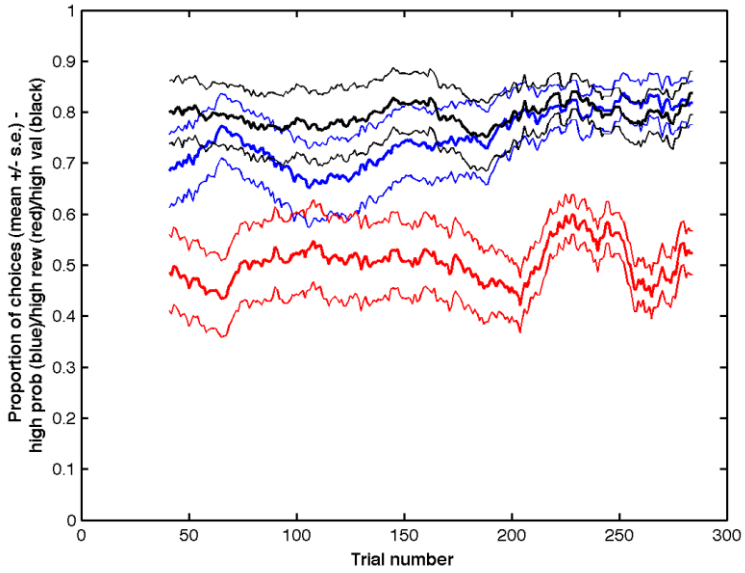

Figure 7. Proportions of trials (mean +/- s.e.) on which subjects chose the option with the highest probability (blue), highest magnitude (red), or highest Pascalian value (black). Running average, smoothed over 40 trials.

Parameter	Half 1	Half 2	Paired T(58)	p value
γ	0.92+/-0.13	0.74+/-0.08	1.22	0.2280 (n.s.)
α	0.53+/-0.05	0.42+/-0.05	1.73	0.0882 (n.s.)
τ	0.41+/-0.11	0.24+/-0.05	1.43	0.1593 (n.s.)
LogRT(ms)	7.27+/-0.04	7.07+/-0.04	3.53	0.0008 (**)

Table 3 – Prospect theory parameter fits (and reaction times) compared between first and second halves of the experiment. Reaction times decrease dramatically in the second half of the experiment, but there is little change in Prospect theory parameters.

4.1.3.6 A cross-subject speed-accuracy tradeoff

We then investigated whether there was a *speed-accuracy tradeoff* in our subjects – that is, did subjects who took longer to decide on average make more accurate decisions? We defined accuracy as the proportion of trials on which subjects chose the option with the higher (subjective) value. We found a highly significant cross-subject correlation of accuracy with median RT (R=0.5499, p<0.005; figure 8).

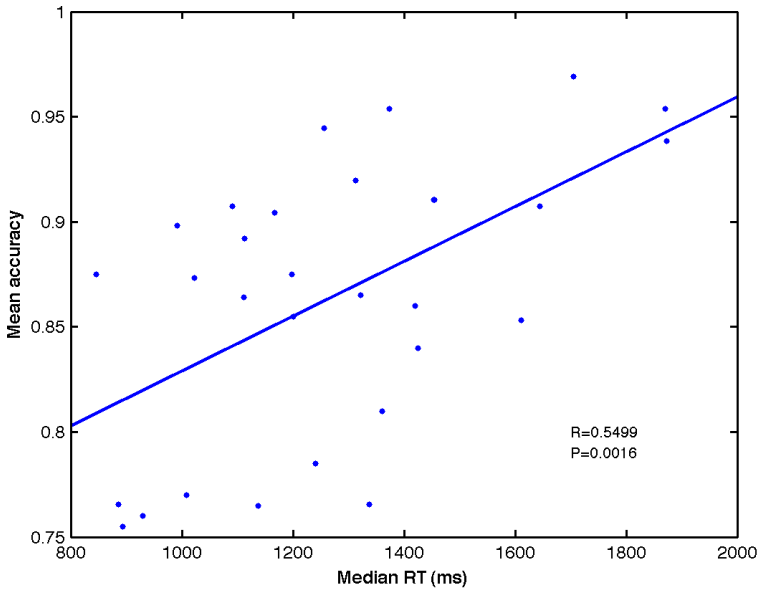

Figure 8. Cross-subject correlation of median reaction time with mean accuracy (percentage of trials on which trials chose option with higher subjective value).

4.1.3.7 Recent feedback slightly affects subject behaviour

Finally, we investigated whether subject choices were different on trials immediately after they had lost, vs. trials immediately after they had won. Although such a difference would be unexpected (as each trial's outcome was independent of the last), we hypothesized that subjects might suffer from the *gambler's fallacy* – that future returns depend upon past performance. We investigated whether Prospect theory parameters were altered on trials immediately after a loss vs. immediately after a win. There was a small effect on the α parameter ($T(29) = 2.1316$; $p<0.05$), meaning that subjects had a slightly more concave utility function after a loss than after a win (figure 9). There was also a small effect on the temperature parameter, τ ($T(29)=-2.3487$; $p<0.05$), but no effect on the probability weighting parameter, γ ($T(29) = 1.0622$; $p=0.29$). Thus, reward feedback has a weakly significant effect on the immediately subsequent trial.

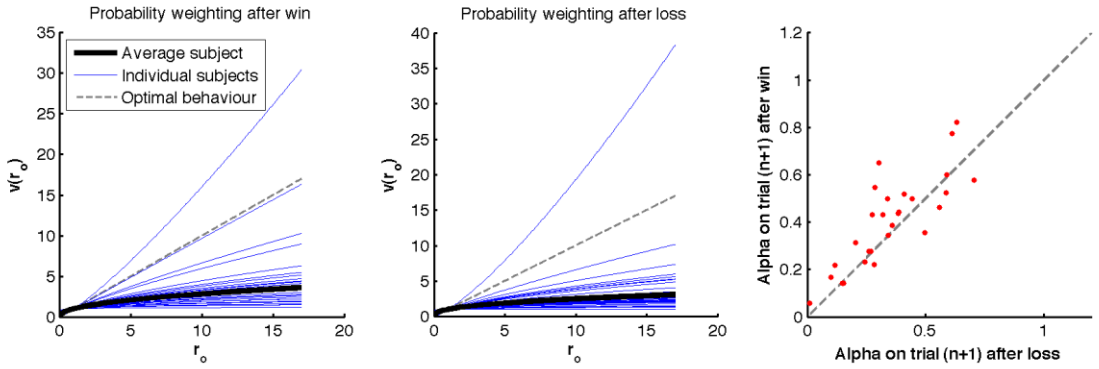

Figure 9. Weakly significant effect of reward feedback on expected utility function in Prospect theory. Left and middle panels show function after win and loss, respectively; right panel shows α parameter after loss and win for each subject.

4.1.4 Discussion

In this section, I introduced a simple choice task based on paradigms from the economics literature, in which subjects had to combine reward probability and magnitude to guide their choices. The paradigm was a 'decision from description' paradigm – that is, a description of probability and magnitude appeared onscreen at each trial. However, subjects also received full feedback at each trial, and so might also use their recent experience to guide their behaviour. In accordance with classic behaviour on tasks involving decisions from description (Kahneman and Tversky, 1979), subjects overweighted their probabilities of winning at small values. This subjective value function was also reflected in behavioural RTs. Subject behaviour remained relatively consistent throughout the experiment, except for RTs, which became significantly faster as the experiment progressed. Subject RTs were also faster on 'no brainer' trials, and on trials with a higher overall value.

One point of note is that subjects were slightly more influenced by reward probabilities than by reward magnitudes in this task. This was also reflected in the concavity of the expected utility function, with the median subject having an α parameter around 0.5; this is slightly lower than typical values of ~0.7 found in most economics studies. This can perhaps be attributed to the fact that reward probability is presented as a number, whereas the reward magnitude is presented as a bar, where perceptual discriminations might be slightly more challenging. Nevertheless, subjects did appear to integrate across both dimensions, and the expected utility function remains monotonically increasing, so this result does not provide too many difficulties for subsequent interpretation of neural data collected using this task.

4.2 Can a reinforcement learning model be used to capture learning about the expected behaviour of a confederate in a social interaction?

4.2.1 Introduction

In addition to information gathered from the environment, an important factor in many value-guided decisions is information from other conspecifics: social information. We shape our actions both in the light of information gathered from others, and also in the expectancy that our actions will have a certain impact on others' behaviour. Social learning is important throughout the animal kingdom, and there are numerous examples of observational learning shaping animal behaviour in mammals, birds and fish (Danchin et al., 2004). However, the number and complexity of social interactions in primates surpasses that of any other species, and social group size is a key determinant of primate neocortical volume – suggesting a critical role for brain size in determining the sophistication of social behaviour (Dunbar, 1993). Social sophistication has been shown to have measurable consequences for primate evolutionary fitness, implying a selection pressure that favours increasing brain size, which may have played an important role in the expansion of the human brain (Silk, 2007). One hypothesis proposes that certain regions of primate and human brains may even have developed as functional specialisations for social behaviour (Brothers, 1990).

It is unclear, however, whether information gathered from other conspecifics is amenable to similar computational strategies to more traditional reward-based learning, such as the associative learning mechanisms discussed in **chapter 2**, or whether it depends upon altogether different strategies. One helpful approach may be to adopt a normative perspective, and design simple interactive games that allow us to investigate the optimal use of social information. Formal studies of cooperative behaviour using one such game, the iterated prisoner's dilemma, have emphasised that very *straightforward* strategies show surprisingly successful and robust performance. One particularly successful strategy, outperforming numerous more complex algorithms, is 'tit-for-tat' – simply returning cooperation from a social partner with cooperation, and defection (non-cooperation) with defection (Axelrod and Hamilton, 1981). Even more successful is a 'win-stay-lose-switch' strategy, which uses only one's own outcome to determine the next behavioural response (Nowak and Sigmund, 1993). The success of such simple strategies suggests little need for sophisticated mechanisms for social inference, and also suggests that social learning may not need to be as sensitive as the more sophisticated associative learning strategies used for learning from the environment.

On the other hand, in slightly more developed social environments, the most successful strategies depend more heavily upon social observation, and employ a weighted average of recent behaviour similar to that seen in learning algorithms (Rendell et al., 2010). Human behaviour in repeated economic games with payoff matrices designed to test interactions other than cooperation is found to more closely match a strategy of reinforcement learning than any particular equilibrium strategy (Erev and Roth, 1998). Moreover, models that successfully describe behaviour may not only track the expected value of particular responses, but also attempt to perform 'belief inference' – that is, use reinforcement learning to predict the behaviour of the other individual at each trial, and shape one's own response accordingly (Camerer and Ho, 1999, Hampton, Bossaerts and O'Doherty, 2008). Such findings suggest that social learning may indeed be amenable to the same learning strategies employed for reward-guided tasks.

It is similarly unclear whether the human brain uses similar or distinct brain regions for learning information from social and non-social sources. Several recent studies have highlighted that social interaction can elicit reinforcement learning-like signals in regions traditionally associated with reward-guided learning, such as the ventral stiatum (King-Casas et al., 2005, Klucharev et al., 2009, Burke et al., 2010). However, it is not known whether these regions are found as a result of inferring something about the behaviour of the social partner, or because of the need to modify one's *own* actions in light of social information. In a separate literature, several brain regions have been proposed as being specialised for social inference, including portions of the temporoparietal junction and dorsomedial prefrontal cortex (Amodio and Frith, 2006, Saxe, 2006). It is surprising that these regions are not isolated in the studies of social interaction found to activate the ventral striatum.

Moreover, dissociations have recently been found between cortical subregions suggestive of their being involved in similar computational roles, but in social and non-social *frames of reference*. As discussed in **chapter 2**, a sulcal portion of the anterior cingulate cortex (ACC) correlates with the estimated level of volatility in a reinforcement learning model during a reward- guided learning task (Behrens et al., 2007). Volatility is important as it determines the value assigned to new pieces of information received during learning, and so controls the *learning rate*. Lesions to this sulcal portion of ACC are found to impair the assignment of the correct value to new pieces of information in an instrumental reward-guided learning task (Kennerley et al., 2006).

However, lesions to the nearby gyral portion of ACC impair the valuation of *social* stimuli, such as the presence of other conspecifics in the environment (Rudebeck et al., 2006). Control macaque monkeys are willing to forego a food reward in order to view a photograph of another monkey (Deaner, Khera and Platt, 2005). After receiving a lesion to the ACC gyrus, however, the monkeys are found to attribute no value to the social stimulus, and instead immediately reach out to take a food reward (Rudebeck et al., 2006). This is not a general deficit in impulsivity, as the lesioned monkeys are equally unwilling to take the food in the presence of a fear-inducing stimulus, such as a snake. Thus, ACC gyrus appears critical to valuation of social information, whereas ACC sulcus appears crucial in ascribing the value assigned to new pieces of reward-based information. These findings suggest the intriguing possibility that social and non-social information may be subject to similar computational strategies, but processed by different brain regions.

To test these ideas, we designed a learning task in which subjects had to combine information from social and non-social sources in order to guide their decisions. We investigated whether subjects used a similar reinforcement learning strategy for the different sources of information, and whether subjects integrated equally across different sources of information in shaping their choices. We subsequently investigated the neural correlates of social and non-social learning using functional MRI, and these results are discussed in **chapter 6**.

4.2.2 Methods

4.2.2.1 Experimental task

24 human subjects (14M/10F, age range 20-62, 4 left-handed) performed a decision-making task (whilst undergoing functional MRI) in which they repeatedly chose between blue and green rectangles in order to accumulate points (figure 10). One subject was excluded from subsequent analysis due to excessive head motion during fMRI data acquisition, leaving 23 subjects remaining for analysis. The point score (a random number between 1 and 100) associated with blue (f_{blue}) and green (f_{green}) was shown in the centre of each rectangle; this number was added to the subject's score if they chose the correct option. Subjects also saw a red bar onscreen, whose length was proportional to their current score; they aimed to reach a silver target to win £10, or a gold target to win £20. Subjects were instructed that either blue or green would be correct on each trial, but that the probability of the two colours being correct was not equal – instead, the chance of each colour being correct depended upon the recent outcome history. Subjects were informed that the probabilities of each colour being

correct were independent of the rewards available. Thus, as a result of the difference in reward magnitudes associated with the blue and green options, subjects often picked the less likely colour if it was associated with a higher reward. As the probability of green being correct (*r*) was always the inverse probability of blue being correct (1-*r*), subjects (and the reinforcement learning model of the task) needed only track one probability.

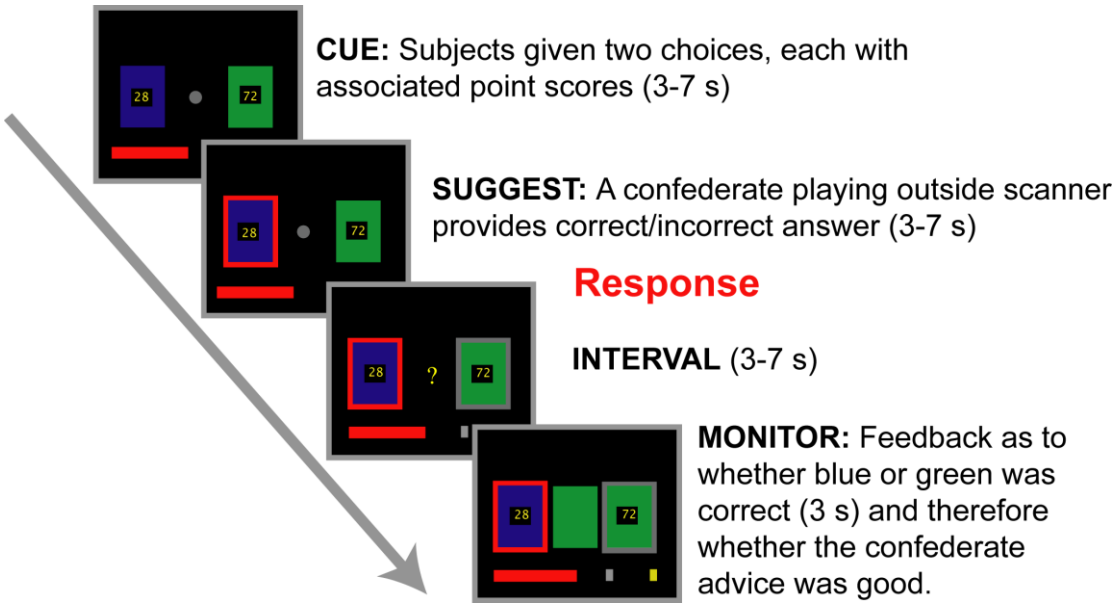

Figure 10. Experimental task. Subjects chose between two options of different reward magnitude, whose probabilities of yielding reward were learnt through time, in order to accumulate as many points as possible. Subjects also received advice from a confederate at each trial, who had the option to provide correct or incorrect advice; the fidelity of the confederate's advice could also be learnt through time.

On each trial, 3-5 seconds after first seeing the stimuli (CUE phase), subjects also received computer-generated advice about which rectangle to choose from a "human confederate", supposedly playing outside the scanner. This advice appeared for 3-7 seconds (SUGGEST phase) before the subject was allowed to make their decision, and remained onscreen until an option was selected. After the subjects had made their choice, there was a 3-7 second interval (INTERVAL phase) before the correct answer was revealed. The correct answer remained onscreen for 3 seconds (MONITOR phase), and was then replaced by a fixation point for 1 second before the next trial began.

Subjects were introduced to an actor before the experiment began, and both subject and actor were taken through the experimental instructions and practised the task together. The confederate had two 'ranges' presented on their screen, gold and silver, which the subject would be unable to see during the experiment (figure 11). In front of the subject, confederates were told that if the subject's red bar ended the

experiment within one of these ranges, the confederate would receive £20 (gold) or £10 (silver). On each trial, the confederate would be given two options: 'Provide correct answer' or 'Provide incorrect answer'. When they made their choice, the correct or incorrect answer would be highlighted on the subject's screen ('SUGGEST, figure 10). It was made clear that the confederate was not able to see whether blue/green was the correct answer, nor see the rewards available on each trial – their advice would therefore be *independent* of these other sources of information. Subjects were also told that the confederate was unable to see whether or not they took the advice, and so they could make use of consistently *unhelpful* advice by going *against* their confederate's suggestions. The only feedback that the confederate would receive was an update, approximately every five trials, of how far advanced the subject's red bar was, and how far through the experiment the subject had progressed.

Figure 11. Confederate 'task'. During the instruction period, it was pointed out that the confederate's aim was to land the subject in one of two 'ranges' that could appear anywhere along the money bar, and which only the confederate could see. This was to motivate the subject into believing that the confederate might or might not provide good advice, and this advice might change during the course of the task. We carefully highlighted that no other information was available to the confederate (e.g. reward magnitudes, subject choices, reward feedback etc.) to avoid any possibility of 'Machiavellian' strategizing by confederate or subject. In reality, the confederate was replaced by a computer who delivered correct advice on predetermined trials.

The ranges could be located anywhere along the length of this bar; they could be close together or far apart. Thus, as in figure 11, situations could easily be designed in which a confederate might reasonably give unhelpful advice initially (to try to land the subject in the gold range by the end of the experiment), but *change* this advice as the subject did better than expected, and the confederate's *motive* changed (to try to land the subject in the silver range instead). Several examples of different situations were given in the initial instructions, to explicitly make clear to the subject that the confederate's motives would depend upon the location of these ranges, and that these motives might change over time. As the subject was unable to see the two ranges, their only insight into the confederate's current motive would be the reliability of the advice that they received on each trial.

In summary, subjects had three sources of information at each trial to guide their choice: (i) the reward magnitude on each option; (ii) whether each option was blue/green (combined with the recent history of reward on blue/green); (iii) which option was suggested by the partner (combined with the recent history of confederate fidelity). At feedback, subjects received two new pieces of information that could guide their future behaviour; each outcome revealed information both about the future probability of blue or green being correct and about the fidelity of the confederate's advice.

Subjects underwent 120 trials in total. During the first 60 trials, the reward history was stable, with a 75% probability of blue being correct. During the next 60 trials, the reward history was volatile, switching between 80% green correct and 80% blue correct every 20 trials. Meanwhile, during the first 30 trials, the social advice given was stable, with 75% of suggestions being correct. During the next 40 trials, the social advice given was volatile, switching between 80% incorrect and 80% correct every 10 trials. During the final 50 trials, the advice given was stable again, with 85% of suggestions being incorrect. In order to counterbalance the design, eleven of the subjects had the advice inverted, such that the first 30 trials were 75% incorrect, and the last 50 trials were 85% correct. Hence, the dashed line in figure 12 (below) refers to the probability of true advice in half the subjects and the probability of false advice in the other half.

4.2.2.2 Reinforcement learning model

We used a reinforcement learning (RL) model to track probabilistic information in the task. Importantly, there were two features of the task that could be tracked using an RL strategy: (i) the probability of blue or green yielding reward at each trial, based on the past reward history, and (ii) the probability of receiving correct advice on each trial, based on previous confederate fidelity. Both social and non-social information were therefore amenable, at least in theory, to the same strategy of learning via reinforcement.

The RL model used was the Bayesian model developed by Behrens and colleagues, discussed in **chapter 2**. The algorithm has been documented in detail in that chapter and elsewhere (Behrens et al., 2007), but we briefly describe its concept here. The model assumes that outcomes are generated with an underlying probability, *r*. The objective is to track *r* as it changes through time. The crucial question addressed by the model is how much the estimate of *r* should be updated when a new positive or negative outcome is observed. In order to know how much to update the estimate of *r* on

witnessing a new outcome, it is necessary to know the rate of change of *r*. If *r* is changing fast on average then an unlikely event is more likely to signify a large change in *r*, so an optimal learner should make a large update to its estimate. The Bayesian model therefore maintains an estimate of the expected rate of change of *r*, referred to as the volatility *v.* In a fast changing environment, the model estimates a high volatility and each new outcome has a large influence on the optimal estimate of the reward rate. Conversely, in a slow changing environment, the model estimates a low volatility and each new outcome has a negligible effect on the model's estimate of *r*.

4.2.2.3 Logistic regression choice analysis

We performed a multivariate logistic regression to establish factors that predicted subject choices. If subjects were performing the task optimally, they would learn a probability associated with blue rather than green being correct based on the history of outcomes. They would also learn a separate probability of the confederate giving correct advice based on the history of correct and incorrect advice at previous trials. When the confederate advice became visible at the current trial, subjects should then combine these probabilities to provide an overall probability that blue (and conversely green) would be the correct option. This probability should then be weighed with the respective reward magnitudes on each option to guide the final decision.

Using the Bayesian reinforcement learning model described above, we generated the optimal estimates of these probabilities based on the same observations witnessed by the subjects in the scanner. If subjects were learning the probability of both the outcome and confederate advice according to such an associative strategy, these two factors should be key in predicting subject behaviour. We also considered as factors two alternative strategies that might predict subject behaviour with respect to the confederate advice. First, subjects might blindly follow confederate advice without learning the probability that this advice would be good; and second subjects might appreciate that the confederate may have a strategy of giving good or bad advice, but subject may fail to integrate this advice over a number of trials in an RL-like fashion - instead relying only on the confederate's most recent behaviour, analogous to common tit-for-tat models of social behaviour (Axelrod and Hamilton, 1981).

We therefore had five factors with which to predict subject choices (coded 1 for occasions when subjects chose blue and 0 for occasions when subjects chose green); these were (i) the difference in reward magnitudes on the two options (f_{blue}-f_{green}); (ii) the RL probability that blue would be correct given the history of outcomes; (iii) the RL probability that blue would be correct given the advice of the confederate at the current

trial and the history of correct confederate advice; (iv) the confederate advice at the current trial, ignoring the confederate's history; (v) the confederate advice at the current trial interacted with the correctness of the confederate advice at the previous trial.

Regressor (iv) has the value 1 whenever the confederate advises blue, and 0 whenever the confederate advises green. Regressor (v) has the value 1 when the confederate advises blue on the current trial, after giving correct advice at the previous trial, or when the confederate advises green, after giving incorrect advice at the previous trial; otherwise this factor has the value 0.

As in section 4.1.3.1, the logistic regression for each subject analysis results in a parameter estimate for each factor, reflecting the extent to which that factor predicts subject choices. Significant effects were also analysed in individual subjects, and assessed with a threshold of $Z>2.3$, $p<0.01$ for each subject.

4.2.2.4 Weighting of different sources of information

Whilst the logistic regression analysis described in section 4.2.2.3 is convenient in that it allows us to perform a statistical test on parameter estimates from the regression, it does not necessarily provide an accurate estimate of the *relative* weightings of each of the factors, as two of the factors represent *probabilities* whilst the third represents reward *magnitude*. We therefore constructed a model that incorporates this knowledge, to gain a more accurate estimate of how much weight is given to each information source. This information becomes important in subsequent cross-subject analyses of the functional MRI data, presented in **chapter 6**.

Optimal behaviour is to compute separately the probability of the next outcome given reward history information (p_r) and the probability of the next outcome given the current confederate advice and the history of confederate truths (p_c). Subjects should then combine these probabilities into an overall probability (p_o) according to Bayes' rule:

$$p_o = \frac{p_r p_c}{p_r p_c + (1-p_r)(1-p_c)}$$

Equation 6

In order to compute the overall value of the action (V_o), subjects should then multiply this overall probability by the reward magnitude available on the action (r_o):

$$EV_o = p_o r_o$$

Equation 7

In order to account for the fact that subject behaviour is guided to different extents by the different sources of information, we included a free parameter for each source of information, that allows subjects to either upweight or downweight this probability with respect to the other, and with respect to reward magnitude. We assume a sigmoidal form for this weighting, such that for each source of information:

$$p = \frac{1}{1 + \exp(-\gamma(p_{opt} - 0.5))}$$

Equation 8

where p is the probability used by the subject and p_{opt} is the probability computed by the optimal model. This equation transforms the optimal probabilities such that they are nearer to 0.5 if γ is small (and hence the source of information has *less* influence on behaviour), and nearer 1 or 0 if γ is large (giving the source of information *more* influence on behaviour). This is again combined with reward magnitude to give the subjective expected value on each option, sEV_o. To reduce the number of free parameters in the model, we assume that the expected utility is linearly proportional to the reward magnitude.

Subjects are then assumed to generate actions stochastically, according to a further sigmoidal probability distribution (as used in section 4.1.2.2) (Sutton and Barto, 1998):

$$P(C = o) = \frac{e^{\frac{sEV_o}{\tau}}}{\sum_{i=1}^{n} e^{\frac{sEV_i}{\tau}}}$$

Equation 9

where n is the number of options available, again 2 in this study. We fit this model using Bayesian estimation techniques (using direct numerical integration) in order to estimate γ for reward information (γ_r) and confederate information (γ_c).

4.2.3 Results

4.2.3.1 Volatility and choice probabilities for social and non-social information are decorrelated

As discussed in **chapter 3**, in order to estimate the contribution of social and non-social information to subject choices, it was important to ensure that these two sources of information were *decorrelated* so that they could explain separate portions of variance in choice behaviour. Similarly, it was important to make sure that the estimated *volatility* of the two sources of information (estimated using the reinforcement learning model) were also decorrelated, as these estimates would subsequently be used to explain variation in BOLD fMRI responses recorded during the task. We therefore designed a task schedule that ensured that these explanatory variables were indeed decorrelated (figure 12). There was no correlation between the probability of choosing green based on past reward history, or based on confederate advice (R=0.0642,*p*=0.48). These measures were also decorrelated from reward magnitude difference between green and blue options (advice and reward: R=-0.0419,*p*=0.64; blue/green history and reward: R=-0.1013,*p*=0.27). Finally, the estimated volatility of information from the confederate advice was decorrelated from the estimated volatility of reward history (R=-0.0752,*p*=0.41).

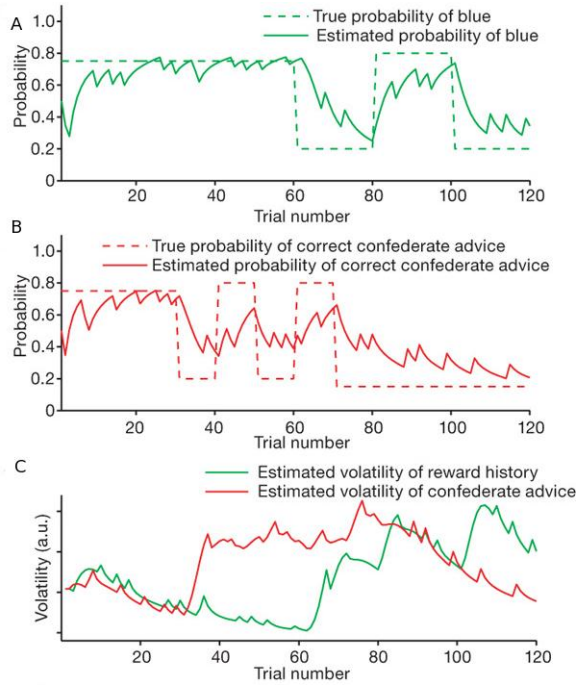

Figure 12. Experimental schedules. (A) True and RL-estimated probability of blue option yielding reward at each trial. (B) True and RL-estimated probability of confederate providing correct advice at each trial (flipped for 11/23 subjects for counterbalancing, such that advice was initially incorrect and later correct). (C) RL-estimated volatility of reward history and confederate advice.

4.2.3.2 Subjects use all three sources of information to guide choices, and use an RL-like strategy to track confederate behaviour

We used logistic regression to estimate the extent to which each source of information had a significant effect on subject choice behaviour. We used the difference in reward magnitude (RMD), the reinforcement learner's estimate of green being correct based on past outcomes (RLO), and the reinforcement learner's estimate of green being correct based on past confederate advice (RLC) as variables to explain the probability of choosing green at each trial. For tracking confederate advice, we also considered two alternative and more straightforward strategies – that subjects either blindly followed confederate suggestions (BFC) or assumed the confederate would perform the same as in the previous trial (CPT). We found a highly significant effect of RMD ($T(22)=9.38, p<0.0001$), RLO ($T(22)=5.15, p<0.0001$) and RLC ($T(22)=10.11, p<0.0001$) on subject behaviour across the population, but little effect of BFC ($T(22)=1.95, p=0.06$) and no effect of CPT ($T(22)=0.35, p=0.73$). We also investigated the extent to which each of these factors influenced individual subject choice behaviour, using a criterion of $p<0.01$ for detecting a significant influence of each factor on choices. This analysis shows that BFC and CPT were each significant in 3/23 subjects, whereas RLO and RLC each showed significant effects in 14/23 subjects and RMD showed a significant effect in 19/23 subjects. These results strongly indicate that subjects used all three sources of information to guide their choices, and adopted an RL-like strategy to estimate the fidelity of the confederate advice during the course of the experiment.

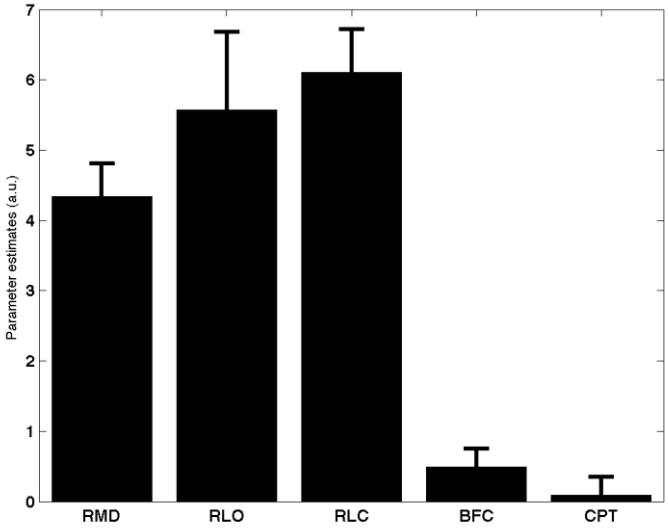

Figure 13. Logistic regression analysis of influence of different sources of information on subject choice behaviour. Explanatory variables were reward magnitude difference (RMD); RL-derived outcome probability (RLO); RL-derived confederate probability (RLC); 'blindly following confederate' strategy (BFC); 'assume confederate behaviour is same as on past trial' strategy (CPT). Bars represent mean +/- s.e. across population.

4.2.3.3 Variation in the extent to which individual subjects used each source of information

Whilst logistic regression provides a useful means of testing the hypothesis that each source of information influences behaviour, it does not provide a particularly accurate estimate of how much subjects weighed each source of information relative to each other. To account for this, we used a simple parametric model (described in section 4.2.2.4) which includes two weighting factors, γ_c and γ_o, that adjust the weighting assigned to the true (RL-derived) probability of outcome/fidelity to reflect the *subjective* weighting function used by individual participants. The subjective weighting functions are shown in figure 14. It can be seen that although on average subjects used the two sources of information to a similar extent, there was also considerable *variation* in the extent to which individuals used each source. This becomes important in subsequent analysis of the fMRI data collected during the task, and how fMRI responses vary across the population (**chapter 6**).

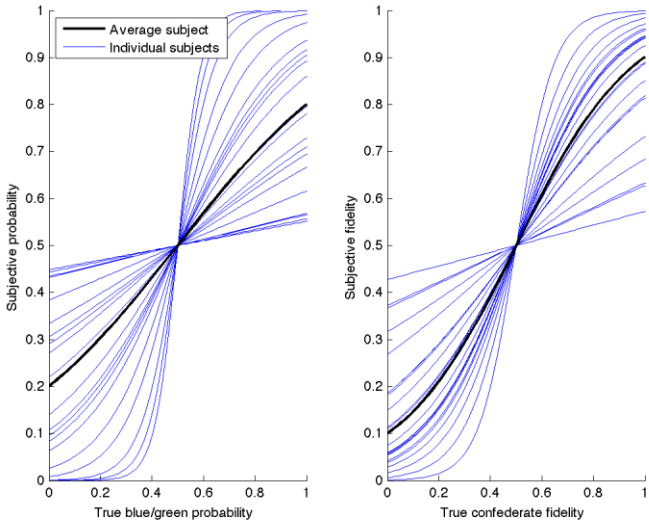

Figure 14. Subjective weighting functions for probability of blue/green reward (left), and probability of confederate fidelity (right). Although the average subject (solid black line) uses similar weighting for both sources of information, there is considerable variation across the population (thin blue lines) in the use of each source.

4.2.4 Discussion

In this section, I introduced a task designed to probe whether subjects used a similar computational strategy when learning about the probability of reward on different options as when making inferences about the advice received from a human confederate. In agreement with some recent analyses in the economics literature, it appears that human behaviour can indeed be explained using a model that tracks partner behaviour using reinforcement learning. Moreover, such a model provides a better account of behaviour than simpler models that fail to integrate over the past

several trials. This suggests that similar computational strategies might be implemented by the brain in learning from social and non-social sources of information; in **chapter 6**, I investigate whether these occur in similar or distinct neural substrates.

Importantly, subjects were found to integrate across multiple sources of information to guide their behaviour. However, the influence of each source of information varied across subjects, with some subjects paying more attention to reward history, and other subjects paying more attention to collaborator advice. These findings might suggest that stronger neural signals relating to each source of information might be found in subjects for whom that information has a more pronounced effect on behaviour. By decorrelating the regressors used for each source of information, we should be able to investigate the neural signals relating to each source independent of any confound from other sources.

4.3 Summary

Models of decision making require accurate estimates of subjective values in order to make accurate predictions of neural activity. Subjective values deviate from objective reward probabilities and magnitudes in a systematic manner, and these biases have been well described by Prospect theory for single-shot economic decisions in the absence of feedback. Here I showed that Prospect theory is also successful in describing subject behaviour in an experiment with several hundred trials, with full feedback. Subjective values can also be learnt through time, using associative learning strategies, and may be influenced by social interactions. Here I showed that social learning might, in some circumstances, be amenable to the same associative learning strategies as non-social stimuli.

In **chapters 5 and 6** I use these experimental tasks as the basis of the body of neuroimaging work conducted for this thesis. **Chapter 5** presents the temporal dynamics of cortical activity across different frequency bands during the value-guided choice paradigm discussed in section 4.1, and compares value correlates during this task to predictions from a biophysical model of decision making. **Chapter 6** examines the neural structures that support associative learning of social value during the choice paradigm discussed in section 4.2. In **chapter 7** I bring together the main findings from these studies, and highlight some of the conceptual advances that they offer to the field of reward-guided decision making.

Chapter 5: Mechanisms underlying cortical activity during value-guided choice

When choosing between two options, correlates of their value are represented in neural activity throughout the brain. It is unclear, however, which of these representations reflects activity fundamental to the computational process of value comparison, as opposed to other computations covarying with value. In this chapter, I investigate activity in a biophysically plausible network model that transforms inputs relating to value into categorical choices. A set of characteristic time-varying signals emerges that reflects value comparison. I test these model predictions in magnetoencephalography data from human subjects performing value-guided decisions. Parietal and prefrontal signals matched closely with model predictions. These results provide a mechanistic explanation of neural signals recorded during value-guided choice, and a means of distinguishing computational roles of different cortical regions whose activity covaries with value.

"By analogy with other macro-description methods, the promise of field-potential research is twofold. I) It may disclose interaction phenomena that are not accessible to single-unit studies and may thereby help to understand higher stages of intracortical information processing. 2) By combining field-potential data with anatomical and single-unit results, it may also be possible to bridge the gap between the micro-description at the single-unit level and the macro-description at the field-potential level."

Ulla Mitzdorf, 1985

5.1 Introduction

There has been widespread recent interest in identifying neural mechanisms that support the ability to choose between competing courses of action. Neural activity has been isolated that correlates with the reward value of making a particular action, and integrates across dimensions that might influence value such as reward magnitude and probability (Rangel, Camerer and Montague, 2008, Rushworth and Behrens, 2008, Kable and Glimcher, 2009). Using single-unit electrophysiological recording and functional magnetic resonance imaging (fMRI), correlates of value have been found in numerous brain regions, including parietal cortex (Platt and Glimcher, 1999, Dorris and Glimcher, 2004, Sugrue, Corrado and Newsome, 2004, Gershman, Pesaran and Daw, 2009), ventromedial (Blair et al., 2006, Plassmann, O'Doherty and Rangel, 2007, Tom et al., 2007, Boorman et al., 2009, Gershman, Pesaran and Daw, 2009), dorsolateral (Plassmann, O'Doherty and Rangel, 2007, Kim, Hwang and Lee, 2008), and orbital (Padoa-Schioppa and Assad,

2006) portions of prefrontal cortex, posterior cingulate cortex (McCoy and Platt, 2005, Kable and Glimcher, 2007), striatum (Knutson et al., 2005, Cai, Kim and Lee, 2011), and even early sensory (Serences, 2008) and late motor (Hernandez, Zainos and Romo, 2002) cortices. A neural decision process should take information reflecting the subjective value of each available action, and transform these inputs into a choice. It is unclear, however, which of these value-coding brain regions perform computations underlying choice, or even what form a choice signal should take. Some regions, such as ventromedial prefrontal cortex, have been the subject of particular debate (Kable and Glimcher, 2009, Noonan et al., 2010); in some fMRI studies this region has been found to signal a difference between chosen and unchosen values (Serences, 2008, Boorman et al., 2009), whilst in others it has appeared to signal the overall value of available reward (Blair et al., 2006), or the value of just the chosen option (Kable and Glimcher, 2007). Here, I use a biophysical model to derive predictions of the temporal dynamics of activity in a region of cortex underlying value comparison. The biophysical model is based upon knowledge of single unit activity, but is used to predict the net post-synaptic activity (i.e. local field potential) that should be observable in a brain region that makes choices. I investigate which regions of cortex show temporal dynamics that match these model predictions.

As discussed in **chapter 1**, one brain region that exhibits single unit activity consistent with a role in value-guided choice is in the macaque lateral intraparietal area (LIP), the homologue of a mid-posterior region of the human intraparietal sulcus. Neural activity in this region precedes saccadic eye movements towards a particular spatial location (Gnadt and Andersen, 1988), and is strongly influenced by the value of making an eye movement towards that location (Platt and Glimcher, 1999, Dorris and Glimcher, 2004, Sugrue, Corrado and Newsome, 2004). When options of differing subjective values are presented in different spatial locations, activity in LIP evolves from initially representing the subjective desirability of each option to eventually representing the probability of saccading towards each option (Louie and Glimcher, 2010).

LIP activity has also been the subject of perceptual decision tasks in which the coherent direction of motion must be determined from within a stereogram of randomly moving dots (Shadlen and Newsome, 2001, Roitman and Shadlen, 2002). In these tasks, LIP activity reflects a steady integration of evidence for directional motion from the noisy stimulus until a decision threshold is reached, at which point a saccade is executed. As discussed in **chapter 2**, this activity resembles components of drift-diffusion and

accumulator models of decision making (Bogacz et al., 2006), but it can also be described using biophysically realistic network models of spiking neurons endowed with N-methyl D-aspartate (NMDA) receptors to allow for slow integration of perceptual evidence (Wang, 2002, Wong and Wang, 2006). In such models, neurons selective for a saccade in a particular direction reciprocally excite similarly selective cells whilst inhibiting neurons selective for other directions through a shared pool of inhibitory interneurons. This recurrent mechanism enables the model to slowly integrate motion-sensitive inputs until a decision is made. The network models bear the advantage that they not only predict characteristics of subject behavior, but also predict the temporal evolution of neural activity in LIP in a biophysically plausible manner (Soltani and Wang, 2010).

It is unclear, however, whether this same integrative process would apply in the context of a value guided choice, in which evidence may not need to be accumulated over time, and where the noise on the inputs to the decision arises within the nervous system, rather than from the stimulus (Lee and Wang, 2008). It is equally unclear whether outside the context of saccadic decision tasks in which macaque subjects have been highly overtrained, these biophysical models might describe neural activity in the numerous cortical regions other than LIP in which value representations have been found. If so, such a model may even allow us to distinguish the roles of these cortical regions, and isolate those fundamental to value comparison.

To test this idea, we employed magnetoencephalography (MEG), a technique that allows us to examine the temporal evolution of activity in multiple cortical areas simultaneously, in human subjects performing a value-based decision task. We used simulations from a biophysical model to derive predictions not of single unit activity, but of the summed postsynaptic potentials of all excitatory cells in the network, as is likely to be isolated using MEG (Hämäläinen et al., 1993). We found regions in both parietal and ventromedial prefrontal cortex that matched well with predictions from the model, whereas other cortical regions showed value correlates that might be explained by appealing to their roles in separate cognitive processes.

5.2 Methods

5.2.1 MEG/MRI data acquisition

MEG data were sampled at 1000Hz on a 306-channel VectorView system (Elekta Neuromag, Helsinki, Finland), with one magnetometer and two orthogonal planar gradiometers at each of 102 locations distributed in a hemispherical helmet across the scalp, in a magnetically shielded room. A band-pass filter of 0.03-330Hz was applied during acquisition. Head position was monitored at the beginning of each run, and at twenty-minute intervals during each run, using four head position indicator (HPI) coils attached to the scalp. Data were acquired in two or three runs, with pauses between blocks to save data acquired. HPI coil locations, headpoints from across the scalp, and 3 anatomical fiducial locations (nasion, left and right pre-auricular points) were digitized using a Polhemus Isotrak II prior to data acquisition. Simultaneous 60-channel EEG data was acquired using a MEG-compatible EEG cap (ANT Neuro, Enschede, Netherlands), but is not discussed here. Vertical EOG and ECG were also measured to detect eye blinks and heartbeat, respectively. In experiment 2, eye location was monitored by high resolution Eyelink 1000 eyetracker (SR Research, Ontario, Canada) using 500Hz monocular recording.

MRI data for forward model generation were acquired using an MP-RAGE sequence on a Siemens 3T TRIO scanner, with voxel resolution 1x1x1 mm^3 on a 176x192x192 grid, TE= 4.53 ms, TI = 900 ms, TR= 2200 ms.

5.2.2 MEG data pre-processing

External noise was removed from MEG data using the signal space separation method (Taulu, Kajola and Simola, 2004), and adjustments in head position across runs (detected using HPI) were compensated for using MaxMove software, both implemented in MaxFilter version 2.1 (Elekta Neuromag, Helsinki, Finland). Continuous data were down-sampled to 200Hz and low-pass filtered at 40Hz, before conversion to SPM8 format (http://www.fil.ion.ucl.ac.uk/spm). Eye blinks were detected from the EOG channel (EOG data was bandpass filtered at 1-15 Hz; local maxima lying more than 3 standard deviations from the mean were considered blinks). Detected eye blinks were used to generate an average eye blink timecourse, on which principle components analysis was run to obtain spatial topographies describing the average eye blink; these were regressed out of the continuous data (as per (Berg and Scherg, 1994), without inclusion of brain source vectors as co-regressors; see http://www.fmrib.ox.ac.uk/~lhunt/artifact_session.zip for an SPM-

based tutorial). Data were epoched with respect to stimulus onset (-1000 to 2000ms around stimulus, with -200 to 0ms pre-stimulus baseline), and buttonpress (-2000 to 1000ms around response, again with -200 to 0ms pre-stimulus baseline). Artifactual epochs and bad channels were detected and rejected via visual inspection, using FieldTrip visual artifact rejection routines (Oostenveld et al., 2011).

5.2.3 Source reconstruction

MRI processing and forward modeling. All source reconstruction was performed in SPM8. MRI images were segmented and spatially normalized to an MNI template brain in Talairach space; the inverse of this normalization was used to warp a cortical mesh derived from the MNI template to each subject's MRI space (Mattout, Henson and Friston, 2007). Digitized scalp locations were registered to head model meshes using an iterative closest point algorithm, to affine register sensor locations to model meshes (Mattout, Henson and Friston, 2007). Forward models were generated based on a single shell using superposition of basis functions which will approximately correspond to the plane tangential to the MEG sensor array (Nolte, 2003). The forward models are implemented in FieldTrip's *forwinv* toolbox (Oostenveld et al., 2011).

Beamforming. Linearly constrained minimum variance (LCMV) beamforming (VanVeen et al., 1997) was used to reconstruct data to a grid across MNI space, sampled with a grid step of 7mm. Beamforming constructs a spatial filter at each grid location, to spatially filter the sensor space data, **y** (N sensors * t timepoints), to the grid location of interest, $\mathbf{r_i}$, with the aim of achieving unit pass band response at the location of interest while minimizing the variance passed from all other locations. The two ingredients in a beamformer are the N*3 lead field matrix at the location of interest, $\mathbf{H(r_i)}$, and the N*N sensor covariance matrix, $\mathbf{C_y}$. The 1*N weights vector, $\mathbf{w(r_i)}$, is given by:

$$w(r_i) = (H^T(r_i)C_y^{-1}H'(r_i))^{-1}H^T(r_i)C_y^{-1}$$

where $\mathbf{H'(r_i)}$ is a reduced N*1 lead field matrix at $\mathbf{r_i}$. The original N*3 lead field matrix is reduced to an optimal single orientation dipole by calculating the projected power, $\mathbf{p(r_i)}$, of the covariance matrix:

$$p(r_i) = (H^T(r_i)C_y^{-1}H(r_i))$$

and multiplying $\mathbf{H(r_i)}$ by the first principle component of $\mathbf{p(r_i)}$. The data at the source location of interest, $\mathbf{d(r_i)}$, is then given by multiplying the weights vector by the original sensor data:

$$d(r_i) = w(r_i) * y$$

This can be repeated across all grid locations to give a whole-brain image.

$\mathbf{C_y}$ was estimated using data pass band-filtered to the frequency band of interest, 2-10Hz, using 0% regularization. For stimulus-locked analyses, we included all non-artifactual trials from stimulus onset to 1 second after stimulus onset. For response-locked analysis, we included all non-artifactual trials from 1.5 seconds prior to response onset to the time of the response.

5.2.4 Computational model

5.2.4.1 Model implementation

We implemented a mean-field reduction (Wong and Wang, 2006, Wong et al., 2007) of the spiking neuronal network model described in (Wang, 2002). This reduced model captures the dynamics of the full model and exhibits neural and behavioral results similar to the full model. The full network model (Wang, 2002) describes a mechanism by which local cortical competition between two selective pools of neurons can be realized. It was originally designed to capture lateral intraparietal single unit activity in the random dot stereogram motion discrimination task. It comprises two selective excitatory populations (240 cells each), a large non-selective excitatory pool (1120 cells), and an inhibitory pool (400), with all-to-all connectivity. The selective neurons receive external inputs proportional to the level of coherent motion towards those neurons' receptive field, and have stronger excitatory connections to other cells within the same selective pool than those outside the pool.

The mean field approximation of this model reduces it from a system containing 7200 dynamical variables to one containing only 2 dynamical variables, greatly reducing the time required for numerical simulation and simplifying the task of exploring changes in model parameters. The model retains biophysically realistic parameters, information about the instantaneous mean firing rate of cells, synaptic input currents, and a slow NMDA gating variable for each selective population. Full details of the reduction are given in (Wong and Wang, 2006, Wong et al., 2007); here, we briefly revisit the reduced model's components, and add parameters used to convert the subjective value of two options into synaptic input currents for each of the selective two populations in the network.

The reduced model consists of two units (i=1,2), selective for one option each, with an excitatory recurrent coupling ($J_{A,ii}$) onto each unit, and an effective inhibitory coupling to the other unit ($J_{A,ij}$). Each unit receives external input currents that are proportional to the value of its favored option, as well as noisy background inputs which resemble endogenous noise in the cortex. The firing rate in each population of selective neurons is a function of the total synaptic input to this pool as follows:

$$r_i = f(I_i) = \frac{aI_i - b}{1 - \exp(-d(aI_i - b))}$$

Equation 1

where a, b and d determines the input-output relationship for a neuronal population and set to 270 Hz/nA, 108 Hz and 0.154s, respectively.

The total synaptic currents to each pool of neurons is set to:

$$I_i = J_{A,ii}S_i - J_{A,ij}S_j + I_0 + J_{A,ext}(r_i + r_{vis}) + I_{noise,i} \text{ (nA)}$$

Equation 2

where S_i is the NMDA synaptic gating variable related to neural pool i. I_0 represents the synaptic input current from external inputs to both pools and is fixed at 0.3297 nA; $I_{noise,i}$ is white noise filtered by a synaptic time constant of 2 msec and an amplitude of 0.009 nA; $J_{A,ext}$ represents the strength of synaptic coupling constant from external sources, and is set at 0.0011215 (nA/Hz); r_{vis} represents input firing rates of neurons which respond to the presentation of the visual stimulus, fixed at 7.5Hz; r_i represents the input firing rate proportional to the value of each option presented, and is given by the equation:

$$r_i = r_{dec}(1 + k_{dec}sEV_i) \text{ (Hz)}$$

Equation 3

where r_{dec} and k_{dec} are constants, and sEV_i the subjective expected value on option i derived above from Prospect theory. We set r_{dec} to be 10 Hz, and k_{dec} to be 0.1125 for the simulations in figure 1. (For a typical subject in experiment 1 (Prospect theory parameters

α=0.63,γ=0.64), r_i would therefore range between 10.63 Hz for the lowest value option on offer in the experiment and 14.03 Hz for the highest value option. Note that the values of r_{vis} and r_{dec} can be scaled as their product with $J_{A,ext}$ determines the selective inputs to neuronal pools in the network.) Finally, $J_{A,ii}$ was set at 0.3539 for the simulations show in figure 1A-D, and varied between 0.3166 and 0.3725 for the cross-subject variation simulations (figure 1E/F). $J_{A,ij}$ was set at 0.0966.

S_i for populations *i=1,2* are dynamical variables representing the slow synaptic currents attributable to NMDA receptor activation, given by the equation:

$$\frac{dS_i}{dt} = -\frac{S_i}{\tau_S} + (1 - S_i)\xi f(I_i)$$

Equation 4

where τ_S is the NMDA receptor decay time constant, set at 60 ms, and ξ is a parameter that relates the presynaptic input firing rate to the synaptic gating variable, set at 0.641. We used a total simulation period of 2500ms, with time step *dt* of 0.2ms. Stimulus onset (I_{vis} = 7.5 nA) was from 500ms, with reward-dependent inputs delivered from 600ms (I_{opt} = 10 nA); both inputs were offset at 2000ms. The decision was made when the firing rate of one of the populations reached a threshold of 30 Hz.

5.2.4.2 Model analysis

When analyzing the model's behavior, we no longer investigated the firing of individual selective neuronal populations (as previously (Wang, 2002)), but instead the summed synaptic inputs to both populations within the network, I_1+I_2. We chose this measure as, for the reasons discussed in **chapter 3**, MEG is more sensitive to the dipolar currents produced by post-synaptic potentials than the quadrupolar currents produced by action potentials, and also because the lack of separation between the neuronal pools means their activity is likely to be mixed when viewed at the macroscopic spatial scale resolved by MEG. However, neuronal firing rates and synaptic input currents are highly correlated within the model, and similar results could be obtained using firing rates as the dependent variable.

For predictions relating to a single subject (figures 1A-D), we simulated 6480 trials generated from the same stimulus set as used in experiment 1, with α=0.63, γ=0.64. In main

figure 1A, we plot the activity of the model as a function of the overall value ($sEV_1 + sEV_2$) of the decision, and in 1B as a function of the value difference ($sEV_{chosen} - sEV_{unchosen}$). We then treat the model outputs, **m**, in the same way as we had done the beamformed data at each location of interest (see section 5.2.3, below). First, we performed a time-frequency decomposition of the data on each trial from 2-10Hz using Morlet wavelets (Morlet factor 5). The decomposed data is then treated as the dependent variable as a function of overall value and value difference:

$$\mathbf{m}^{tr,f,t} = \beta_0^{f,t} + \beta_1^{f,t} * OV^{tr} + \beta_2^{f,t} * VD^{tr}$$

Equation 5

where $\beta_0^{f,t}$, $\beta_1^{f,t}$, $\beta_2^{f,t}$, and their associated variances, $var(\beta_0^{f,t})$, $var(\beta_1^{f,t})$ and $var(\beta_2^{f,t})$ are estimated using ordinary least squares regression. Figure 1C shows the T-statistic (=β/sqrt(var(β))) for overall value and value difference; figure 1D shows the averaged response across the relevant frequencies, after Z-transformation of the T-statistics.

For predictions relating to 'cross-subject' variation in model behavior (figures 1E/F), we simulated 1620 trials per instantiation of the model ('subject') for each of 30 subjects, varying $J_{A,ii}$ between 0.3166 and 0.3725 but keeping α and γ fixed. In main figure 1E we plot mean accuracy (%trials where $EV_{chosen} > EV_{unchosen}$) as a function of median reaction time. Linear regression was run on each "subject's" time-frequency decomposed data, and parameter estimates for each subject were then submitted to a second-level general linear model in which they were treated as dependent variables of median reaction time:

$$\beta_1^{s,f,t} = \kappa_0^{f,t} + \kappa_1^{f,t} * mRT^s$$

Equation 6

where $\kappa_0^{f,t}$ and $\kappa_1^{f,t}$ and their associated variances are estimated using ordinary least squares regression. Figure 1F shows the Z-statistic across frequencies from 2-10Hz for κ_1.

5.2.5 Experimental task

For this chapter, we used experimental data collected from the MEG experiment presented in **chapter 4**. The behavioural analysis is described in section 4.1.

5.2.6 Frequency domain analyses of beamformed MEG data

5.2.6.1 Frequency decomposition and linear regression

We used multiple regression to estimate the contribution of overall value and value difference to power in each frequency band at each timepoint through the decision. At each trial, the source-reconstructed data $\mathbf{d(r_i)}$ was decomposed into 10 time-frequency bins linearly spaced between 2 and 10 Hz, by convolving the data with Morlet wavelets (Morlet factor 5) (Tallon-Baudry et al., 1997). This yielded, at each trial, **tr**, frequency **f**, and timepoint, **t**, an instantaneous estimate of the power at that frequency.

Linear regression was then used to estimate the contribution of experimental variables that varied across trials to this value:

$$\mathbf{d}(\mathbf{r}_i)^{tr,f,t} = \beta_0^{f,t} + \beta_1^{f,t} * OV^{tr} + \beta_2^{f,t} * VD^{tr}$$

Equation 7

where OV is the subjective overall value (=sEV_{chosen} + s$EV_{unchosen}$) and VD is the subjective value difference (=sEV_{chosen} - s$EV_{unchosen}$). Overall value and value difference were normalized prior to regression, so they occupied a similar range of values across subjects. In a separate analysis, we subdivided all trials and regressors for the first and second halves of the experiment, and performed a contrast of parameter estimates to compare activity and value responses in the two halves of the experiment. In response-locked data from experiment 1 (see section 5.2), an additional coregressor describing whether the left or right option had been chosen was included as a coregressor of no interest. $\beta_0^{f,t}$, $\beta_1^{f,t}$, $\beta_2^{f,t}$, and their associated variances, var($\beta_0^{f,t}$), var($\beta_1^{f,t}$) and var($\beta_2^{f,t}$) were estimated using ordinary least squares regression. The parameter estimates, normalized by the their variances, were submitted to a group-level one-sample T-test to test for significant effects of OV and VD.

5.2.6.2. Region of interest analysis

We first performed a whole-brain analysis of task vs. baseline, focusing on activity in the 2-10Hz frequency range.

We then performed the linear regression described above (section 5.2.2.1) on data extracted from the clusters identified in the whole brain analysis, to identify regions whose

activity matched predictions from the biophysical model. Importantly, we only performed statistical inference on tests *orthogonal* to those originally used to identify the region of interest – namely, the main effect of task vs. baseline. Frequency-domain analysis was performed as in the whole brain analysis; timecourses show the group Z-statistic of the averaged (normalized) β-values from 2-4.5 Hz (for value difference) and 3-9Hz (for overall value), based on predictions from the biophysical model.

5.2.6.3 Statistical inference on region of interest analysis

For inference on the effects of overall value and value difference on region of interest data, we performed a cluster-based permutation test at the group level after collapsing across the relevant frequencies. We generated 5000 randomly permuted T-statistics for each timepoint, by randomly sign-flipping the group design matrix 5000 times. We then thresholded each permutation's T-statistic timeseries at a threshold of $T(29)>2.1$ (equivalent to $p<0.05$ uncorrected), and measured the maximum size of any cluster passing this threshold in the timeseries, to build a null distribution of cluster sizes. We then compared the size of clusters from the true T-statistic timeseries to those from the null distribution. We report clusters at a significance level of $p<0.05$, corrected for multiple comparisons across time.

5.3 Results

5.3.1 Biophysical model predictions

We used a mean-field version (Wong et al., 2007) of a biophysical cortical attractor network model (Wang, 2002) to derive predictions of the temporal dynamics of activity in a cortical region that selects between inputs reflecting the value of two options. The model comprises two populations of excitatory pyramidal cells selective for each option, with strong recurrent excitation between cells of similar selectivity, and effective inhibition between the two pools mediated by γ-aminobutyric acid (GABA)-ergic interneurons (Wang, 2002). This effective inhibition mediates a competition between the two excitatory pools, with one pool ending up in a high firing attractor state (chosen option), and the other pool staying in a low firing attractor state (unchosen option). Neurons selective for option *o* receive inputs r_o at firing rates proportional to the subjective value of that option, sEV_o (see equation 3, section 5.2.1). The neurons further receive background noise inputs and

currents from other cells in the network. Importantly, the network has very few free parameters that are not otherwise constrained by their biophysical plausibility[1]. The behavior of single units in the network has been described elsewhere (Wang, 2002, Wong and Wang, 2006); here, we focus on predictions suited to investigation with MEG, namely behavior of the summed input currents to all pyramidal cells.

We simulated network behavior using a set of trials with varying sEV_o (as used in the human experiment, below). We sorted trials by overall value (sEV_1+sEV_2; fig 1A top) and value difference (sEV_{chosen}-$sEV_{unchosen}$; figure 1A bottom). In both cases, the network attracted faster to a decision when overall value or value difference were higher, yielding the prediction of decreased reaction times (RTs) under these conditions. We tested this prediction more formally using a multiple regression in which model RTs were predicted as a function of both overall value and value difference; both variables were found to have a negative effect on RT (figure 1B). We then used Morlet wavelets to perform a time-frequency decomposition of network activity on each trial (Tallon-Baudry et al., 1997), and regressed the decomposed data onto overall value and value difference (figure 1C). This revealed a key prediction of the biophysical model: the time at which overall value produced most significant variation in network responses was earlier than the time at which value difference produced variation in network responses. Network transitions typically took several hundred milliseconds to occur, and so most key model predictions were limited to frequencies ranging from approximately 2-10Hz (figure 1C). Overall value had a broadband effect on model activity in the 3-10 Hz frequency range starting soon after selective inputs were delivered to the network (figure 1C, top), whereas value difference had a later and slightly lower-frequency effect, predominantly in the 2-4 Hz range (figure 1C, bottom). If we collapsed across the relevant frequencies (figure 1D), this temporal progression could be clearly seen, and another prediction could be made: on trials where the network model made an error (i.e. $sEV_{chosen}<sEV_{unchosen}$), there was an effect of overall value on the model's activity, but no clear effect of value difference (figure 1D, dashed lines).

Finally, by varying the strength of recurrent excitation in the network model, we found that networks with stronger recurrent excitation relative to inhibition were faster in

[1] The free parameters in the model are the delay for value related inputs to reach the network and for motor response execution; a parameter J(A,ii) that controls the level of recurrent excitation within the network, and the gain parameter k(opt) that determines the scaling of input currents by value. We selected parameters that matched the reaction time distribution of our subjects, but the precise selection of these parameters did not affect the main qualitative predictions of the model. Parameters used are given in section 5.2.4.1.

making decisions but typically made more mistakes (Wong and Wang, 2006), giving rise to a speed-accuracy tradeoff (figure 1E). Assuming that such a mechanism might account for any cross-subject variability in task performance, we could use this to generate model predictions that relate to how network activity might vary across subjects. To do this, we took the parameter estimates from the first-level regression of value difference for each of the different network model instantiations, and regressed these parameter estimates onto the models' median RTs. This yielded the prediction that networks with a higher median reaction time had first a weaker correlate of value difference (blue area in figure 1F), and then a stronger one (red area in figure 1F), at these relatively low frequencies. This is equivalent to saying that the effect of value difference is later to emerge in network models that exhibit slower temporal dynamics.

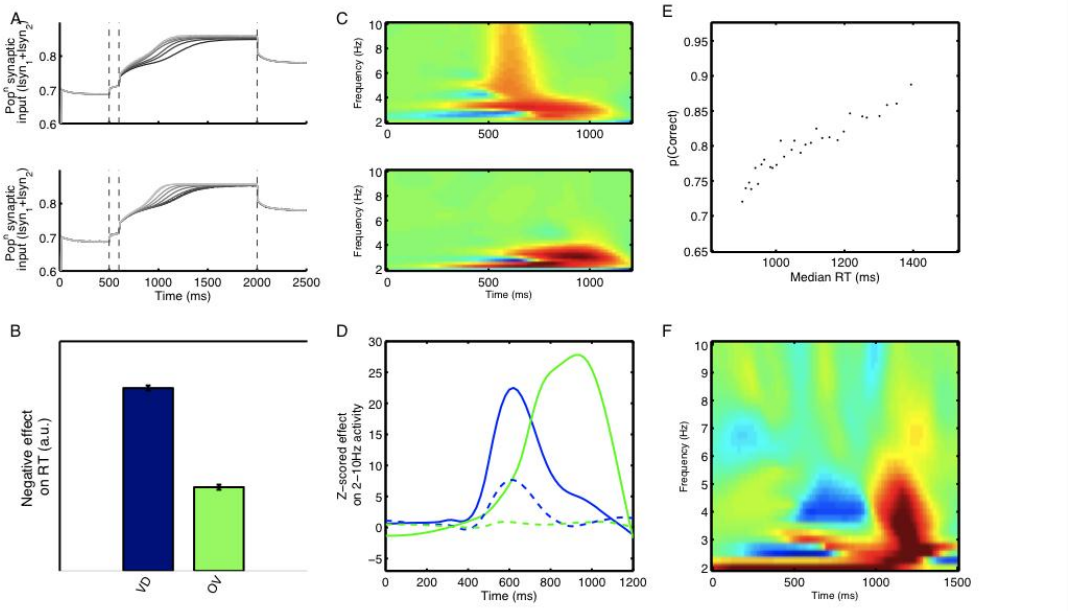

Figure 1. ***Predictions of neural activity from cortical attractor network model*****. (A) Top panel: summed network postsynaptic currents as a function of time through trial, sorted and binned into trials with high overall value (lighter shades of grey) through trials with low overall value (dark grey/black). Bottom panel: As top panel, resorted and binned by value difference between chosen and unchosen options. (B) Effect of value difference (VD) and overall value (OV) on reaction time, estimated using multiple regression (mean +/- s.e. of effect size; Y-axis is flipped, so positive values equate to a negative effect on RTs). (C) Time-frequency spectra of effects of overall value (top panel) and value difference (bottom panel) on network model activity, estimated with multiple regression. Color indicates Z-statistic. (D) Z-scored effect of overall value (on frequency range 3-9Hz; blue lines) and value difference (on frequency range 2-4.5 Hz; green lines); solid lines are correct trials, dashed lines incorrect trials. (E) 'Cross-subject' speed-accuracy tradeoff elicited by varying network connectivity; p(Correct) is proportion of trials on which network chose option with higher value. (F) 'Cross-subject' regression of median reaction time on value difference, derived from biophysical model; the plot reflects the result of a first level regression of value difference onto network activity, and then submitting the parameter estimates from this regression to a second-level regression with model median RT as the independent variable. Color indicates group Z-statistic.**

5.3.2 A distributed network of task-sensitive areas at 2-10 Hz

We used the MEG data collected for the value-guided choice paradigm, presented in **chapter 4**, to test the model predictions. Details of the experimental task and subject behaviour are presented in section 4.1 of this thesis. Briefly, 30 subjects repeatedly chose between two options of varying reward magnitude and probability to obtain monetary reward whilst undergoing MEG. Subjects integrated across both dimensions to guide their choices, and choice behaviour was well described by Prospect theory. As predicted by the biophysical model, subject reaction times correlated negatively with both the *value difference* between chosen and unchosen options, and the *overall value* of the two options (see figure 4.4). Reaction times were additionally slower in earlier trials relative to late trials, and in trials where probability and magnitude advocated opposing choices relative to 'no brainer' trials. Finally, as predicted by the biophysical model, there was a cross-subject speed accuracy tradeoff, with subjects who were on average slower also being more accurate (see figure 4.8).

We used linearly constrained minimum variance beamforming (VanVeen et al., 1997) to spatially filter MEG data to locations in source space. We epoched data with respect to both stimulus onset and subject response; we then focused our analyses on responses in the 2-10 Hz frequency range, in accordance with model predictions. We first used a whole-brain analysis to look for areas showing a main effect of performing the task relative to a pre-stimulus (-300ms to -100ms) or post-response (+100ms to +300ms) baseline. We hypothesized that, in addition to areas important to stimulus valuation such as ventromedial prefrontal cortex, the stimulus-locked analysis would reveal early visual areas involved in basic stimulus processing, and the response-locked analysis would reveal areas involved in visually guided manual movements in parietal and premotor cortices (Johnson et al., 1996), in addition to primary motor areas.

A distributed network of areas was found to be task-sensitive at these frequencies (figure 2). Stimulus-locked, early visual cortex activation (figure 2A) was followed by slowly ramping bilateral activation at the frontal pole and ventromedial prefrontal cortex (figure 2B). Whilst 2-10Hz activity in these frontal regions peaked relatively late in the trial (1000ms after stimulus onset), it ramped from a much earlier point in the trial (see figures 3D and 4, discussed below). Response-locked, a prolonged activation spread from a mid-posterior portion of the superior parietal lobule, which extended medially into the marginal ramus of the posterior cingulate sulcus (figure 2C), to a bilateral medial portion of the mid-

intraparietal sulcus (IPS) (figure 2D). This was followed by bilateral activation of the angular/supramarginal gyri (figure 2D) and right premotor cortex (figure 2E) before finally bilateral inferior frontal sulci and primary sensorimotor cortices (figure 2F) were activated at the time of the response.

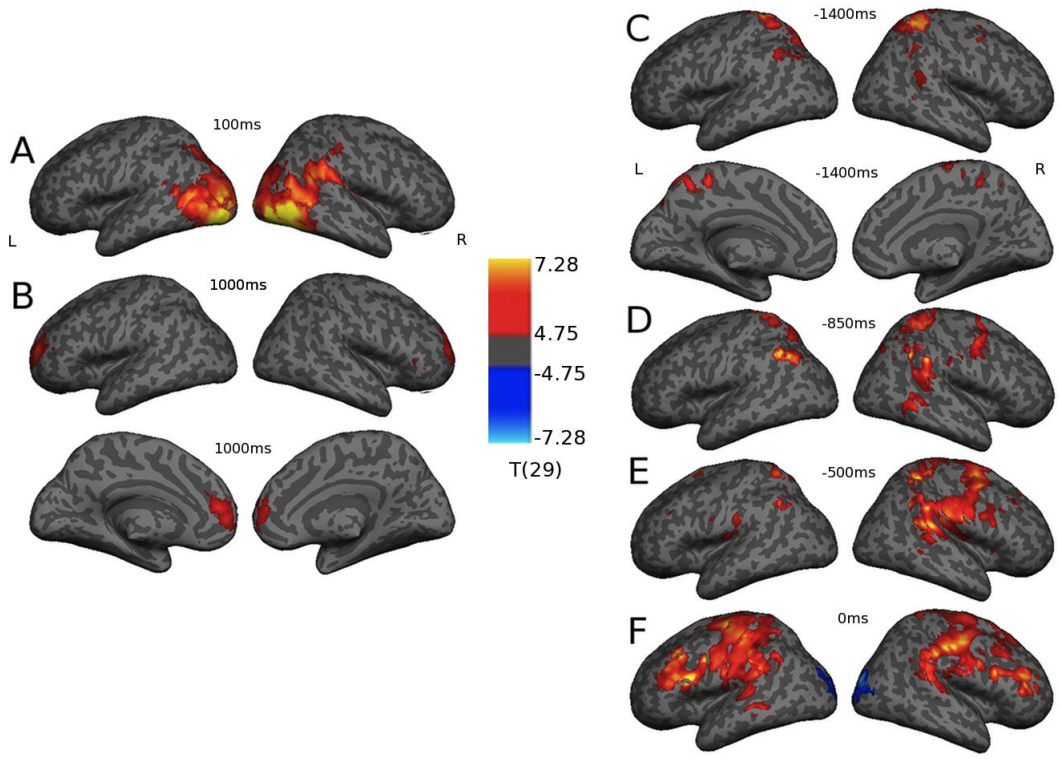

Figure 2. Main effect of task performance on activity in 2-10Hz frequency range. (A)/(B) Stimulus locked activity. Group T-map of effect of task performance relative to a -300 to -100 ms (pre-stimulus) baseline; (A) 100ms post-stimulus, early visual activation (peak T(29)=10.00, 100ms, MNI (40,-74,6)); (B) 1000ms post-stimulus, activation at frontal pole (T(29)=7.23, 1125ms, MNI (22,58,26) and ventromedial prefrontal cortex (T(29)=5.20, 1000ms, MNI (43,60,35)). (C)-(F) Response locked activity. Effect of task performance relative to a +100ms to +300ms (post-response) baseline; (C) 1400ms pre-response, activation at pSPL/posterior cingulate (T(29)=7.05, -1625ms (pre-response), MNI(18,-44,62) and mid-IPS (T(29)=8.20, -525ms, MNI(30,-46,56) (right) and T(29)=7.55,-700ms, MNI (-24,-42,74) (left)); (D) 850ms pre-response, activation at angular/supramarginal gyri (T(29)=8.46, -725ms, MNI (56,-50,40) (right) and T(29)=8.69, -725ms, MNI (-50,-60,42) (left)); (E) 500ms pre-response, premotor activation (T(29)=7.35,-450 ms, MNI (38,-2,64); (F) time of response, activation at inferior frontal sulci (T(29)=8.02, 0ms, MNI (-54, 12, 28) (left) and T(29)=7.55, -75ms, MNI (48,10,30) (right)) and sensorimotor cortices (T(29)=7.57, -75ms, MNI (-50,-28,58) (left) and T(29)=8.02, 0ms, MNI (-54,12,28) (right). All images are thresholded at T>4.75 (p<5*10^{-5} uncorrected) for display purposes.

5.3.3 Value-dependent activity in posterior parietal and medial prefrontal cortex and comparison to the network model

Having isolated areas that showed changes in activity relative to baseline at 2-10 Hz frequencies, we then examined whether activity within these regions co-varied with

decision values, and where this activity matched with predictions derived from the biophysical decision model. Importantly, by selecting regions based on the main effect of task versus baseline, we ensured we would not be subject to a selection bias when examining these regions for value-related activity. We also investigated activity in several *a priori* defined areas commonly found to be important in functional MRI studies of decision making, bearing in mind that value correlates might not be restricted to regions showing a main effect of task versus baseline. The key predictions (figure 1) we sought to test were: (i) the temporal evolution from an 'overall' value signal to a 'difference' value signal; (ii) a difference in the frequency of the response, with value difference dominating responses at lower frequencies than overall value; (iii) the presence of an overall, but little or no difference signal, on error trials; and (iv) the cross-subject prediction that subjects with longer median RTs would exhibit first a weaker, and then a stronger, value difference signal.

We found that activity in the right posterior superior parietal lobule (pSPL) bore several hallmarks of the biophysical model (figure 3). On trials where subjects chose the option with higher subjective value ('correct' trials), activity in pSPL showed a broad correlate of overall value between 3 and 10 Hz ($p<0.0005$, permutation test, cluster corrected for multiple comparisons across time), followed by a lower frequency (2-4Hz) correlate of value difference (figure 3A) ($p<0.01$, corrected), as predicted by the model (cf. figure 1C). When we collapsed across the relevant frequencies (figure 3B), activity on these correct trials differed from error trials; error trials showed a positive correlate of overall value (dashed blue line, fig 3B) ($p<0.05$, corrected), but no such positive correlate of value difference (dashed green line in fig 3B; compare with figure 1D) ($p>0.5$). We also tested the model's cross-subject predictions that might explain why subjects with longer mean RTs were more accurate in performing the task. We found that in the pSPL, these subjects had a more negative correlate of value difference initially, followed by a more positive one (figure 3C), matching predictions elicited by varying the degree of self-excitation relative to inhibition within the model's selective neuronal pools (cf. figure 1E/F).

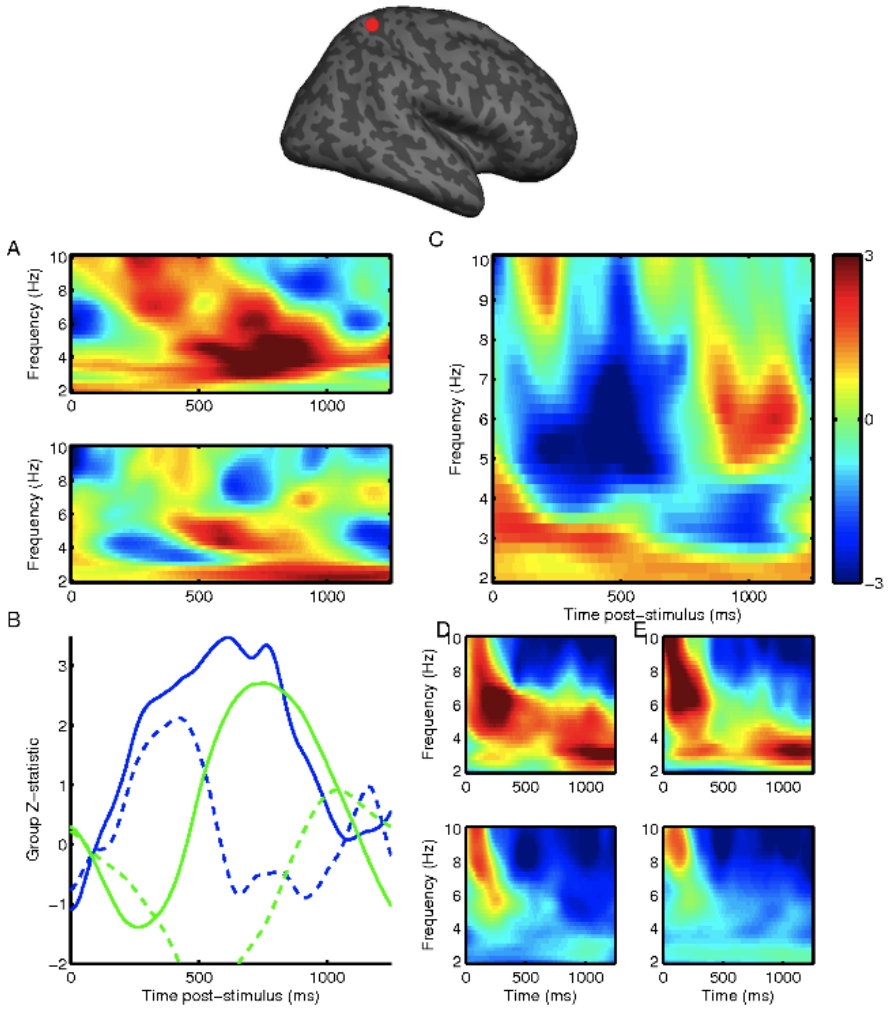

Figure 3. pSPL (MNI 18, -44, 62mm) shows several value-related hallmarks of the biophysical network model. (A) Time-frequency spectra of effects of overall value (top panel) and value difference (bottom panel) on activity in pSPL, estimated using multiple regression. Color indicates group Z-statistic. (B) Effect of overall value (3-9Hz, blue) and value difference (2-4.5 Hz, green) on correct/error trials (solid/dashed lines respectively). (C) Cross-subject effect of reaction time on value difference time-frequency spectrum, analysis equivalent to figure 1F. Color indicates group Z-statistic. (D) Main effect of task performance relative to pre-stimulus baseline on first half of trials (top panel) and second half of trials (bottom panel). (E) Main effect of task performance on trials where reward magnitude and probability advocate opposing choices (top panel), and 'no brainer' trials (bottom panel).

We also investigated whether the main effects of task performance in this region was affected by factors shown behaviorally to modulate reaction time independently of value. We looked for changes in activity in early trials relative to late trials (where reaction time was speeded, as discussed in section 4.1.3.2), and also compared activity in trials where reward magnitude and probability advocated opposing choices (in which a 'common currency' valuation may need to be formed) with activity on 'no brainer' trials (where an

additional bonus to reaction time was present beyond that explained by overall value or value difference, see section 4.1.3.2). There was some difference between the patterns of activity in pSPL on these trials; an increase in 2-5 Hz power relative to baseline that was present on the first half of trials (figure 3D, top panel) was largely absent on the second half of trials (figure 3D, bottom panel). A similar distinction could be seen between activity on trials where reward magnitude and probability advocated opposing choices, and a 'common currency' representation might need to be formed (figure 3E, top panel) and 'no brainer trials' (figure 3E, bottom panel).

We also investigated value-related activity in ventromedial prefrontal cortex (VMPFC), focusing our analyses on a subregion that has often been shown to signal value-related metrics during decision tasks (Knutson et al., 2005, Plassmann, O'Doherty and Rangel, 2007, Tom et al., 2007, Serences, 2008, Boorman et al., 2009, Gershman, Pesaran and Daw, 2009). In this region, there was an even more striking distinction between those situations where subjects would be more deliberative and exhibit slower RTs (figure 4D and E, top panels) versus later (figure 4D, bottom panel) or 'no brainer' (figure 4E, bottom panel) trials. VMPFC recruitment steadily decreased through the task, as could be seen more clearly when trials were further subdivided into separate quartiles of the experiment (figure 5). We found that this region transitioned from signaling overall value ($p<0.05$, corrected) to value difference ($p<0.05$, corrected) (figure 4A/B) specifically if we restricted our analysis to the first half of trials in which it was task active (figure 4D). When we directly contrasted the effect of overall value and value difference on early and late trials (figure 4C), we found that only the value difference signal was significantly stronger on earlier trials in this region ($p<0.05$, corrected). There was not a significant effect of either overall value or value difference on error trials (figure 4D, dashed lines), although the somewhat weaker signals in this region relative to pSPL may result from the relative insensitivity of MEG to deep, anterior sources, as opposed to posterior, superficial ones (Hillebrand and Barnes, 2002, Marinkovic et al., 2004), and from the analysis including only half the number of trials.

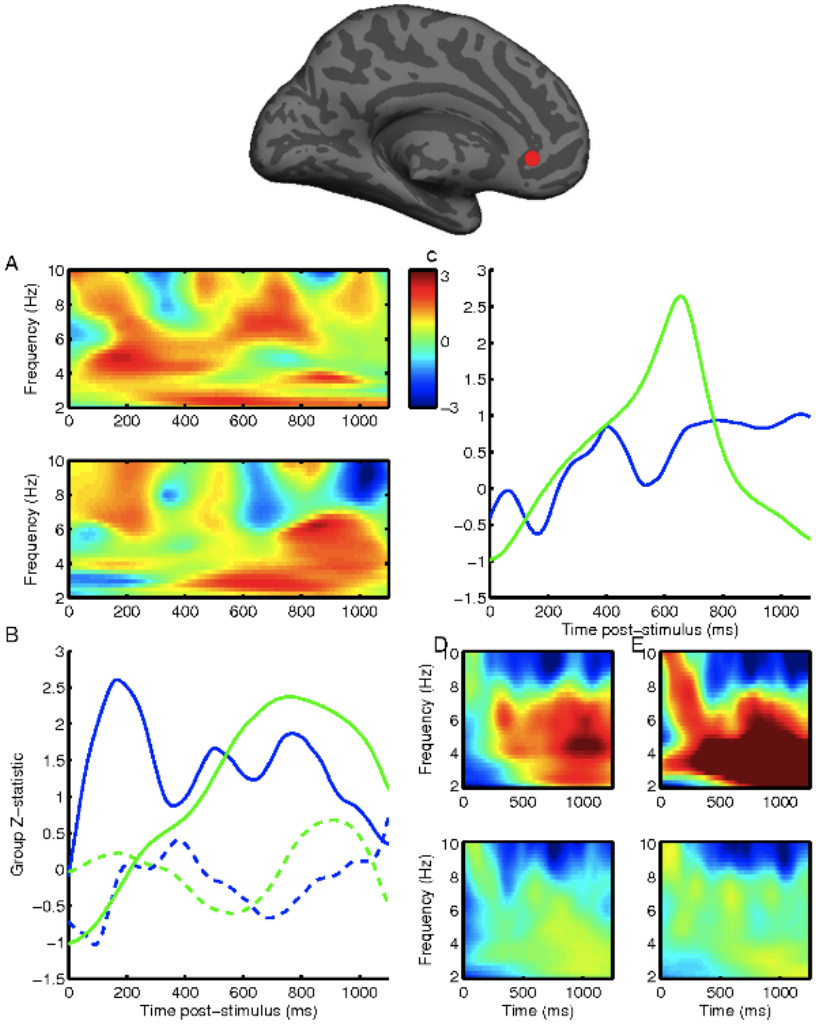

Figure 4. VMPFC (MNI 6, 28, -8 mm) shows similarities to biophysical model, specifically on early trials. (A) Time-frequency spectra of effects of overall value (top panel) and value difference (bottom panel) on activity in VMPFC for first half of trials, estimated using multiple regression. Color indicates group Z-statistic. (B) Effect of overall value (3-9Hz, blue) and value difference (2-4.5 Hz, green) on correct/error trials (solid/dashed lines respectively), for first half of trials only. (C) Direct contrast of effects of overall value (3-9Hz, blue) and value difference (2-4.5 Hz, green) between first half and second half of trials. Positive Z-statistics denote stronger value-related signal in first half of experiment. (D) Main effect of task performance relative to pre-stimulus baseline on first half of trials (top panel) and second half of trials (bottom panel). (E) Main effect of task performance on trials where reward magnitude and probability advocate opposing choices (top panel), and 'no brainer' trials (bottom panel).

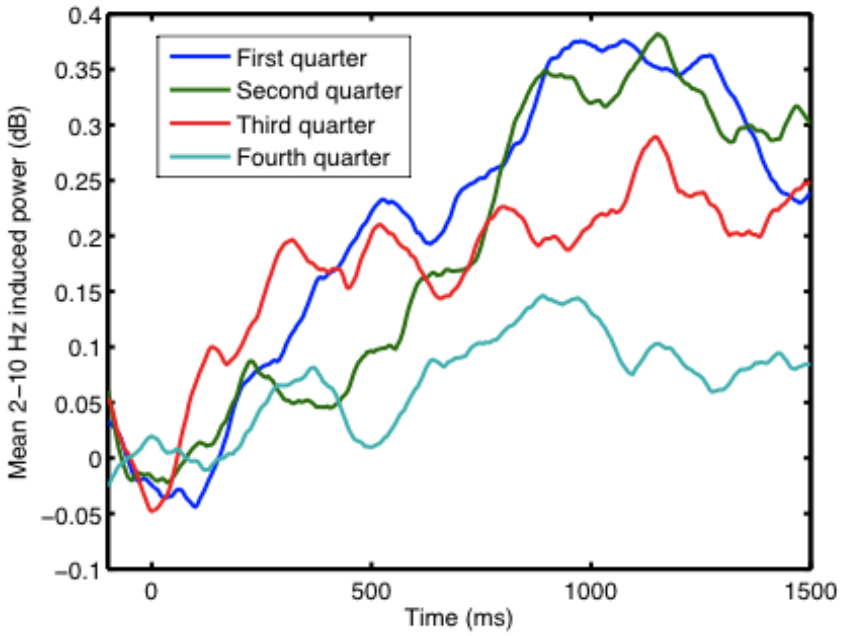

Figure 5. Effect of task performance, stimulus-locked, on activity in 2-10Hz range in VMPFC (relative to -300ms to -100ms pre-stimulus baseline). Trials are subdivided into first, second, third and fourth quarters of the experiment. A steady decrease in the recruitment of VMPFC through the experiment can be seen.

5.3.4 Value-related effects and difficult vs. no-brainer comparisons in other regions

One possible concern with the differences between the first and second halves of the experiment is that it might reflect more trivial cognitive differences, such as subject fatigue, rather than a change in the cortical networks underlying choice behavior. To address these concerns, we performed an additional whole-brain analysis in which we searched for regions coding more strongly for value difference in the second than in the first half of the experiment - that is, the opposite pattern of activity witnessed in VMPFC. A bilateral portion of the anterolateral intraparietal sulcus - more lateral to the main effect pSPL activation described above - selectively reflected value difference in the second half of trials (figure 6). In this region, there were also no clear differences between the main effect of task performance on early versus late trials, or on harder trials versus 'no brainers' (figure 7).

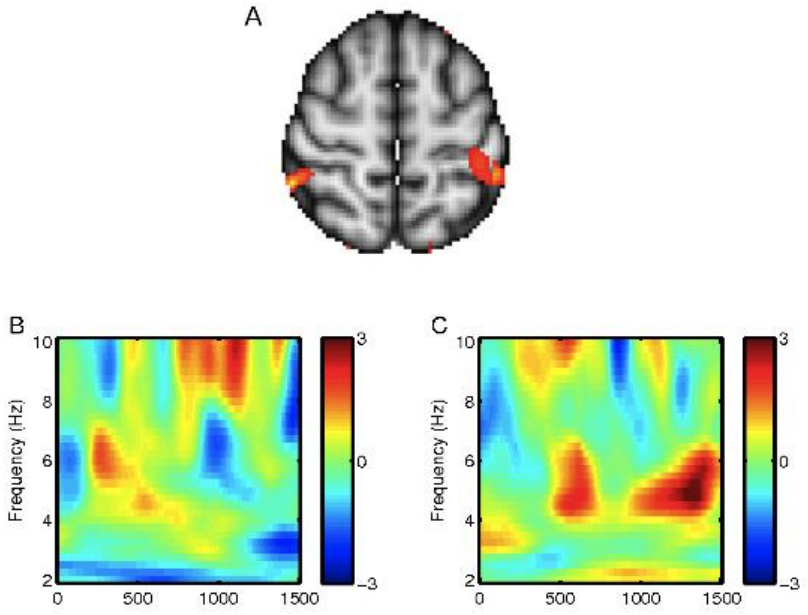

Figure 6. Lateral IPS correlates with value difference more strongly in second half of experiment. (A) Results from a whole brain analysis looking for brain regions coding value difference more strongly in the second half of the experiment than the first half. Cross-section at MNI z=58mm, t=725ms post-stimulus onset. Thresholded at T(29)>2.3 (peak T(29)=3.79 (675 ms post stimulus), MNI (50,-46,48) (left); T(29) = 2.69 (750ms post stimulus), MNI (-50,-38,56) (right)). (B) Time-frequency effect of value difference in left IPS peak in first half of experiment; color indicates group Z-statistic. (C) Time frequency effect of value difference at same location in second half of experiment.

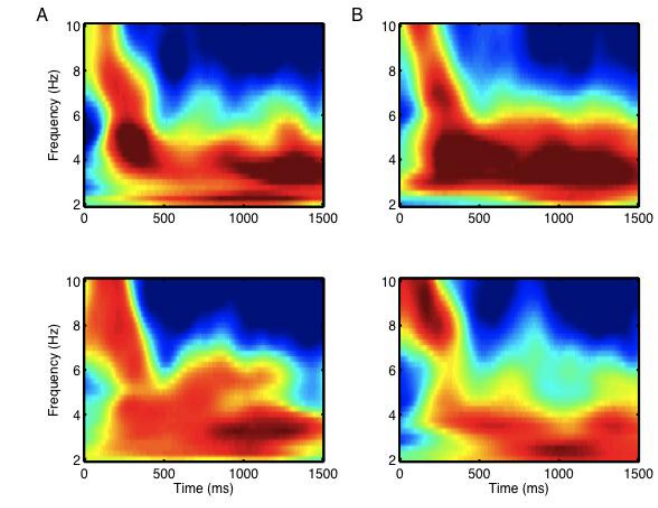

Figure 7. Lateral IPS main effect responses are stable throughout experiment. Responses are from left lateral IPS peak presented in figure 6. (A) Main effect of task performance relative to pre-stimulus baseline. Top panel=first half of trials; bottom panel=second half of trials. (B) Main effect of task performance relative to pre-stimulus baseline. Top panel=harder trials; bottom panel='no brainer' trials. Color indicates group Z-statistic.

Lastly, we also searched for effects of value in other regions identified in the main effect contrast of task vs. baseline (figure 2), and in several regions defined *a priori* from previous fMRI studies of value-based choice. In these analyses, we found that several areas exhibited value-dependent activity, but none of these regions matched well with predictions from the biophysical decision model (figure 8). We hypothesize that the value correlates in these regions might be better described by appealing to their role in other computational processes that are likely to covary with value, such as attention or response preparation.

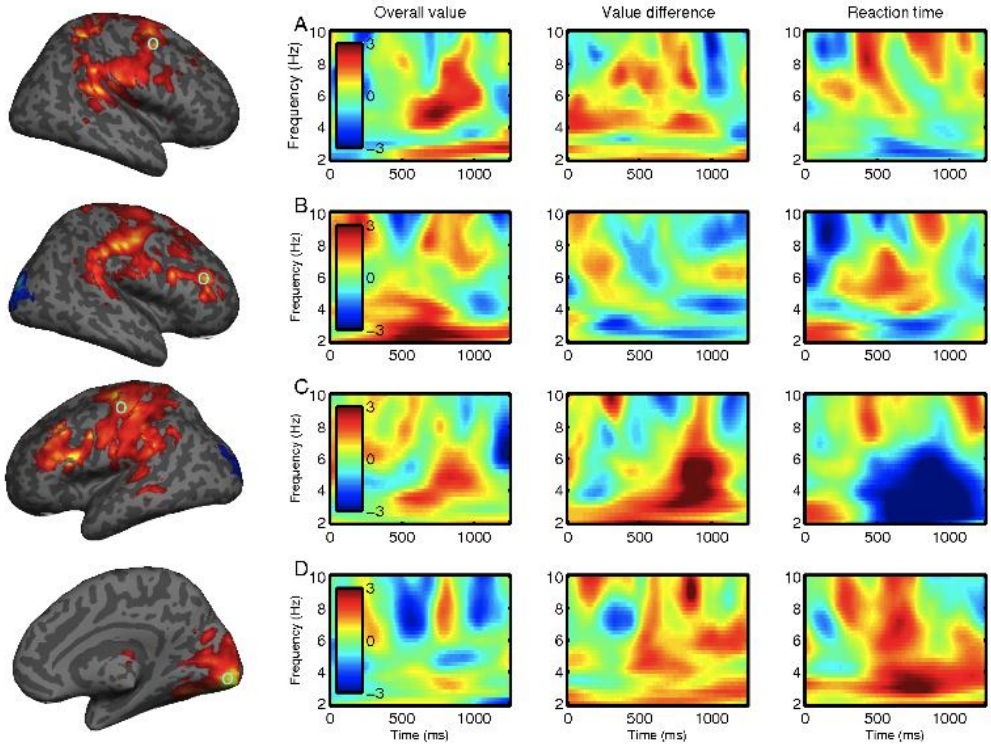

Figure 8. Value-related and reaction-time dependent signals in other cortical regions. Value correlates can be seen elsewhere, but do not match with predictions from the biophysical model. Time frequency spectra show group Z-statistics for effects of overall value (left), value difference (middle) and reaction time (right) on cortical activity in 2-10 Hz range, estimated using multiple regression. (A) Right premotor cortex, MNI (38, -2, 64 mm) – shows a correlate of overall value, but no difference signal; (B) right inferior frontal sulcus, MNI (50, 10, 30 mm) – shows a correlate of overall value, but no difference signal; (C) left sensorimotor cortex, MNI (-50, -28, 58 mm) shows a correlate of value difference, but this signal is noticeably at the same time as a corresponding negative correlate of reaction time, suggestive of a role in response execution; (D) left primary visual cortex, MNI (-12, -94, -2 mm) –shows a weak correlate of value difference, but a stronger positive correlate of correlate of reaction time, suggestive of increased attentional demands on trials of longer duration.

5.4 Discussion

The cortical correlates of value during decision under risk are typically spread over a distributed network of areas, but the unique contribution of each of these areas to choice is unclear. A region involved in value comparison should receive inputs relating to the value of available options, and transform these inputs into a choice. We used a biophysically plausible model that exhibits this property to derive novel predictions of the temporal dynamics of cortical activity. This led to a set of characteristic time-varying signals that could be used to isolate relevant brain regions; these responses typically occurred in low frequencies (<10 Hz), consistent with a slow integrative process. This is in agreement with predictions that long latency responses in regions anatomically distant from sensory cortex will typically exhibit low frequency responses (Hari, Parkkonen and Nangini, 2010). In source-reconstructed MEG data, a distributed network of areas were task-sensitive at these frequencies, but only pSPL and VMPFC closely matched predictions of the biophysical model, with the latter doing so selectively in trials early in the experiment.

A key feature of the biophysical model is the ability to slowly integrate value-related inputs, afforded by its recurrent excitatory structure and long synaptic time constants mediated by NMDA receptors. It is not immediately obvious that value comparison should be subject to a process of integration in the same way as a noisy sensory stimulus. However, the observed distribution of reaction times fits well with a process of integration, as has been investigated more closely in numerical comparison tasks, in which a drift diffusion model (DDM) readily fits subject RTs (Sigman and Dehaene, 2005). Notably, in our task, a DDM would be unlikely to predict the observed effect of overall value on subject RTs or on neural activity, as the rate of integration in a DDM is typically determined by the difference between the two options. We tested this more formally by comparing the biophysical model with predictions from several other commonly used mathematical models of decision making (see supplementary information). Whichever model is used, we hypothesize that whereas in a perceptual decision task noise may be intrinsic to the stimulus, in a numerical or value comparison task noise will arise from within the nervous system, and so an integration process may still be required to successfully compare alternatives.

The predictions from the model also form a striking example of the distinction between two types of representation - 'content' and 'functional' representations - in cortical circuits (deCharms and Zador, 2000). To the external observer, recording with an imaging technique (or an electrode), the content of the network first appears to 'represent' the

overall value and later the value difference between the two options. By contrast, the *functional* representations in the network - those used by the brain - are quite different. There is a representation of option values on the input to the network, and a representation of choice on the outputs of the network, as should be decoded by a suitable downstream observer. The reason that the network shows value-related activity is simply that the same network transitions occur faster on high value and high value difference trials. Hence, whilst neural activity in the network may covary with the overall value and value difference, this content need never be decoded by another brain region. Thus, the extent to which the network can be said to functionally 'represent' these two quantities in a meaningful way is questionable (deCharms and Zador, 2000).

The region in pSPL isolated as matching with model predictions is close to the cytoarchitectonic region hIP3 (Scheperjans et al., 2008), which may be the human homologue of the medial intraparietal area (MIP). It is also referred to as IPS4 (Swisher et al., 2007) and DIPSA (Vanduffel et al., 2002), which resembles macaque MIP (Mars et al., 2011). In the macaque, this region is often implicated in visually-guided movements of the forelimbs (Johnson et al., 1996). It may therefore have a role in integrating information in order to guide limb movements that is analogous to the role of LIP in generating saccades.

The region in VMPFC is often found using fMRI to be responsive to the value of stimuli during decision tasks (Knutson et al., 2005, Plassmann, O'Doherty and Rangel, 2007, Tom et al., 2007, Serences, 2008, Boorman et al., 2009, Gershman, Pesaran and Daw, 2009), but its precise role has been debated (Kable and Glimcher, 2009, Noonan et al., 2010), perhaps as a result of the relative absence of published single-unit recording data in comparison to the nearby lateral orbitofrontal cortex (Padoa-Schioppa and Assad, 2006, Rushworth and Behrens, 2008). In early trials, this region was found to transition from signaling overall value to signaling value difference. Strikingly, this same transition was also recently found in single-unit recordings from the most ventral portion of the striatum (Cai, Kim and Lee, 2011), which receives a particularly dense projection from VMPFC (Haber et al., 2006). Similarly to the present study, this task required monkeys to combine two stimulus properties to form their decision, namely the reward magnitude and the delay to reward delivery. In our task, VMPFC was selectively activated in trials where subjects had to combine probability and magnitude information to choose accurately. This is also consistent with the finding that lesions to this area, but not nearby lateral orbitofrontal cortex, produce impairments in value comparison (Noonan et al., 2010), and

more specifically produce changes in tasks where multiple dimensions have to be considered in forming a choice (Fellows, 2006). In these tasks, VMPFC-lesioned subjects have been found to switch to a strategy of accumulating information about each alternative, rather than comparing alternatives across attributes. This is consistent with the idea that the VMPFC might be particularly important for selection over behavioral goals, whereas other brain regions (for example, in parietal cortex) might frame decisions in the context of possible actions. In the absence of a VMPFC, subjects might be less able to compare options by selecting over internal goals, but will still be able to make a decision by slowly constructing an action value for each available option.

Some previous studies have attempted to apply a modeling approach to capture signals from distributed cortical regions during choice, measured using BOLD fMRI. These studies have made predictions based either on drift diffusion models (Basten et al., 2010) or biophysically plausible networks (Rolls, Grabenhorst and Deco, 2010) but the predictions of these models are heavily dependent upon whether BOLD signal is assumed to reflect activity from all timepoints including after a decision has been formed (Rolls, Grabenhorst and Deco, 2010), or whether it only reflects activity until the decision threshold is reached (Basten et al., 2010). Moreover, several key predictions of these models also relate to how their activity evolves over time as a decision is made, and the slow hemodynamic response means fMRI is limited in how well it can tease apart these predictions of temporal dynamics. We argue that it is important to use a time-resolved technique, such as MEG, to test these predictions. Other studies have used MEG to investigate the temporal dynamics of activity during choice behavior, but these studies (Guggisberg et al., 2007) have typically focused only on the main effect of task performance as opposed to value-related modulations, and have examined activity in frequencies higher than those that would be associated with a slow integrative process.

In the present study, we used a model designed to make predictions of single-unit activity in a particular cognitive process in order to derive novel predictions of the temporal dynamics of activity that might be seen with MEG (or electroencephalography (EEG)). Biophysically-inspired models have previously been used to infer the structure of connections between or within different cortical areas from M/EEG data (Kiebel et al., 2009). However, these studies have not made been used to make any inferences about the specific neuronal mechanism underlying a particular cognitive process, as we have proposed here. The key feature of the present model is that it performs the necessary

computation of transforming value-related inputs into a choice, and does so in a way that has captured single-unit activity during decision tasks. However, the application of this biophysical modeling approach may not be limited to decision making paradigms. Novel predictions might be derived from biophysical models already designed to capture, for example, single unit data in inhibitory control (Lo et al., 2009) or working memory processes (Compte et al., 2000). In models of working memory, for instance, gamma-band (30-70 Hz) responses can be elicited, and parametric modulation of input to these models may explain variation in gamma-band frequencies seen during working memory tasks in frontal cortex (Meltzer et al., 2008). Alternatively, by varying internal parameters of a biophysical model, novel predictions can be derived of the effects of this variation on cortical responses that are measurable with M/EEG (Brunel and Wang, 2003). Because these parameters relate to specific biophysical properties such as the density of network connectivity or the concentration of a specific neurotransmitter, they can be directly related to cross-subject variation in these properties, for instance via local measurements of neurotransmitter concentrations (Muthukumaraswamy et al., 2009), or perhaps genetic polymorphism or pharmacological challenge.

5.5 Supplementary information

5.5.1 A comparison of different decision models

It is important to note that the main objective of the present study was not to make specific claims about the accuracy (or otherwise) of the biophysical decision model, but was instead to use a whole-brain imaging technique to compare predictions from this model with activity recorded from across the entire cortex, and infer which cortical regions were involved in value comparison. It was not our intention to perform a strong 'model comparison' to infer that the biophysical decision model was better at explaining cortical activity than other decision models.

Nevertheless, alternative models are commonly used to explain neural activity in other studies, so it is interesting to see which of the predictions from the biophysical decision model are present in other classes of decision model, particularly as they have been shown to be closely related in terms of behavioural predictions[4]. We considered three other classes of decision model, using the simplified descriptions taken from [4]. These were:

(i) the drift diffusion model (DDM) (where the difference in subjective value determined the drift rate of the decision particle)[5];

(ii) a race model, in which two (non-competing) units accumulate information for each alternative (where the subjective value of each option determined the rate of accumulation for each unit)[6];

(iii) a feedforward inhibition model, in which value-related inputs both excite the selective unit and inhibit the non-selective unit[7].

Results from this analysis are presented in figure 9. We found that the simple DDM made starkly different predictions to the biophysical decision model, in that it predicted only an effect of value difference on neural activity and reaction times, and no effect of overall value (fig 9B). This is, in many ways, to be expected, as the two values are subtracted before being used to determine the drift rate in the DDM, so the model has *no information* about the overall value of the trial. Similar predictions were made by an extended DDM[8] (although we did not consider whether alternative DDMs, such as those containing a time-varying urgency signal, would make different predictions).

By contrast with the DDM, the race model made the opposite prediction: a strong effect of overall value on neural activity, but no effect of value difference (fig 9C). A similar set of predictions were made by the feedforward inhibition model (fig 9D). Thus, the temporal dynamics shown by the biophysical decision model appear to be unique in exhibition a transition from an overall value signal to value difference. Such predictions may also generalize to other models of mutual inhibition[9], but we did not test these predictions here.

We make this claim with a few caveats. We did not exhaustively test the entire parameter space of the models presented in figure 9 (instead we selected parameters that provided a reasonable fit to our behavioural data), nor did we considered more sophisticated variants of these models, of which there are many examples in the literature. We also assigned the same noise structure to data prior to the decision and after reaching the decision bound, so that the measured signal was related to decision-related activity rather than noise. An exhaustive test of variations of model parameters and model structures is beyond the scope of this paper. It is possible that further adaptations or variations of the DDM, race and feedforward inhibition models might be able to make similar predictions to the biophysical decision model, but we leave this topic to future studies.

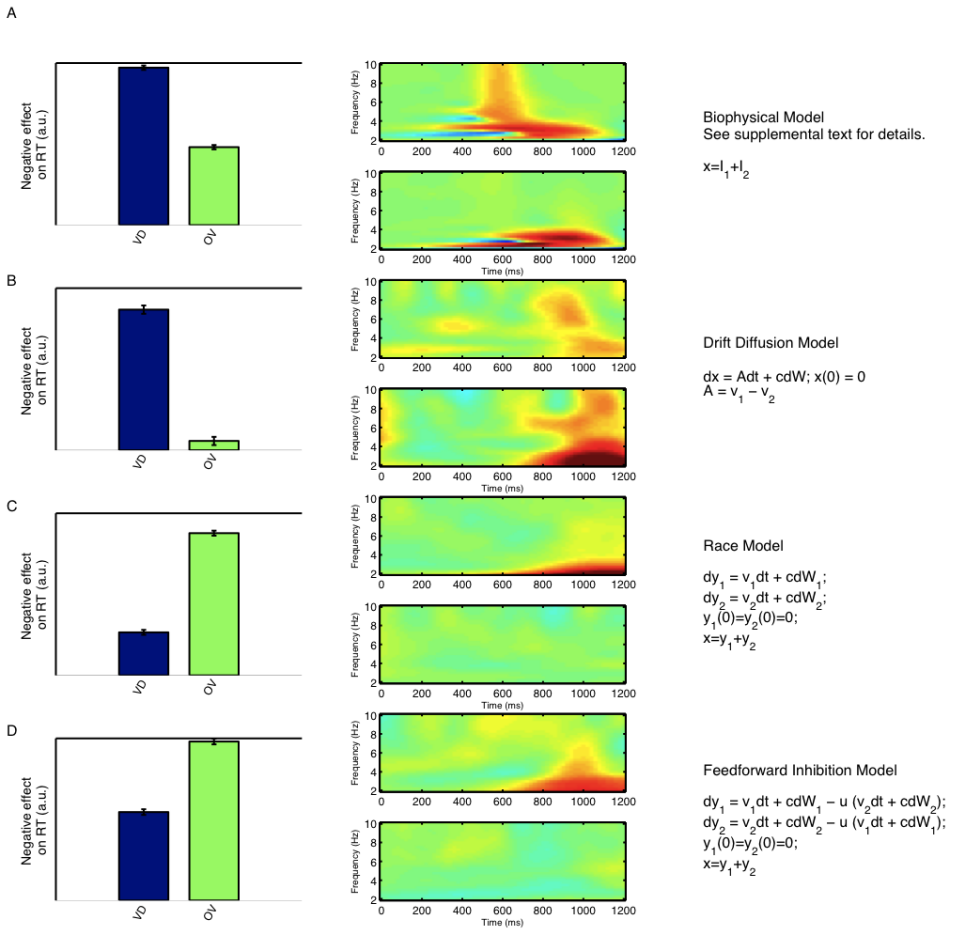

Figure 9. Comparison of different decision models. Left column: Effect of value difference (VD) and overall value (OV) on reaction times for each model, estimated using linear regression. Y-axis is flipped, so positive values equate to a negative effect on reaction time. See figure 1B. Middle column: Effect of overall value (top panel for each model) and value difference (bottom panel for each model) on time-frequency decomposed model data. See figure 1C. Right column: Model details. For further details of model parameters, see (Bogacz et al., 2006). In each case, x is the variable submitted to time-frequency decomposition and linear regression; cdW is white noise, normally distributed with mean 0 and variance c2dt. v1 and v2 are values of options 1 and 2. (A) Biophysical model, as used in main paper; (B) Simple drift diffusion model; (C) Race model; (D) Feedforward inhibition model.

Chapter 6: Associative learning of social value

Our decisions are guided by information learnt from our environment. Such information can come from personal experiences of reward or vicariously – derived from the actions of social partners. It is often assumed that these sources of information rely on different neural systems and are processed in very different fashions – competing with one another to drive decision-making. In this chapter, we demonstrate that social and reward-based information may be learnt using similar computational strategies - key computational variables for learning in the social and reward domains are processed in a similar fashion, but in parallel neural processing streams. Whereas errors in prediction of reward are reflected in BOLD fMRI signal measured from ventral striatum, errors in predictions of social behaviour are found in brain regions previously implicated in theory of mind, such as dorsomedial prefrontal cortex and the temporoparietal junction. Two neighbouring divisions of the anterior cingulate cortex are central to learning about social and reward-based information, and for determining the extent to which each source of information guides behaviour. At the time of the decision, however, information is combined from social and non-social sources in a common region – ventromedial prefrontal cortex – to guide current behaviour.

6.1 Introduction

Learning about the actions of others generates a rich source of information for making decisions in a social setting (Maynard Smith, 1982, Fehr and Fischbacher, 2003). It is widely held that such social learning is distinct from other forms of learning in its mechanism and neural implementation. Social learning and evaluation mechanisms are often assumed to compete with simpler mechanisms, such as associative learning, to drive behaviour (Delgado, Frank and Phelps, 2005). Recently, however, neural signals have been observed during social exchange reminiscent of prediction error signals seen routinely in associative learning paradigms (King-Casas et al., 2005, Hampton, Bossaerts and O'Doherty, 2008, Burke et al., 2010). Here, we test whether social information can be acquired by the *same* associative processes that are commonly assumed to underlie experienced-based learning, even in situations where social partners may have different motives. We use BOLD fMRI to examine whether learning parameters that are key to such associative processes are coded in a similar fashion in social and non-social domains, and whether the neural processing of learning in the two domains recruits common or specialised neural circuitry. Finally, we examine how the two sources of information are combined together to guide current decisions.

Computational models of reinforcement learning (RL) have had considerable success in predicting behavioural data in associative learning tasks outside the social domain. As discussed in **chapter 3**, the simplest RL models suggest that when new information is observed the *value* of the action or stimulus is updated by the product of the *prediction error* and the *learning rate* (Sutton and Barto, 1998). The prediction error represents the difference between expected and actual outcomes. The learning rate represents the expected *value of information* available at the current trial, and depends on the agent's current level of understanding of the environment (Dayan, Kakade and Montague, 2000, Courville, Daw and Touretzky, 2006, Behrens et al., 2007). In situations where the agent is *uncertain* about the environment, new information is more valuable. In keeping with this idea, humans optimally adapt their learning rate when moving between a *stable* environment (in which they can be confident in their understanding, and consequently have a low learning rate) and a *volatile* environment (in which they must be uncertain, and consequently have a high learning rate) (Behrens et al., 2007).

Neural responses have been found that appear to code for key parameters in such models (Schultz, Dayan and Montague, 1997, Waelti, Dickinson and Schultz, 2001, O'Doherty et al., 2004, Tanaka et al., 2004, Tobler, Fiorillo and Schultz, 2005, Daw et al., 2006, Behrens et al., 2007, Matsumoto et al., 2007). Reward prediction error signals are reported in dopamine neurons in the ventral tegmental area (VTA) of the macaque monkey, and are considered a critical neural correlate of an RL-like learning strategy (Schultz, Dayan and Montague, 1997, Waelti, Dickinson and Schultz, 2001, Bayer and Glimcher, 2005). Such brain regions are often too small to detect in FMRI data without dedicated technical strategies (D'Ardenne et al., 2008) but reward prediction error signals have been reported in the striatum, a key projection target of the VTA (O'Doherty et al., 2003, O'Doherty et al., 2004, Haruno and Kawato, 2006). FMRI correlates of the learning-rate in a reward association task have been reported in the sulcal division of the anterior cingulate cortex (ACC) (Behrens et al., 2007). At the time when the decision is made, FMRI signals correlating with the RL-derived *expected value* of the chosen action have been reported in ventromedial prefrontal cortex (VMPFC) (Daw et al., 2006, Hampton, Bossaerts and O'Doherty, 2006).

In the present study, we designed an experiment in which we could look for distinct correlates of prediction error, learning rate and expected value in both reward-guided and social domains. *Reward-guided* learning was framed in the same context as in a previous experiment (Behrens et al., 2007); the reinforcement learning model tracked the *probability of reward* on one of two options. *Social* learning was framed in

terms of tracking the intentions of another individual; the reinforcement learning model tracked the *probability that a social partner would give reliable advice* at each trial.

Unlike previous experiments, which have identified regions of the brain involved in social inference on the basis of a comparison between different conditions (Fletcher et al., 1995, Rilling et al., 2004), we here provided *both* social and non-social information to subjects within the same condition. Crucially, we made sure that the regressors for each source of information were decorrelated from one another, allowing us to distinguish the neural correlates of social *and* non-social inference within a single task design. In line with previous studies, we found that ventral striatum and ACC sulcus correlated with prediction error and learning rate respectively for reward-guided learning. By contrast, regions previously implicated in theory of mind (dorsomedial PFC/temporoparietal junction) (Amodio and Frith, 2006, Saxe, 2006) and social valuation (ACC gyrus) (Rudebeck et al., 2006a) correlated with prediction error and learning rate for social inference. At the time of the decision, both sources of information were combined in VMPFC to guide behaviour at the current trial.

6.2 Methods

6.2.1 Experimental task

Full details of the task design and subject behaviour are given in **section 4.2**. Briefly, in order to compare subject learning strategies for social and reward-based information, we constructed a task in which each trial outcome revealed information both about likely future trial outcomes (reward-based information) and about the level of trust that should be assigned to future advice from a confederate (social information). 24 subjects performed the decision–making task whilst undergoing FMRI, repeatedly choosing between blue and green rectangles, each of which had a different point score (reward magnitude) available on each trial. The chance of the correct colour being blue or green depended on the recent outcome history. Prior to the experiment, subjects were introduced to a confederate who would play the game alongside them. At each trial, the confederate would choose between supplying the subject with the correct or incorrect option, unaware of the number of points available. The subject's goal was to *maximise* the number of points gained during the experiment. In contrast, the confederate's goal was to ensure that the eventual score would lie within one of two pre-defined ranges, known to the confederate but not the subject. The confederate might therefore reasonably give consistently helpful or unhelpful advice, but this advice might change as the game progressed. Both the subject and the confederate were made aware of the task rules in an in-depth briefing at which both players were present.

During the experiment, the confederate was replaced by a computer that gave correct advice on a prescribed set of trials. Subjects knew that the trial outcomes were determined by an inanimate computer program, but believed that the social advice came from an animate agent's decision. Following the experiment, subjects were debriefed to ensure that they maintained the belief that they were playing with a human confederate throughout.

Subjects needed to combine information from three sources to make successful decisions: (i) the *reward magnitude* of each option (generated randomly at each trial); (ii) the likely correct response (blue or green) based on *their own experience* of how frequently these options yielded reward; and (iii) the *confederate's advice*, and how trustworthy the confederate currently was. Whenever a new outcome was witnessed, subjects could update their estimate of the reward environment depending on the colour of the outcome, but could also update their trust in the advice of the confederate depending on whether the advice at the current trial was good.

Optimal behaviour in this task requires the subject to track the probability of the correct action and the probability of correct confederate advice independently, and to combine these two probabilities into an overall probability of the correct response, as outlined in **section 4.2.2.4**. The overall probability of each response being correct should then be multiplied by the magnitude of reward available to give the Pascalian value. The subject should select the response with the greater value.

We have previously described a Bayesian reinforcement learning (RL) approach for optimal tracking of reward probabilities in a changing environment (Behrens et al., 2007). In this study, we used this model to generate the optimal estimates of outcome probability both based on previous outcomes, and based on past confederate advice.

6.2.2 FMRI experimental design

We designed schedules for the trial outcomes and the accuracy of confederate advice such that they both went through periods of stability and of volatility (as outlined in **section 4.2.3.1**). We derived optimal estimates of the volatility of each source of information from the Bayesian RL model. These estimates represent the respective values of the same outcomes for learning about the two different pieces of information. Depending on the recent stability of the two information sources, any single outcome may be of high value for learning about one source, but contain little information about the other. We took care to ensure that the two volatility estimates were decorrelated, so that we would be able to attribute neural signals to each of them unambiguously. We also obtained model estimates of two prediction errors in each trial: The *reward*

prediction error (actual reward – expected value); and the *confederate prediction error* (confederate fidelity – predicted confederate fidelity). Again, we ensured that these prediction error signals were decorrelated from each other, and from the volatility signals reflecting the respective learning rate in each domain. Finally, we obtained model estimates of the two probabilities of the chosen option yielding reward: the current probability of whether the chosen option was correct based on the subjects' *own experience*, and the probability this option was correct based on the current *confederate advice (and their recent fidelity)*.

Each trial was divided into 4 phases: CUE (when the trial was presented, 3-5 seconds), SUGGEST (when the confederate advice appeared, 3-5 seconds), INTERVAL (post-decision phase, 3-5 seconds), and MONITOR (when the outcome was displayed, 3 seconds). This allowed us to test whether *decision*-related BOLD signal changes were present at the relevant times – CUE and SUGGEST - and whether *learning*-related BOLD signal changes were present at the relevant time – MONITOR. We analysed the data using FSL (Smith et al., 2004). Using a general linear model, we looked for learning-related activity by including regressors representing each of these 4 phases and the interaction of the MONITOR phase with each of the volatility and prediction error signals. In a separate general linear model, we analysed decision-related activity by including regressors for each of the 4 phases and the interaction of the CUE and SUGGEST phases with (i) the relevant probabilities (ii) the reward magnitude of the chosen option (iii) the reward magnitude of the unchosen option. Full GLM details are given in **sections 6.2.4.2** and **6.2.4.3**.

6.2.3 FMRI data acquisition

FMRI data were acquired in 24 subjects on a 3T Siemens TRIO scanner. Data were excluded from one subject due to rapid head motion. The remaining 23 subjects were included in the analysis.

FMRI data were acquired with a voxel resolution of 3x3x3 mm3, TR=3s, TE=30ms, Flip angle=87˚. The slice angle was set to 15˚ and a local z-shim was applied around the orbitofrontal cortex to minimize signal dropout in this region (Deichmann et al., 2003), which had been implicated in other aspects of decision-making in previous studies. The number of volumes acquired depended on the behaviour of the subject. The mean number of volumes was 943, giving a total experiment time of approximately 47 minutes. Stimulus presentation/subject button presses were registered and time-locked to FMRI data using Presentation (Neurobehavioral Systems, USA). Field Maps were acquired using a dual echo 2D gradient echo sequence with echoes at 5.19 and 7.65 ms,

and repetition time of 444ms. Data were acquired on a 64x64x40 grid, with a voxel resolution of 3mm isotropic.

T1-weighted structural images were acquired for subject alignment using an MPRAGE sequence with the following parameters: Voxel resolution 1x1x1 mm^3 on a 176x192x192 grid, Echo time(TE)=4.53 ms, Inversion time(TI)=900 ms, Repetition time (TR)=2200 ms.

6.2.4 FMRI data analysis

6.2.4.1 Preprocessing

Data were preprocessed using FSL default options: motion correction was applied using rigid body registration to the central volume (Jenkinson et al., 2002); Gaussian spatial smoothing was applied with a full width half maximum of 5mm; brain matter was segmented from non-brain using a mesh deformation approach (Smith, 2002); high pass temporal filtering was applied using a Gaussian-weighted running lines filter, with a 3dB cut-off of 100s. Susceptibility-related distortions were corrected as far as possible using FSL field-map correction routines (Jenkinson, 2003).

6.2.4.2 Model estimation (learning-related activity)

A general linear model was fit in pre-whitened data space (to account for autocorrelation in the FMRI residuals) (Woolrich et al., 2001). The following regressors (plus their temporal derivatives) were included in the model:

1. CUE – times when options and reward values were presented onscreen, but not social advice;
2. ADVICE – times when options, reward values and social advice were all presented onscreen;
3. INTERVAL – times between making a response and the outcome being revealed;
4. MONITOR – times when the outcome of the trial was presented onscreen;
5. MONITOR x REWARD HISTORY VOLATILITY – monitor phase, modulated by the estimated volatility in the reward history on each trial;
6. MONITOR x CONFEDERATE ADVICE HISTORY VOLATILITY – monitor phase, modulated by the estimated volatility in the confederate advice history on each trial.
7. MONITOR x REWARD PREDICTION ERROR – monitor phase, modulated by the prediction error in the frame of reference of the reward;
8 MONITOR x CONFEDERATE PREDICTION ERROR – monitor phase, modulated by the prediction error in the frame of reference of fidelity of the confederate advice;

These regressors were convolved with the FSL default haemodynamic response function (Gamma function, delay=6s, standard deviation =3s), and filtered by the same high pass filter as the data.

6.4.2.3 Model estimation (decision-related activity)

A separate general linear model was fit in pre-whitened data space (to account for autocorrelation in the FMRI residuals) (Woolrich et al., 2001). We computed two potential values of the subject's chosen option, each one based only on either social or non-social information (i.e. (i) the probability of a reward based only on experience and (ii) the probability of a reward based only on confederate advice). We used these two values, together with information about the reward magnitude, as regressors in our analysis. Information about reward magnitude and experience-based probability was available to subjects from the beginning of each trial (from the CUE phase onwards), whereas information about the collaborator-based probability was only available to subjects once the suggestion had been presented (SUGGEST phase). Each regressor was therefore interacted with the time the information was available.

The following regressors (plus their temporal derivatives) were therefore included in the model:

1. CUE – times when options and reward values were presented onscreen, but not social advice;
2. ADVICE – times when options, reward values and social advice were all presented onscreen;
3. INTERVAL – times between making a response and the outcome being revealed;
4. MONITOR – times when the outcome of the trial was presented onscreen;
5. CUE x EXPERIENCE-BASED PROBABILITY – cue phase, modulated by the logarithm of the probability of the chosen action based on subjects' previous experience;
6. SUGGEST x EXPERIENCE-BASED PROBABILITY – suggest phase, modulated by the logarithm of the probability of the chosen action based on subjects' previous experience;
7. SUGGEST x CONFEDERATE ADVICE-BASED PROBABILITY – suggest phase, modulated by the logarithm of the probability of the chosen action based on current confederate advice and previous confederate fidelity;
8. CUE x CHOSEN REWARD MAGNITUDE – cue phase, modulated by the logarithm of the reward magnitude of the chosen action;

9. SUGGEST x CHOSEN REWARD MAGNITUDE – suggest phase, modulated by the logarithm of the reward magnitude of the chosen action;
10. CUE x UNCHOSEN REWARD MAGNITUDE – cue phase, modulated by the logarithm of the reward magnitude of the unchosen action;
11. SUGGEST x UNCHOSEN REWARD MAGNITUDE – suggest phase, modulated by the logarithm of the reward magnitude of the unchosen action.

Note that in order to compute an overall probability the subjects must (approximately) multiply the two sources of information – experience based probability and advice-based probability. This overall probability should then be multiplied by the reward magnitude to obtain the Pascalian value of each option (see section **4.2.2.4**). In order to linearise this problem for FMRI, we therefore entered as regressors the logarithm of these three values.

These regressors were convolved with the FSL default haemodynamic response function (Gamma function, delay=6s, standard deviation =3s), and filtered by the same high pass filter as the data.

6.4.2.4 Group data processing

Subjects were aligned to the MNI152 template using affine registration (Jenkinson and Smith, 2001). A general linear model was fit to the effects of the regressors described above (Woolrich et al., 2004). In the case of the analysis of expected value (figure 4 in results), a general linear model was also fit to the effect of the combination of regressors 5, 6, and 7 in the decision-related analysis shown above.

This group GLM contained three factors:
1. A group mean.
2. The weight for reward history information based on each subject's behaviour (calculated using the method described in **section 4.2.2.4**).
3. The weight for confederate information based on each subject's behaviour (calculated using the method described in **section 4.2.2.4**).

6.4.2.5 Inference

Volatility effects: Effects of volatility (figure 3) were hypothesized to be present in the anterior cingulate cortex (ACC) based on previous data (Behrens et al., 2007). We therefore performed cluster inference (Z>3.1) correcting for multiple comparisons at

$p<0.05$ within a hand-drawn mask of the ACC. This required that there be more than 25 contiguous voxels.

Prediction Error effects: Prediction error effects (figure 2; table 1) are reported for clusters of greater than 50 contiguous voxels (for social prediction error) and greater than 100 contiguous voxels (for reward prediction error) at $Z>3.5$.

Expected Value effects: Effects of the two individual expected value signals (figure 5) are reported at $Z>2.6$ ($p<0.01$ uncorrected; $p<0.05$ cluster-corrected). Effects of the combination of the two probabilities (figure 4) were hypothesized to be present in ventromedial prefrontal cortex (vmPFC) based on previous data (Daw et al., 2006, Hampton, Bossaerts and O'Doherty, 2006, Kable and Glimcher, 2007). We therefore performed cluster inference correcting for multiple comparisons at $p<0.05$ within a hand-drawn mask of vmPFC. This required that there be more than 21 contiguous voxels.

6.4.2.6 Post-hoc fMRI region of interest analysis

Region of interest analyses were performed on activations reflecting the prediction errors on confederate and reward information. These analyses were performed in order to determine the nature of the BOLD signal fluctuations and their relationship to the expected fluctuations induced by prediction and prediction error signals, which could be accounted for by several potential confounding regressors. Figure 1 shows an outline of this analysis, which is described in detail below.

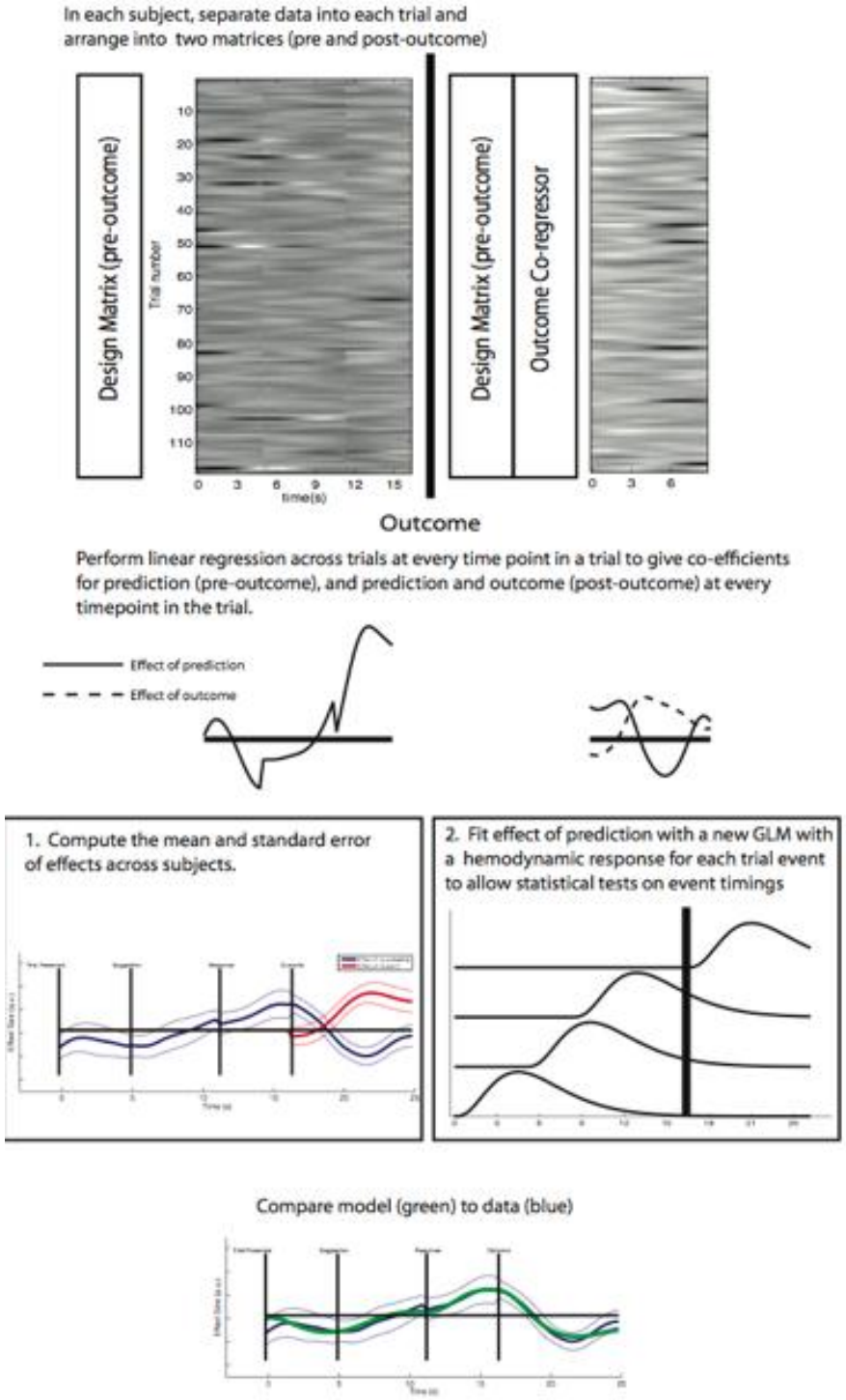

Figure 1. ROI analysis of fMRI timeseries, using linear regression. Details provided in text below.

We took BOLD data in each subject from masks back-projected from the each group prediction error region. We separated each subject's timeseries into each trial, and resampled each trial to a duration of 25s, such that the decision was presented at 0s, the confederate advice was presented at 5s, the response was given at 12s and the outcome was presented from 17s-20s. (These timings were the mean timings across all trials in all subjects.) The resampling resolution was 100ms. We then performed two

separate GLMs across trials in each subject. The first GLM included a regressor for the prediction (the estimated probability of a confederate lie in figure 2b, and the expected value of the trial in figure 2d). The second GLM included regressors for the prediction and for the outcome (in figure 2b, the outcome was the event of a collaborator lie (1 for lie and 0 for truth). In figure 2d the outcome was the reward itself. We then calculated the group average effect sizes (i.e. the mean of the effect across subjects) at each timepoint, and their standard errors. The graphs in the top panels of main figures 2b,d, therefore show a timeseries of effect sizes for the prediction throughout the trial (blue) and for the outcome after the outcome period (red). In each case, a prediction signal should therefore show a positive effect in the blue curve before the outcome. A prediction error signal (outcome – expectation) should show a positive effect of the red curve and a negative effect of the blue curve after the outcome.

6.3 Results

6.3.1 Prediction errors for social information

If subjects learn about the reliability of the confederate in an associative fashion, they should update their current estimate of this reliability using the *confederate prediction error*. Such a signal may be thought of as a learning signal about the *motive* of the confederate. FMRI correlates of the ascription of motive to stimuli have previously been reported in a network of brain regions including the superior temporal sulcus (STS), middle temporal gyrus (MTG), the temporoparietal junction (TPJ) and the dorsomedial prefrontal cortex (DMPFC) in the vicinity of the pregenual paracingulate sulcus (Amodio and Frith, 2006, Saxe, 2006, Van Overwalle, 2009). Such activations have been thought critical in studies of theory of mind.

We observed BOLD correlates of the confederate prediction error in DMPFC (MNI x=2,y=54,16, max Z=4.73), right MTG (MNI x=45,y=-30,z=-16, max Z=3.81), and right STS/TPJ (MNI x=54,y=-48,z=30, max Z=4.23) (figure 2, Z>3.1, cluster size > 50 voxels; table 1). Equivalent signals were present in the left hemisphere at the same threshold, but did not pass the extent criterion of 50 voxels. We also observed a similar effect bilaterally in the cerebellum (see table 1). Notably, these regions showed a characteristic pattern of activation similar to known dopaminergic activity in reward learning (Waelti, Dickinson and Schultz, 2001), but for social information. Activity correlated with the estimated probability of a confederate lie after the subject decision but before the outcome was revealed (a prediction signal). When the subjects observed the trial outcome, activity correlated negatively with this same probability, but positively with the actual event of a confederate lie (figure 2b). This outcome signal

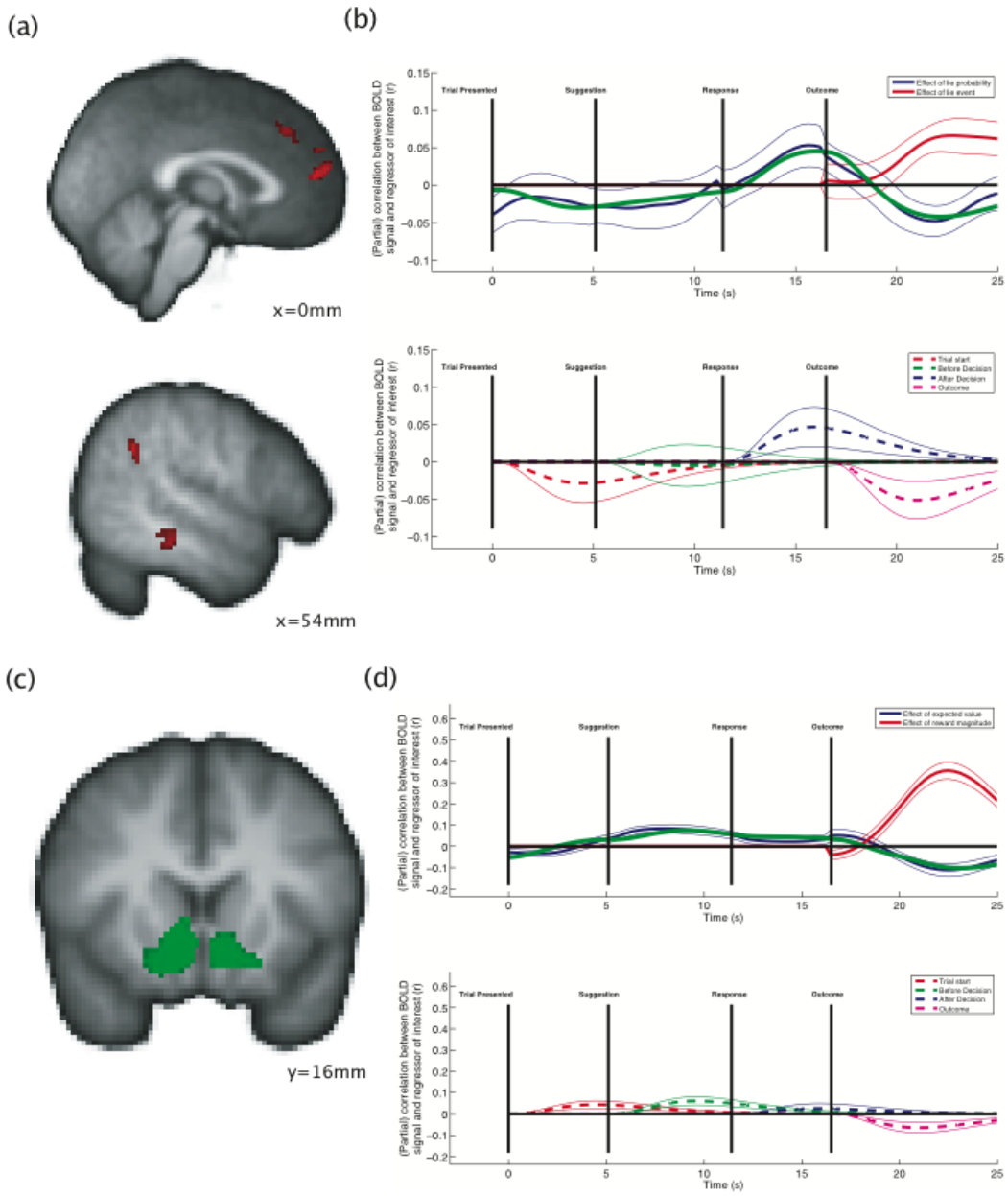

Figure 2. Time courses show (partial) correlations +/- s.e.m. (a) Activation in the DMPFC, right TPJ/STS and MTG correlate with the social prediction error at the outcome (threshold set at Z>3.1, cluster size>50 voxels). (b) Deconstruction of signal change in the DMPFC. Similar results were found in the MTG and TPJ/STS. Top: following the outcome, areas that encode prediction error correlate positively with the outcome and negatively with the predicted probability. Red, effect size of the confederate lie outcome (1 for lie, 0 for truth); blue, effect size of the predicted confederate lie probability. To perform inference, we fit a haemodynamic model in each subject to the time course of this effect (that is, to the blue line). The green line in the top panel shows the mean overall fit of this haemodynamic model (for comparison with the blue line). Bottom: the effect of lie probability (blue line from top panel) is decomposed into a haemodynamic response function at each trial event (corresponding to the four colours in the bottom panel) (see figure 1). Dashed and solid lines show mean responses +/- s.e.m. (c) Ventral striatum is taken as an example of a number of regions revealed by the voxel-wise analysis of reward prediction error (threshold set at Z . 3.1, cluster size .100 voxels). (d) Panels are exactly as in (b), but coded in terms of reward and not in terms of confederate fidelity. The top panel shows the parameter estimate relating to the expected value of the trial (blue line) and, after the outcome, the parameter estimate relating to the magnitude of these rewards (red line). To test for prediction error coding, we again fit a haemodynamic model to the expectation parameter estimate (shown by the green line, for comparison with blue line).

therefore reflects a prediction error signal for social information, as both components of the prediction error are represented: The outcome (lie or truth) minus the expectation of this outcome (figure 2b). These signals cannot be influenced by reward prediction error processing as the two types of prediction error are carefully controlled to be orthogonal in the task design, and furthermore compete against one another to explain variance in the FMRI data. The presence of this prediction error signal in the brain is a prerequisite for any theory of an RL-like strategy for social valuation.

We performed a similar analysis for prediction errors on reward information (reward minus expected reward). We found a significant effect of reward prediction error in the ventral striatum (MNI x=8,y=14,z=-10 max Z=5.33) (figure 2c, Z>3.1, cluster size > 50 voxels), the ventromedial prefrontal cortex, and posterior cingulate sulcus (see table 2). As in the social domain, we observed significant effect of all three elements of the reward prediction error: a positive effect of reward expectation prior to the outcome, a positive effect of delivered reward at the time of the outcome, and a negative effect of reward expectation at the time of the outcome (Figure 2d).

Cluster size (voxels)	Max Z	MNI x(mm)	MNI y(mm)	MNI z(mm)	Location
216	4.11	-26	-72	-36	Left Cerebellum
198	4.73	2	54	16	dmPFC
182	4.3	-8	44	38	dmPFC
111	3.98	32	-60	-48	Right Cerebellum
84	3.81	54	-30	-16	Right MTG
70	4.23	54	-48	30	Right STS/TPJ

Table 1. Activations for social prediction error at MONITOR (feedback) time, thresholded at Z>3.1.

Cluster size (voxels)	Max Z	MNI x(mm)	MNI y(mm)	MNI z(mm)	Location
1444	5.33	8	14	-10	Ventral Striatum
841	4.74	52	-68	-14	Extra-striate cortex
836	4.54	34	-14	48	Precentral gyrus
766	4.98	2	52	-8	vmPFC
313	4.61	6	-28	48	Posterior Cingulate sulcus
244	3.92	44	-34	56	Extra-striate cortex
197	4.52	-6	-80	-16	Striate cortex
189	4.21	16	-76	56	Dorsal parietal cortex
184	4.23	-20	-66	-48	Left cerebellum
121	4.61	14	-48	-54	Right cerebellum

Table 2. Activations for reward prediction error at MONITOR (feedback) time, thresholded at Z>3.1.

In order to ascertain whether the signals were really prediction error signals, we performed a hemodynamic convolution of the effect of the prediction (blue line, figure 2). These effects can be seen in the bottom panels of figures 2b and 2d. We assumed that the trial could be modelled by hemodynamic response functions (HRFs) at 5 characteristic times: 1) the initial cue starting the trial; 2) the confederate suggestion; 3)

the decision time; 4) the outcome. 5) The ITI (not shown). In each subject we then fit the BOLD effect of prediction (blue line in top panel) with these 5 HRFs using a general linear model. A prediction error signal should show a significant negative effect of the 4th hrf (after the outcome was revealed). This was true for the social prediction error signals ($t(22)=2.68$ ($p<0.005$), 2.35 ($p<0.05$), 3.27 ($p<0.005$) for DMPFC, right STS/MTG and right TPJ respectively), and for the ventral striatal signal for reward prediction error ($t(22)=2.50$, $p<0.05$). The social regions showed a significant positive effect of lie prediction during the 3rd hrf ($t(22)=1.96$ ($p<0.05$), 1.73($p<0.05$), 1.74($p<0.05$) for DMPFC, right STS/MTG and right TPJ respectively). The ventral striatum showed a significant positive effect of reward prediction during the 2nd hrf ($t(22)=3.32$ ($p<0.002$)).

In order to verify the model fit of the 5 HRF model, we plotted the predicted timecourse of effect sizes from these HRFs on top of the observed timecourses (green line in top panels).

6.3.2 Agency-specific learning rates dissociate in the ACC

Anatomically distinct brain regions encode information necessary for learning in social and non-social domains. Information about personal experiences must derive from the actions produced by the motor system, whereas information concerning another agent is more likely to reflect activity in a network of brain areas encoding social information. One cortical region with access to both forms of information via direct connections is the Anterior Cingulate Cortex (ACC). However, in the macaque monkey, there is an anatomical dissociation within the ACC with respect to connections with these two systems. Connections with motor regions lie predominantly in the ACC sulcus (ACCs). Connections with visceral and social regions, such as the STS, hypothalamus and amygdala, lie predominantly in the ACC gyrus (ACCg) (Van Hoesen, Morecraft and Vogt, 1993). In macaque monkeys, selective lesions to ACCs but not ACCg impair reward-guided decision-making in the non-social domain (Kennerley et al., 2006, Rudebeck et al., 2008). In the social domain, male macaques will forego food opportunities in order to acquire information about other individuals (Deaner, Khera and Platt, 2005, Shepherd, Deaner and Platt, 2006). Selective lesions to ACCg but not ACCs (or other regions of prefrontal cortex) abolish this effect (Rudebeck et al., 2006a, Noonan et al., 2010).

As previously demonstrated (Behrens et al., 2007), the volatility of action-outcome associations predicted BOLD signal in a circumscribed region of ACCs (figure 3a). This effect varied across people such that those whose behaviour relied more on

their own experiences showed a greater volatility related signal in this region (figure 3b) (max Z=3.7, MNI x=-16,y=12,z=40, p<0.05 cluster-corrected for ACC). By contrast, the volatility of confederate advice correlated with BOLD signal in a circumscribed region in the adjacent ACCg (figure 3a). Subjects whose behaviour relied more on this advice showed greater signal change in this region (figure 3c - max Z=4.1, MNI x=-6,y=12,z=26, p<0.05 cluster-corrected for ACC). BOLD signals in these two regions therefore reflect the respective values of the same outcome for learning about the two different sources of information. The fact that they correlate with the same computational learning parameter in the two different contexts, and that these correlations both drive behaviour, suggests that similar processes are employed in parallel in ACC for learning about social partners and for learning from experience derived from one's own actions.

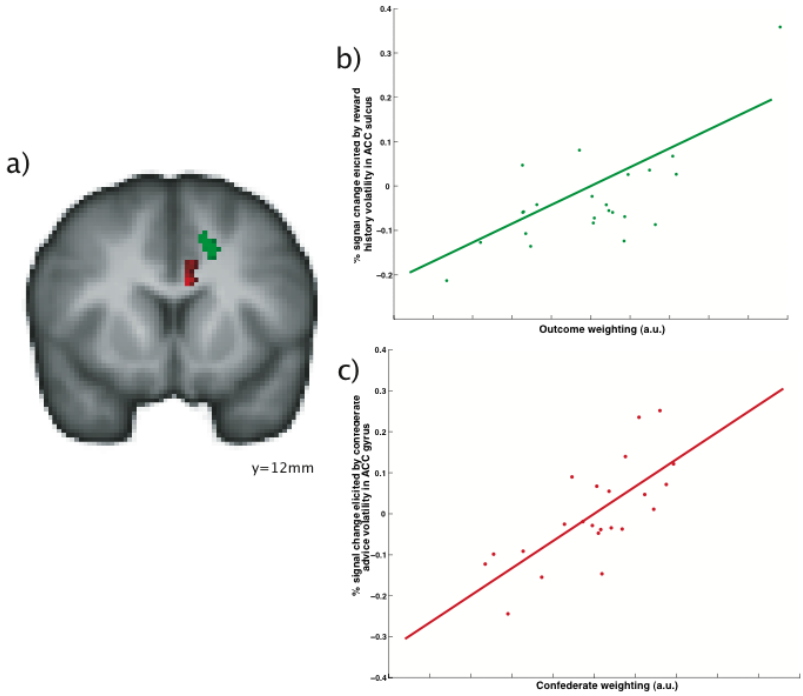

Figure 3. Agency-specific learning rates dissociate in the ACC. (a) Regions where the BOLD correlates of reward (green) and confederate (red) volatility predict the influence that each source of information has on subject behaviour (Z>3.1, P<0.05 cluster-corrected for cingulate cortex). (b) Subjects with high BOLD signal changes in response to reward volatility in the ACCs are guided strongly by reward history information (correlation R=0.7163, P<0.0001 (note R value and P value are subject to circularity bias (see Vul et al., 2009; Kriegeskorte et al., 2009) - they are reported for completeness, but crucially are *not* used for inference). (c) Subjects with high BOLD signal changes in response to confederate advice volatility in the ACCg are guided strongly by social information (correlation R=0.7252, P<0.0001, subject to circularity bias).

6.3.3 Combining different sources of information

Learning about reward probability from vicarious and personal experiences recruits distinct neural systems (figures 1 and 2), but subjects combine information across both sources when making decisions (see **section 4.2.3.2**). For this to be achieved, one might expect to find a signal that represents the combination of these two sources at the time of the decision. A ventromedial portion of the prefrontal cortex (VMPFC) has been shown to code such an expected value signal during decision-making (Daw et al., 2006, Hampton, Bossaerts and O'Doherty, 2006, Kable and Glimcher, 2007, Plassmann, O'Doherty and Rangel, 2007, Boorman et al., 2009). It has been suggested that VMPFC activity might represent a common currency in which the value of different types of items might be encoded (O'Doherty, 2004).

We computed two potential probabilities of the subject's chosen option, each based on only one of the two sources of information (i.e. (i) based only on experience and (ii) based only on confederate advice), and used these probabilities as regressors in our analysis. Signal in a circumscribed region in the VMPFC was significantly correlated with the combination of the two probabilities (figure 4a, max Z = 4.51, MNI x=-2,y=26,z=-18, $p<0.005$ cluster-corrected for VMPFC). Activations were also found for both probabilities (experience-based and confederate advice-based) separately (figure 5), demonstrating that both are represented simultaneously in VMPFC. However, there was subject variability in whether the VMPFC signal better reflected the reward probability based on outcome history or on social information. The extent to which the VMPFC data reflected each source of information (at the time of the decision) was predicted by the ACCs/ACCg response to outcome/social volatility (at the time when the outcomes were witnessed) (figure 4b,c).

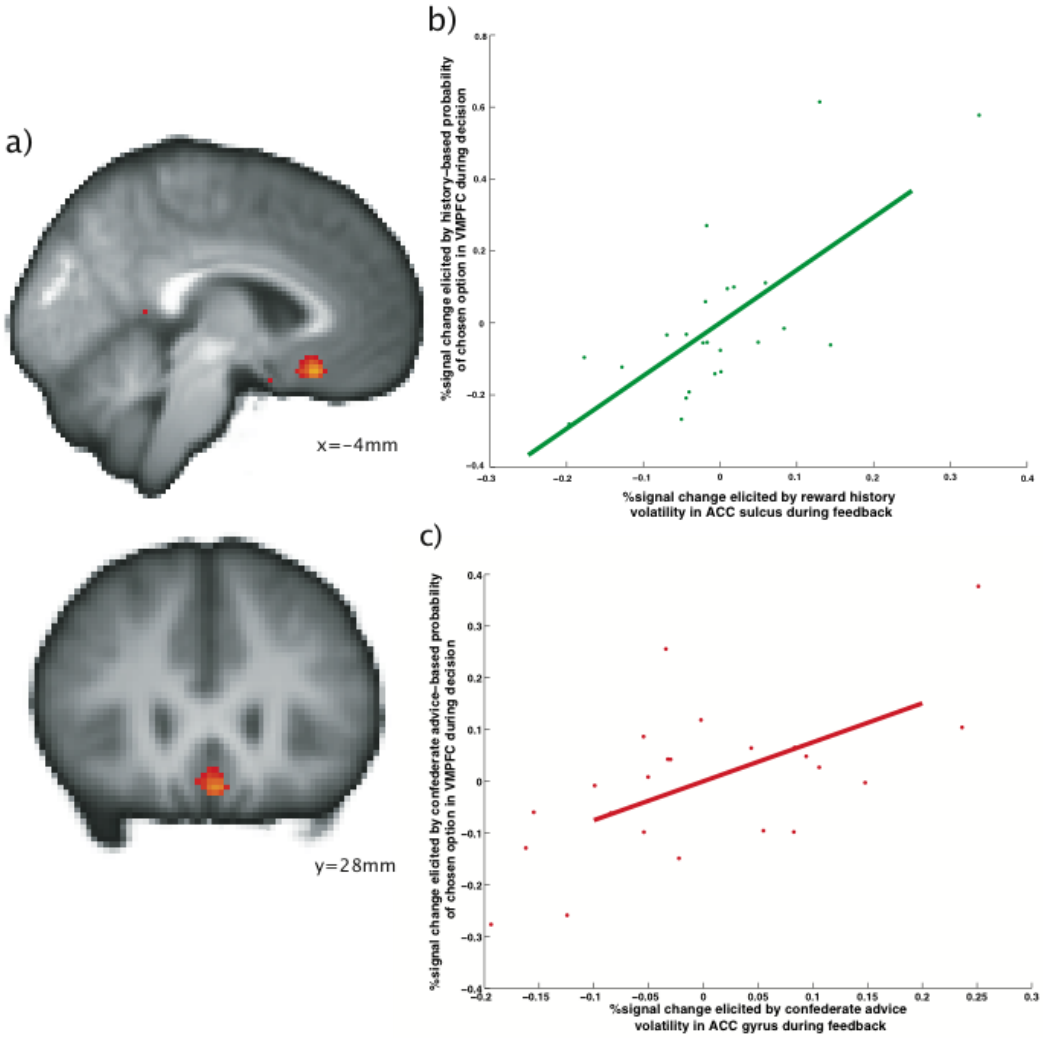

Figure 4. a, Activation for the combination (mean contrast) of experience-based probability during the CUE and SUGGEST phases, and advice-based probability during the SUGGEST phase (threshold set at Z>3.1, P<0.05 cluster-corrected for VMPFC). These phases represent the times at which subjects had these probabilities available to them. b, Correlation between the effect of outcome-based probability in VMPFC during the decision and the effect of outcome volatility in ACCs during the MONITOR phase (R=0.7113, P<0.0002, *not* subject to circularity bias). c, Correlation between the effect of confederate-based probability in VMPFC during the decision and the effect of confederate volatility in ACCg during the MONITOR phase (R= 0.6119, P<0.002).

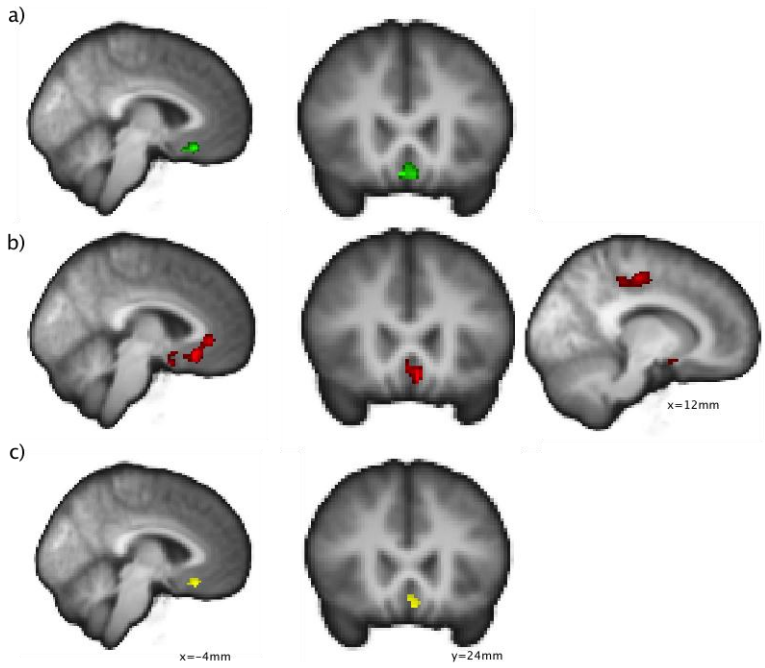

Figure 5. Responses to the reward probabilities computed on the basis of each source of information individually. (a) Response to reward probability during CUE and SUGGEST computed only on the basis of past outcomes (thresholded at Z>2.6, p<0.05 cluster-corrected for VMPFC (cluster size >100 voxels)). No other clusters of >20 voxels were present elsewhere in the brain at this Z-threshold. (b) Response to social advice probability during SUGGEST computed only on the basis of present and past confederate advice (thresholded at Z>2.6, p<0.05 whole-brain cluster corrected). One other region (right posterior cingulate cortex (peak Z=4.56 at MNI 12mm, -20mm, 50mm), shown in right panel) also survived this cluster-based thresholding. (c) Overlap of (a) and (b).

6.4 Discussion

Social interactions provide a rich source of information for guiding behaviour in many species (King-Casas et al., 2005, Amodio and Frith, 2006, Singer et al., 2006, Tomlin et al., 2006, Silk, 2007). Previous studies have shown that monkeys and humans assign different weight to information from different individuals, but neither how such weight is assigned, nor the mechanism by which it might be changed after new experiences of any one individual has been clear (Deaner, Khera and Platt, 2005, Rudebeck et al., 2006a). Here, we show that the weighting of social information in humans is subject to learning and continual update via an RL mechanism. We use techniques that predict behaviour in the context of learning from personal experiences to show that similar mechanisms explain behaviour in a social context. Furthermore, we demonstrate that key reinforcement learning parameters are coded neurally in the context of social information in the same fashion previously shown for information derived from personal experiences of reward. Despite employing the same learning mechanisms, distinct anatomical structures code learning parameters in the two

different domains. However, information from both is combined in ventromedial prefrontal cortex when making a decision.

In our experiment, we use a common technique in FMRI studies of social cognition to dissociate social from non-social processing (Gallagher et al., 2002, Rilling et al., 2004). We use two sources of information that are similar in every respect except that one is perceived to come from a computer, and the other from a confederate. In previous experiments this control has been ensured by having separate experiments or trials. Here, we have been able to ensure the same strong control by continuously manipulating the extent to which one or other source of information is relevant at each trial. As in previous studies, the crucial difference between the two information sources is that the confederate is perceived to have a complex motive behind his actions.

By comparing the two sources of information, we find that social prediction error signals similar to those reported in dopamine neurons for reward-based learning are coded in the superior temporal sulcus, temporoparietal junction and dorsomedial prefrontal cortex in the social domain. BOLD signal fluctuations in these regions are often seen in social tasks (Van Overwalle, 2009), and in tasks which involve the attribution of *motive* to stimuli (Castelli et al., 2002). That such regions, central to many different social tasks, should code quantitative prediction and prediction error signals about a confederate, lends more weight to the argument that social evaluation mechanisms are able to rely on simple associative processes. Notably, these findings are similar to those reported in a recent study of mentalising-related computations (Hampton, Bossaerts and O'Doherty, 2008) but distinct from prediction errors found in ventral striatum in other social tasks (King-Casas et al., 2005, Klucharev et al., 2009, Burke et al., 2010). However, careful consideration of the *frame of reference* used to analyse the data shows that these findings are not in fact mutually incompatible with one another. We will expand upon this point in **chapter 7**.

A second crucial parameter in reinforcement learning models is the *learning rate*, reflecting the value of each new piece of information. In the context of reward-based learning, this parameter predicts BOLD signal fluctuations in the sulcal division of the ACC at the crucial time for learning (Behrens et al., 2007) – a finding that is replicated here. We further demonstrate that the exact same computational parameter, in the context of social learning, is reflected in BOLD fluctuations in the neighbouring gyral portion of ACC. In our study, learning in these two contexts is driven by the very same outcomes. Any given outcome may, however, have quite a different importance for learning about each source of information – the fidelity of the advisor or the subject's own experience-based estimate – in a manner that depends only on differences in the

recent stability of the two sources. That these different implications of the same outcomes predict BOLD signal in neighbouring regions of ACCg and ACCs suggests that parallel streams of learning occur within ACC for social and non-social information respectively.

Previous accounts of ACC have emphasised its importance in both goal-directed learning and social cognition. Several studies have suggested a dissociation between the processing of socially relevant vs. non-socially relevant information, but the degree to which the different regions operate on the basis of a shared principle, albeit in different domains, has been more difficult to ascertain (Etkin et al., 2006). Recent investigations of macaque ACC demonstrate an anatomical distinction between social and non-social processing in ventral supracallosal and more anterior pregenual parts of ACCg and the ACCs respectively (Kennerley et al., 2006, Rudebeck et al., 2006b). The evidence presented here is consistent with a similar anatomical separation between social and non-social processing domains. More importantly, however, the present findings suggest that the two regions process the two types of information in the same fashion; although concerned with different domains of information, both ACCs and ACCg activity increase as the value of a piece of information increases.

It has been suggested that VMPFC activity might represent a common currency in which the value of different types of items might be encoded (O'Doherty, 2004). Here we show that the same portion of the VMPFC represents the expected value of a decision based on the combination of information from social and experiential sources. However, the extent to which the VMPFC signal reflects each source of information during a decision is predicted by the extent to which the ACCs and ACCg modulate their activity at the point when information is learnt. If, as is suggested, the VMPFC response codes the expected value of a decision, then the ACCs response to each new outcome predicts the extent that this outcome will determine future valuation of an action; the ACCg response predicts the extent to which this outcome will determine future valuation of an individual.

Chapter 7: Conclusions and general discussion

Computational modeling holds significant promise in dissecting which cortical regions subserve different component processes in human decision under risk and uncertainty. This thesis has provided two examples of how such models might be used when analysing neuroimaging data. In this final chapter, we draw together some key principles from these two streams of research, and consider them in a more general setting of other recent and relevant findings from other laboratories.

Because neural representations exist to perform functions, our understanding of representation and coding cannot end with the exploration of neural signals themselves but must explain how signaling mechanisms underlie cognitive and behavioural processes. Attempting to understand neuronal representation in the absence of cognition is akin to analyzing the mathematical patterns in a musical score without ever listening to the music – it misses entirely the reasons that particular patterns exist.

DeCharms and Zador, 2000 (Annual Rev Neurosci).

This thesis has used computational modeling approaches to investigate the neural correlates of subjective values during reward-guided choice and learning via reinforcement. We have drawn upon some of the models discussed in **chapter 2** when modeling subjects' behaviour (**chapter 4**). In **chapter 5**, we used a technique with high temporal resolution, magnetoencephalography (MEG), to test predictions that evolved through time from a pre-existing biophysical decision model. Combined with some of the source reconstruction techniques discussed in **chapter 3**, this allowed us to compare activity to model predictions from across the cerebral cortex. In **chapter 6**, we used a technique with high spatial resolution, functional MRI, to disentangle brain regions involved in intentional inference during a social interaction from those involved in reward-guided learning.

In this final chapter, we briefly discuss some general principles that can be drawn from these two studies, and how they might be interpreted in the context of other previous studies.

7.1 What decision variables are 'represented' during choice?

In **chapter 1** we considered how, at the time of making a decision, there is some *heterogeneity* in terms of the signal measured using brain imaging and single unit recording techniques. For instance, at the time of the decision, 'chosen value' and 'state value' representations are frequently found. Why might these signals be represented at the time when a decision is made, when they are of little use to the organism, rather than at the time of feedback, when they might be used to compute a prediction error signal? Are these representations of explicit use to the organism in guiding its

behaviour? Or is there an alternative framework in which we can understand why such representations might be necessary for the animal?

Before we go hunting for an alternative framework, it is useful to step back and consider what we mean by a 'representation'. It is possible that signals that covary with an experimental variable might be of little *functional* relevance to the animal. Consider the toy example in figure 1 (deCharms and Zador, 2000). Neurons B1 and B2 receive identical input from a stimulus neuron, A, and so their firing rate should be identical to one another. To the experimenter, listening in with a recording electrode (or an imaging technique), the 'representation' of the stimulus is identical. That is to say, their *content* is identical to one another. However, neuron B1 performs some computation based on this content, and sends the output to neuron C. This *transformation* reflects the *function* of the neuron B1 – it takes the stimulus and modifies it to produce behaviour. By contrast, neuron B2 does not affect behaviour – and so it has little (if any) *function* at all to play in this (highly simplified) microcircuit.

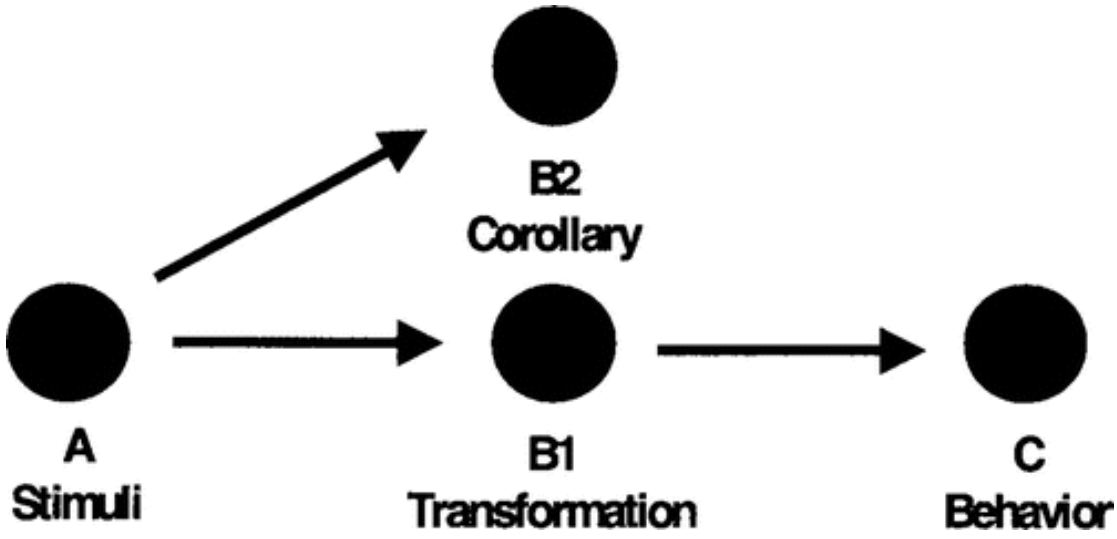

Figure 1. A highly simplified microcircuit that transforms stimuli (received by neuron A) into behaviour (elicited via neuron C). Neuron B1 is involved in this transformation, whereas neuron B2 has no axonal projections associated with this behaviour, and so forms a 'corollary discharge' role. To make this toy example a bit more concrete, we could consider A to be a motion sensitive neuron in area MT, B1/B2 to be neurons in area LIP, and C to be neurons in saccade-controlling output regions (such as the superior colliculus).

The key distinction here is that there are '*content*' representations – those that are measured using a neuroimaging technique on the macroscopic scale, or a microelectrode on the single unit scale – but these might or might not reflect the '*functional*' representation of the brain region or neurons – those which are decoded by a downstream brain region after some computation has been performed. The content

and function of a neural signal are often implicitly or explicitly conflated when considering data recorded during reward-guided choice.

Although the toy example is unrealistic, it proves useful when considering the output and predictions of the more realistic biophysical model of a cortical circuit discussed in **chapter 2** and **chapter 5**. The circuit makes a value-guided decision in a stochastic fashion that matches human behaviour on the task. As such, the model performs a specific computation - transforming inputs reflecting the value of available options to outputs reflecting a categorical choice. We therefore know *a priori* what are the 'functional' representations of the microcircuit; a downstream 'decoder' would need to read out a categorical commitment to a course of action (Lo and Wang, 2006). Crucially, however, we have seen that other 'content' representations naturally fall out of the biophysical models – those relating to 'overall value' and 'value difference' – at the time when the decision is made. This simply reflects the fact that network transitions (from one attractor state to another) occur at different speeds for decisions of different values. Although these content representations may be measured using MEG (**chapter 5**), they do not reflect the function of the microcircuit.

Future work might expand on this idea and start to re-examine some of the neurophysiological (i.e. single unit) data in the light of such a framework. The key assumption that needs to be avoided is that the presence of a 'content' representation – something that covaries with an experimental variable - during a choice implies that this is the signal that must be *decoded* from this region by a downstream neuron. The use and development of mechanistic models may also allow for us to distinguish between different brain regions in terms of the role that they play during the course of a value-guided choice. Moreover, different models may make unique predictions as to the effects of *perturbation* of neural activity – using either microstimulation or transcranial magnetic stimulation – on choice behaviour. Such causal manipulations have a strong role to play in disentangling 'content' and 'functional' representations in different cortical microcircuits.

7.2 'Frames of reference' in social decision making

In **chapter 6** we saw that in a social learning game, in which the fidelity of confederate advice is learnt slowly through time via reinforcement, a reinforcement learning (RL) model can be used to make predictions of neural activity in brain regions previously implicated in 'mentalising' or 'theory of mind' – namely, the dorsomedial prefrontal cortex (DMPFC) and the right temporoparietal junction (TPJ).

Another recent study employed an iterated inspection game, in which an 'inspector' chooses whether or not to monitor the behaviour of a 'worker', to look for similar signals (Hampton, Bossaerts and O'Doherty, 2008). In this game, inspecting is costly if the worker is already working, whilst working is costly if the inspector fails to inspect. If both players were to play the task optimally, the best strategy would be to adopt a mixed strategy of assigning a certain probability to each action, and selecting from these probabilities at random. However, if either player is suboptimal, human subjects might track the *previous* behaviour of the partner, and use this to *infer* a strategy that exploits the other subject's behaviour. A yet more sophisticated strategy would incorporate the *influence* of each player's current action on the next move that the partner would take. Quantitative RL models can be built that deploy each of these strategies; both superior temporal sulcus and DMPFC signal the 'influence update term' at the time critical for learning, and activity in DMPFC correlates with the likelihood that the sophisticated influence model is being used.

Importantly, in both this study and the study presented in **chapter 6**, neural activity can be thought of in one of two *frames of reference.* Firstly, activity can vary in the frame of reference of the *other player's behaviour*, and this seems to affect activity in areas such as DMPFC, TPJ and STS. Secondly, activity can also vary in the frame of reference of *reward for oneself*. By design, these two quantities can be kept orthogonal to one another. Notably, reward prediction errors in the 'self' frame of reference (rather than the 'other' frame of reference) are found in regions more traditionally associated with reward and reinforcement, such as the ventral striatum and ventromedial prefrontal cortex.

Several other recent studies of social cognition have also presented prediction error-like activity in the ventral striatum. King-Casas and colleagues have carefully examined data collected from subjects interacting in an iterated version of the trust game, that allows for the building of a reputation between investor and trustee (King-Casas et al., 2005). In the trustee's brain, they find increased activity in the head of the caudate nucleus when the investor reciprocates their past behaviour in a generous fashion ('benevolent' reciprocity) as compared to trials when they fail to do so ('malevolent' reciprocity). This activity could be a prediction error in the frame of reference of the *investor's* future behaviour – an adjustment of the trustee's expectations of the investor – or alternatively could be a prediction error in the frame of reference of the *trustee's* future behaviour – as benevolent reciprocity is more likely to induce an increase in trust. King-Casas *et al*. show clear evidence for the latter proposition – the

activity in striatum is increased selectively on trials in which the trustee is to increase his *own* level of trusting behaviour in future rounds.

Klucharev and colleagues scanned subjects as they rated the attractiveness of photographs of individuals in a 'hot or not'-style task; they then presented the average rating of a group of other individuals who had rated the picture (Klucharev et al., 2009). As expected, later ratings of the same photographs were highly influenced by what others thought of the photo, and the striatum and ACC were both found to be influenced by conflict between one's own opinions and that of others. Again, however, this signal (likened by the authors to a prediction error) is in the frame of reference of one's own behaviour, as evidenced by the fact that it is stronger when one's own behaviour is modified by the conflict than when it is not.

The main conclusion from these studies thus far is that when one is learning about the intentions of another individual, then a 'social network' of brain regions is recruited, whereas when one is learning about one's own behaviour, a 'reward network' of regions is recruited. The key to understanding neural activity in each of these studies is to carefully consider whether they are in the 'self' or 'other' frames of reference. Although traditionally this has been done by having separate conditions (e.g. where one is interacting with a computer)(Rilling et al., 2004), by keeping activity in each frame of reference orthogonal to the other, it is possible to examine neural activity in both frames of reference *simultaneously*, as in **chapter 6**. Future studies might build on this work by further interrogating activity in 'social' brain regions, and see how they are recruited during 'higher-order' theory of mind (see (Yoshida et al., 2010) for a recent example), or when having to 'mentalise' about one's own behaviour whilst acting on behalf of another individual.

References

Allais M (1953) Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine. Econometrica 21:503-546.

Amiez C, Joseph JP, Procyk E (2005) Anterior cingulate error-related activity is modulated by predicted reward. Eur J Neurosci 21:3447-3452.

Amiez C, Joseph JP, Procyk E (2006) Reward encoding in the monkey anterior cingulate cortex. Cereb Cortex 16:1040-1055.

Amodio DM, Frith CD (2006) Meeting of minds: the medial frontal cortex and social cognition. Nat Rev Neurosci 7:268-277.

An X, Bandler R, Ongür D, Price JL (1998) Prefrontal cortical projections to longitudinal columns in the midbrain periaqueductal gray in macaque monkeys. J Comp Neurol 401:455-479.

Andersen RA, Buneo CA (2002) Intentional maps in posterior parietal cortex. Annu Rev Neurosci 25:189-220.

Andrade A, Paradis AL, Rouquette S, Poline JB (1999) Ambiguous results in functional neuroimaging data analysis due to covariate correlation. Neuroimage 10:483-486.

Apicella P, Scarnati E, Ljungberg T, Schultz W (1992) Neuronal activity in monkey striatum related to the expectation of predictable environmental events. J Neurophysiol 68:945-960.

Axelrod R, Hamilton WD (1981) The evolution of cooperation. Science 211:1390-1396.

Baddeley A (1992) Working memory. Science 255:556-559.

Baillet S, Mosher JC, Leahy RM (2001) Electromagnetic brain mapping. Signal Processing Magazine.

Baker SN, Curio G, Lemon RN (2003) EEG oscillations at 600 Hz are macroscopic markers for cortical spike bursts. J Physiol 550:529-534.

Balleine BW, Daw ND, O'Doherty J (2008) Multiple forms of value learning and the function of dopamine. In: Neuroeconomics: decision making and the brain(Glimcher, P. et al., eds), pp 365-386: Academic Press.

Balleine BW, Delgado MR, Hikosaka O (2007) The role of the dorsal striatum in reward and decision-making. J Neurosci 27:8161-8165.

Barkley GL, Baumgartner C (2003) MEG and EEG in epilepsy. J Clin Neurophysiol 20:163-178.

Barron G, Erev I (2003) Small feedback-based decisions and their limited correspondence to description-based decisions. J Behav Dec Making 16:215-233.

Barth DS (2003) Submillisecond synchronization of fast electrical oscillations in neocortex. J Neurosci 23:2502-2510.

Basten U, Biele G, Heekeren HR, Fiebach CJ (2010) How the brain integrates costs and benefits during decision making. Proc Natl Acad Sci U S A 107:21767-21772.

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.

Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47:129-141.

Baylis LL, Gaffan D (1991) Amygdalectomy and ventromedial prefrontal ablation produce similar deficits in food choice and in simple object discrimination learning for an unseen reward. Exp Brain Res 86:617-622.

Bechara A, Damasio H, Tranel D, Damasio AR (1997) Deciding advantageously before knowing the advantageous strategy. Science 275:1293-1295.

Behrens TEJ, Hunt LT, Rushworth MSF (2009) The computation of social behavior. Science 324:1160-1164.

Behrens TEJ, Hunt LT, Woolrich MW, Rushworth MSF (2008) Associative learning of social value. Nature 456:245-249.

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

Beninger RJ, Hahn BL (1983) Pimozide blocks establishment but not expression of amphetamine-produced environment-specific conditioning. Science 220:1304-1306.

Berg P, Scherg M (1994a) A fast method for forward computation of multiple-shell spherical head models. Electroencephalogr Clin Neurophysiol 90:58-64.

Berg P, Scherg M (1994b) A multiple source approach to the correction of eye artifacts. Electroencephalogr Clin Neurophysiol 90:229-241.

Berger H (1929) Über das elektrenkephalogramm des menschen. Archiv für psychiatrie 527-570.

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.

Bernoulli D (1738) Exposition of a new theory on the measurement of risk (trans. Sommer, L, 1954). Econometrica 22:22-36.

Berns GS, McClure SM, Pagnoni G, Montague PR (2001) Predictability modulates human brain response to reward. J Neurosci 21:2793-2798.

Berridge KC (1996) Food reward: brain substrates of wanting and liking. Neurosci Biobehav Rev 20:1-25.

Bisiach E, Luzzatti C (1978) Unilateral neglect of representational space. Cortex 14:129-133.
Bisley JW, Goldberg ME (2003) Neuronal activity in the lateral intraparietal area and spatial attention. Science 299:81-86.
Blair K, Marsh AA, Morton J, Vythilingam M, Jones M, Mondillo K, Pine DC, Drevets WC, Blair JR (2006) Choosing the lesser of two evils, the better of two goods: specifying the roles of ventromedial prefrontal cortex and dorsal anterior cingulate in object choice. J Neurosci 26:11379-11386.
Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD (2006) The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. Psychol Rev 113:700-765.
Boorman ED, Behrens TEJ, Woolrich MW, Rushworth MSF (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.
Boorman ED, Sallet J (2009) Mean-variance or prospect theory? The nature of value representations in the human brain. J Neurosci 29:7945-7947.
Botvinick M, Nystrom LE, Fissell K, Carter CS, Cohen JD (1999) Conflict monitoring versus selection-for-action in anterior cingulate cortex. Nature 402:179-181.
Botvinick MM, Cohen JD, Carter CS (2004) Conflict monitoring and anterior cingulate cortex: an update. Trends Cogn Sci 8:539-546.
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.
Bower GH (1994) A turning point in mathematical learning theory. Psychol Rev 101:290-300.
Bragin A, Mody I, Wilson CL, Engel J, Jr. (2002) Local generation of fast ripples in epileptic brain. J Neurosci 22:2012-2021.
Brookes MJ, Gibson AM, Hall SD, Furlong PL, Barnes GR, Hillebrand A, Singh KD, Holliday IE, Francis ST, Morris PG (2004) A general linear model for MEG beamformer imaging. Neuroimage 23:936-946.
Brothers L (1990) The social brain: a project for integrating primate behaviour and neurophysiology in a new domain. Concepts Neurosci 1:27-51.
Brown JW, Braver TS (2005) Learned predictions of error likelihood in the anterior cingulate cortex. Science 307:1118-1121.
Brunel N, Wang XJ (2003) What determines the frequency of fast network oscillations with irregular neural discharges? I. Synaptic dynamics and excitation-inhibition balance. J Neurophysiol 90:415-430.
Burke CJ, Tobler PN, Baddeley M, Schultz W (2010) Neural mechanisms of observational learning. Proc Natl Acad Sci U S A 107:14431-14436.
Busemeyer JR, Townsend JT (1993) Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. Psychol Rev 100:432-459.
Bush RR, Mosteller F (1951) A mathematical model for simple learning. Psychol Rev 58:313-323.
Cai X, Kim S, Lee D (2011) Heterogeneous Coding of Temporally Discounted Values in the Dorsal and Ventral Striatum during Intertemporal Choice. Neuron 69:170-182.
Camerer C (2000) Prospect theory in the wild: evidence from the field. Choices, Frames and Values.
Camerer C, Ho TH (1999) Experience-weighted attraction learning in normal form games. Econometrica 67:827-874.
Carmichael ST, Price JL (1995) Sensory and premotor connections of the orbital and medial prefrontal cortex of macaque monkeys. J Comp Neurol 363:642-664.
Carmichael ST, Price JL (1996) Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. J Comp Neurol 371:179-207.
Carpenter RH, Williams ML (1995) Neural computation of log likelihood in control of saccadic eye movements. Nature 377:59-62.
Castelli F, Frith C, Happe F, Frith U (2002) Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. Brain 125:1839-1849.
Churchland AK, Kiani R, Chaudhuri R, Wang XJ, Pouget A, Shadlen MN (2011) Variance as a signature of neural computations during decision making. Neuron 69:818-831.
Churchland AK, Kiani R, Shadlen MN (2008) Decision-making with multiple alternatives. Nat Neurosci 11:693-702.
Cohen D (1972) Magnetoencephalography: detection of the brain's electrical activity with a superconducting magnetometer. Science 175:664-666.
Cohen MX, Elger CE, Ranganath C (2007) Reward expectation modulates feedback-related negativity and EEG spectra. Neuroimage 35:968-978.
Colander D (2007) Retrospectives: Edgeworth's hedonimeter and the quest to measure utility. J Econ Persp 21:215-225.
Colle LM, Wise RA (1988) Effects of nucleus accumbens amphetamine on lateral hypothalamic brain stimulation reward. Brain Res 459:361-368.
Compte A, Brunel N, Goldman-Rakic PS, Wang XJ (2000) Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. Cereb Cortex 10:910-923.
Corbit LH, Balleine BW (2003) The role of prelimbic cortex in instrumental conditioning. Behav Brain Res 146:145-157.
Courville AC, Daw ND, Touretzky DS (2006) Bayesian theories of conditioning in a changing world. Trends Cogn Sci (Regul Ed) 10:294-300.

Croxson PL, Johansen-Berg H, Behrens TEJ, Robson MD, Pinsk MA, Gross CG, Richter W, Richter MC, Kastner S, Rushworth MSF (2005) Quantitative investigation of connections of the prefrontal cortex in the human and macaque using probabilistic diffusion tractography. J Neurosci 25:8854-8866.
Curio G (2000) Linking 600-Hz "spikelike" EEG/MEG wavelets ("sigma-bursts") to cellular substrates: concepts and caveats. J Clin Neurophysiol 17:377-396.
D'Ardenne K, McClure SM, Nystrom LE, Cohen JD (2008) BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. Science 319:1264-1267.
Dalal SS, Guggisberg AG, Edwards E, Sekihara K, Findlay AM, Canolty RT, Berger MS, Knight RT, Barbaro NM, Kirsch HE, Nagarajan SS (2008) Five-dimensional neuroimaging: localization of the time-frequency dynamics of cortical activity. Neuroimage 40:1686-1700.
Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9:179-194.
Damasio A (1994) Descartes' error: emotion, reason, and the human brain: Putnam Publishing.
Danchin E, Giraldeau LA, Valone TJ, Wagner RH (2004) Public information: from nosy neighbors to cultural evolution. Science 305:487-491.
Darvas F, Pantazis D, Kucukaltun-Yildirim E, Leahy RM (2004) Mapping human brain function with MEG and EEG: methods and validation. Neuroimage 23 Suppl 1:S289-299.
Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8:1704-1711.
Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. Nature 441:876-879.
Dayan P, Kakade S, Montague PR (2000) Learning and selective attention. Nat Neurosci 3 Suppl:1218-1223.
de Lange FP, Jensen O, Dehaene S (2010) Accumulation of evidence during sequential decision making: the importance of top-down factors. J Neurosci 30:731-738.
De Martino B, Kumaran D, Seymour B, Dolan RJ (2006) Frames, biases, and rational decision-making in the human brain. Science 313:684-687.
Deaner RO, Khera AV, Platt ML (2005) Monkeys pay per view: adaptive valuation of social images by rhesus macaques. Curr Biol 15:543-548.
Debener S, Ullsperger M, Siegel M, Fiehler K, von Cramon DY, Engel AK (2005) Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. J Neurosci 25:11730-11737.
deCharms RC, Zador A (2000) Neural representation and the cortical code. Annu Rev Neurosci 23:613-647.
Deichmann R, Gottfried JA, Hutton C, Turner R (2003) Optimized EPI for fMRI studies of the orbitofrontal cortex. Neuroimage 19:430-441.
Delgado MR, Frank RH, Phelps EA (2005) Perceptions of moral character modulate the neural systems of reward during the trust game. Nat Neurosci 8:1611-1618.
Delorme A, Sejnowski T, Makeig S (2007) Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. Neuroimage 34:1443-1449.
den Ouden HE, Friston KJ, Daw ND, McIntosh AR, Stephan KE (2009) A dual role for prediction error in associative learning. Cereb Cortex 19:1175-1185.
Dias R, Robbins TW, Roberts AC (1996) Dissociation in prefrontal cortex of affective and attentional shifts. Nature 380:69-72.
Ditterich J, Mazurek ME, Shadlen MN (2003) Microstimulation of visual cortex affects the speed of perceptual decisions. Nat Neurosci 6:891-898.
Dorris MC, Glimcher PW (2004) Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. Neuron 44:365-378.
Dum RP, Strick PL (1991) The origin of corticospinal projections from the premotor areas in the frontal lobe. J Neurosci 11:667-689.
Dunbar RIM (1993) Coevoltion of neocortical size, group size and language in humans. Behav Brain Sci 16:681-735.
Elliott R, Dolan RJ, Frith CD (2000) Dissociable functions in the medial and lateral orbitofrontal cortex: evidence from human neuroimaging studies. Cereb Cortex 10:308-317.
Elliott R, Friston KJ, Dolan RJ (2000) Dissociable neural responses in human reward systems. J Neurosci 20:6159-6165.
Ellsberg D (1961) Risk, ambiguity and the Savage axioms. Q J Econ 75:643-669.
Erev I, Roth AE (1998) Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. The American Economic Review 88:848-881.
Eslinger PJ, Damasio AR (1985) Severe disturbance of higher cognition after bilateral frontal lobe ablation: patient EVR. Neurology 35:1731-1741.
Estes WK (1950) Toward a statistical theory of learning. Psychol Rev 57:94-107.
Etkin A, Egner T, Peraza DM, Kandel ER, Hirsch J (2006) Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala. Neuron 51:871-882.
Fehr E, Fischbacher U (2003) The nature of human altruism. Nature 425:785-791.
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 (2003) Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm. Brain 126:1830-1837.
Fellows LK, Farah MJ (2005) Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans. Cereb Cortex 15:58-63.
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.
Ferster CB, Skinner BF (1957) Schedules of reinforcement. East Norwalk, CT: Appleton-Century-Crofts.
Fibiger HC, LePiane FG, Jakubovic A, Phillips AG (1987) The role of dopamine in intracranial self-stimulation of the ventral tegmental area. J Neurosci 7:3888-3896.
FitzGerald TH, Seymour B, Dolan RJ (2009) The role of human orbitofrontal cortex in value comparison for incommensurable objects. J Neurosci 29:8388-8395.
Fletcher PC, Happe F, Frith U, Baker SC, Dolan RJ, Frackowiak RS, Frith CD (1995) Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. Cognition 57:109-128.
Forstmann BU, Dutilh G, Brown S, Neumann J, von Cramon DY, Ridderinkhof KR, Wagenmakers EJ (2008) Striatum and pre-SMA facilitate decision-making under time pressure. Proc Natl Acad Sci USA 105:17538-17542.
Fourier D (1822) Théorie analytique de la chaleur.
Fouriezos G, Wise RA (1976) Pimozide-induced extinction of intracranial self-stimulation: response patterns rule out motor or performance deficits. Brain Res 103:377-380.
Fox CR, Hadar L (2006) "Decisions from experience" = sampling error + prospect theory. Reconsidering Hertwig, Barron, Weber & Erev (2004). Judgment and Decision Making 1:159-161.
Fox CR, Poldrack RA (2008) Prospect theory and the brain. In: Neuroeconomics: decision making and the brainAcademic Press(Glimcher, P. et al., eds), pp 145-173.
Frank MJ, Seeberger LC, O'Reilly R C (2004) By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306:1940-1943.
Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360:815-836.
Friston K (2009) The free-energy principle: a rough guide to the brain? Trends Cogn Sci (Regul Ed) 13:293-301.
Friston K, Harrison L, Daunizeau J, Kiebel S, Phillips C, Trujillo-Barreto N, Henson R, Flandin G, Mattout J (2008) Multiple sparse priors for the M/EEG inverse problem. Neuroimage 39:1104-1120.
Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. Neuroimage 19:1273-1302.
Friston KJ, Holmes AP, Poline JB, Grasby PJ, Williams SC, Frackowiak RS, Turner R (1995) Analysis of fMRI time-series revisited. Neuroimage 2:45-53.
Friston KJ, Price CJ, Fletcher P, Moore C, Frackowiak RS, Dolan RJ (1996) The trouble with cognitive subtraction. Neuroimage 4:97-104.
Furman M, Wang XJ (2008) Similarity effect and optimal control of multiple-choice decision making. Neuron 60:1153-1168.
Gallagher HL, Jack AI, Roepstorff A, Frith CD (2002) Imaging the intentional stance in a competitive game. Neuroimage 16:814-821.
Gallagher M, McMahan RW, Schoenbaum G (1999) Orbitofrontal cortex and representation of incentive value in associative learning. J Neurosci 19:6610-6614.
Gallese V, Keysers C, Rizzolatti G (2004) A unifying view of the basis of social cognition. Trends Cogn Sci 8:396-403.
Garrido MI, Kilner JM, Kiebel SJ, Friston KJ (2007) Evoked brain responses are generated by feedback loops. Proc Natl Acad Sci USA 104:20961-20966.
Gehring WJ, Coles MG, Meyer DE, Donchin E (1995) A brain potential manifestation of error-related processing. Electroencephalogr Clin Neurophysiol Suppl 44:261-272.
Gershman SJ, Pesaran B, Daw ND (2009) Human reinforcement learning subdivides structured action spaces by learning effector-specific values. J Neurosci 29:13524-13531.
Glascher J, Hampton AN, O'Doherty JP (2009) Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. Cereb Cortex 19:483-495.
Glimcher PW, Camerer C, Fehr E, Poldrack RA (2008) A brief history of neuroeconomics. In: Neuroeconomics: decision making and the brain(Glimcher, P. et al., eds), pp 1-12: Academic Press.
Gnadt JW, Andersen RA (1988) Memory related motor planning activity in posterior parietal cortex of macaque. Exp Brain Res 70:216-220.
Gold JI, Shadlen MN (2000) Representation of a perceptual decision in developing oculomotor commands. Nature 404:390-394.
Gold JI, Shadlen MN (2002) Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. Neuron 36:299-308.
Goldman-Rakic PS (1992) Working memory and the mind. Sci Am 267:110-117.
Gonzalez R, Wu G (1999) On the shape of the probability weighting function. Cognitive Psychology 38:129-166.

Good IJ (1979) Studies in the history of probability and statistics: A.M. Turing's statistical work in World War II. Biometrika 66:393-378.
Gottfried JA, O'Doherty J, Dolan RJ (2003) Encoding predictive reward value in human amygdala and orbitofrontal cortex. Science 301:1104-1107.
Gottlieb JP, Kusunoki M, Goldberg ME (1998) The representation of visual salience in monkey parietal cortex. Nature 391:481-484.
Grabenhorst F, Rolls ET, Parris BA, d'Souza AA (2010) How the brain represents the reward value of fat in the mouth. Cereb Cortex 20:1082-1091.
Green DM, Swets JA (1966) Signal detection theory and psychophysics. New York: Wiley.
Guggisberg AG, Dalal SS, Findlay AM, Nagarajan SS (2007) High-frequency oscillations in distributed neural networks reveal the dynamics of human decision making. Front Hum Neurosci 1:14.
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.
Hadjipapas A, Hillebrand A, Holliday IE, Singh KD, Barnes GR (2005) Assessing interactions of linear and nonlinear neuronal sources using MEG beamformers: a proof of concept. Clin Neurophysiol 116:1300-1313.
Hadland KA, Rushworth MF, Gaffan D, Passingham RE (2003) The anterior cingulate and reward-guided selection of actions. J Neurophysiol 89:1161-1164.
Hämäläinen MS, Hari R, Ilmoniemi RJ, Knuutila J (1993) Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the human brain. Reviews of Modern Physics 65:413-497.
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.
Hampton AN, Bossaerts P, O'Doherty JP (2008) Neural correlates of mentalizing-related computations during strategic interactions in humans. Proc Natl Acad Sci USA 105:6741-6746.
Hanes DP, Schall JD (1996) Neural control of voluntary movement initiation. Science 274:427-430.
Hanks TD, Ditterich J, Shadlen MN (2006) Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. Nat Neurosci 9:682-689.
Hare TA, O'Doherty JP, Camerer CF, Schultz W, Rangel A (2008) Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. J Neurosci 28:5623-5630.
Hari R, Parkkonen L, Nangini C (2010) The brain in time: insights from neuromagnetic recordings. Ann N Y Acad Sci 1191:89-109.
Harlow JM (1848) Passage of an iron rod through the head. Boston Medical and Surgical Journal 39:389-393.
Haruno M, Kawato M (2006) Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. J Neurophysiol 95:948-959.
Hayasaka S, Nichols TE (2003) Validating cluster size inference: random field and permutation methods. Neuroimage 20:2343-2356.
Hayden BY, Heilbronner SR, Pearson JM, Platt ML (2011) Surprise signals in anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. J Neurosci 31:4178-4187.
Hayden BY, Pearson JM, Platt ML (2009) Fictive reward signals in the anterior cingulate cortex. Science 324:948-950.
Hayden BY, Platt ML (2010) Neurons in anterior cingulate cortex multiplex information about reward and action. J Neurosci 30:3339-3346.
Heekeren HR, Marrett S, Bandettini PA, Ungerleider LG (2004) A general mechanism for perceptual decision-making in the human brain. Nature 431:859-862.
Henson RN, Mattout J, Phillips C, Friston KJ (2009) Selecting forward models for MEG source-reconstruction using model-evidence. Neuroimage 46:168-176.
Henson RN, Mattout J, Singh KD, Barnes GR, Hillebrand A, Friston K (2007) Population-level inferences for distributed MEG source localization under multiple constraints: application to face-evoked fields. Neuroimage 38:422-438.
Henson RN, Mouchlianitis E, Friston KJ (2009) MEG and EEG data fusion: simultaneous localisation of face-evoked responses. Neuroimage 47:581-589.
Hernandez A, Zainos A, Romo R (2002) Temporal evolution of a decision-making process in medial premotor cortex. Neuron 33:959-972.
Herrnstein RJ (1961) Relative and absolute strength of response as a function of frequency of reinforcement. J Exp Anal Behav 4:267-272.
Hertwig R, Barron G, Weber EU, Erev I (2004) Decisions from experience and the effect of rare events in risky choice. Psychol Sci 15:534-539.
Hertwig R, Erev I (2009) The description-experience gap in risky choice. Trends Cogn Sci (Regul Ed) 13:517-523.

Hikosaka K, Watanabe M (2000) Delay activity of orbital and lateral prefrontal neurons of the monkey varying with different rewards. Cereb Cortex 10:263-271.
Hillebrand A, Barnes G (2005) Beamformer analysis of MEG data. Int Rev Neurobiology 68:149-171.
Hillebrand A, Barnes GR (2002) A quantitative assessment of the sensitivity of whole-head MEG to activity in the adult human cortex. Neuroimage 16:638-650.
Hillyard SA, Teder-Salejarvi WA, Munte TF (1998) Temporal dynamics of early perceptual processing. Curr Opin Neurobiol 8:202-210.
Hodges A (1992) Alan Turing: the enigma. London: Vintage.
Hodgkin AL, Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. J Physiol 117:500-544.
Holland PC, Rescorla RA (1975) The effect of two ways of devaluing the unconditioned stimulus after first- and second-order appeititive conditioning. J Exp Psych: Animal Behavior Processes 1:355-363.
Hollerman JR, Schultz W (1998) Dopamine neurons report an error in the temporal prediction of reward during learning. Nat Neurosci 1:304-309.
Holroyd CB, Coles MG (2002) The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychol Rev 109:679-709.
Hsu M, Krajbich I, Zhao C, Camerer CF (2009) Neural Response to Reward Anticipation under Risk Is Nonlinear in Probabilities. J Neurosci 29:2231-2237.
Huang MX, Shih JJ, Lee RR, Harrington DL, Thoma RJ, Weisend MP, Hanlon F, Paulson KM, Li T, Martin K, Millers GA, Canive JM (2004) Commonalities and differences among vectorized beamformers in electromagnetic source imaging. Brain Topogr 16:139-158.
Huettel SA, Song AW, McCarthy G (2005) Decisions under uncertainty: probabilistic context influences activation of prefrontal and parietal cortices. J Neurosci 25:3304-3311.
Huk AC, Shadlen MN (2005) Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. J Neurosci 25:10420-10436.
Hunt LT (2008) Distinctive roles for the ventral striatum and ventral prefrontal cortex during decision-making. J Neurosci 28:8658-8659.
Husain M, Rorden C (2003) Non-spatially lateralized mechanisms in hemispatial neglect. Nat Rev Neurosci 4:26-36.
Ito S, Stuphorn V, Brown JW, Schall JD (2003) Performance monitoring by the anterior cingulate cortex during saccade countermanding. Science 302:120-122.
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.
Jenkinson M (2003) Fast, automated, N-dimensional phase-unwrapping algorithm. Magn Reson Med 49:193-197.
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.
Jenkinson M, Smith S (2001) A global optimisation method for robust affine registration of brain images. Med Image Anal 5:143-156.
Jerbi K, Mosher JC, Baillet S, Leahy RM (2002) On MEG forward modelling using multipolar expansions. Phys Med Biol 47:523-555.
Jessup RK, Bishara AJ, Busemeyer JR (2008) Feedback produces divergence from prospect theory in descriptive choice. Psychol Sci 19:1015-1022.
Jezzard P, Smith SM, Matthews PM (2003) Functional MRI: an introduction to methods. Oxford: Oxford University Press.
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. Exp Neurol 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, Knetsch JL, Thaler RH (1991) Anomalies: the endowment effect, loss aversion, and status quo bias. J Econ Persp 5:193-206.
Kahneman D, Tversky A (1979) Prospect Theory: an Analysis of Decision under Risk. Econometrica 47:263-291.
Kamin LJ (1969) Selective association and conditioning. In: Fundamental issues in instrumental learning(Mackintosh, N. J. and Honig, W. K., eds), pp 42-64 Halifax, Canada: Dalhousie University Press.
Kawagoe R, Takikawa Y, Hikosaka O (1998) Expectation of reward modulates cognitive signals in the basal ganglia. Nat Neurosci 1:411-416.
Kennerley SW, Dahmubed AF, Lara AH, Wallis JD (2009) Neurons in the frontal lobe encode the value of multiple decision variables. Journal of Cognitive Neuroscience 21:1162-1178.

Kennerley SW, Walton ME, Behrens TEJ, Buckley MJ, Rushworth MSF (2006) Optimal decision making and the anterior cingulate cortex. Nat Neurosci 9:940-947.
Kiani R, Hanks TD, Shadlen MN (2008) Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. J Neurosci 28:3017-3029.
Kiebel SJ, Daunizeau J, Phillips C, Friston KJ (2008) Variational Bayesian inversion of the equivalent current dipole model in EEG/MEG. Neuroimage 39:728-741.
Kiebel SJ, Garrido MI, Moran R, Chen CC, Friston KJ (2009) Dynamic causal modeling for EEG and MEG. Human Brain Mapping 30:1866-1876.
Killcross S, Coutureau E (2003) Coordination of actions and habits in the medial prefrontal cortex of rats. Cereb Cortex 13:400-408.
Kilner JM, Kiebel SJ, Friston KJ (2005) Applications of random field theory to electrophysiology. Neurosci Lett 374:174-178.
Kim JN, Shadlen MN (1999) Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. Nat Neurosci 2:176-185.
Kim S, Hwang J, Lee D (2008) Prefrontal coding of temporally discounted values during intertemporal choice. Neuron 59:161-172.
King-Casas B, Tomlin D, Anen C, Camerer CF, Quartz SR, Montague PR (2005) Getting to know you: reputation and trust in a two-person economic exchange. Science 308:78-83.
Kitazawa S, Kimura T, Yin PB (1998) Cerebellar complex spikes encode both destinations and errors in arm movements. Nature 392:494-497.
Klucharev V, Hytönen K, Rijpkema M, Smidts A, Fernández G (2009) Reinforcement learning signal predicts social conformity. Neuron 61:140-151.
Knight F (1921) Risk, uncertainty and profit. Boston, MA: Houghton-Mifflin.
Knowlton BJ, Mangels JA, Squire LR (1996) A neostriatal habit learning system in humans. Science 273:1399-1402.
Knutson B, Taylor J, Kaufman M, Peterson R, Glover G (2005) Distributed neural representation of expected value. J Neurosci 25:4806-4812.
Knuutila JET, Ahonen AI, Hamalainen MS, Kajola MJ, Laine PP, Lounasmaa OV, Parkkonen LT, Simola JTA, Tesche CD (1993) A 122-channel whole-cortex SQUID system for measuring the brain's magnetic fields. IEEE Trans Magnetics 29:3315-3320.
Konorski J (1967) Integrative activity of the brain. Chicago: University of Chicago Press.
Kording KP, Wolpert DM (2004) Bayesian integration in sensorimotor learning. Nature 427:244-247.
Krajbich I, Armel C, Rangel A (2010) Visual fixations and the computation and comparison of value in simple choice. Nat Neurosci.
Kringelbach ML, Rolls ET (2004) The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. Prog Neurobiol 72:341-372.
Krugel LK, Biele G, Mohr PN, Li SC, Heekeren HR (2009) Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. Proc Natl Acad Sci USA 106:17951-17956.
Kumaran D, Summerfield JJ, Hassabis D, Maguire EA (2009) Tracking the emergence of conceptual knowledge during human decision making. Neuron 63:889-901.
Kunishio K, Haber SN (1994) Primate cingulostriatal projection: limbic striatal versus sensorimotor striatal input. J Comp Neurol 350:337-356.
Kutas M, Dale A (1997) Electrical and magnetic readings of mental functions. Cognitive Neuroscience (ed Rugg).
Kutas M, McCarthy G, Donchin E (1977) Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. Science 197:792-795.
Lau B, Glimcher PW (2008) Value representations in the primate striatum during matching behavior. Neuron 58:451-463.
Lee D, Wang XJ (2008) Mechanisms for stochastic decision-making in the primate frontal cortex: single-neuron recording and circuit modelling. In: Neuroeconomics: decision making and the brain(Glimcher, P. et al., eds), pp 487-499 San Diego, CA: Academic Press.
Lemus L, Hernandez A, Romo R (2009) Neural codes for perceptual discrimination of acoustic flutter in the primate auditory cortex. Proc Natl Acad Sci U S A 106:9471-9476.
Lewis DA, Campbell MJ, Foote SL, Goldstein M, Morrison JH (1987) The distribution of tyrosine hydroxylase-immunoreactive fibers in primate neocortex is widespread but regionally specific. J Neurosci 7:279-290.
Lin FH, Witzel T, Ahlfors SP, Stufflebeam SM, Belliveau JW, Hamalainen MS (2006) Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. Neuroimage 31:160-171.
Lo CC, Boucher L, Pare M, Schall JD, Wang XJ (2009) Proactive inhibitory control and attractor dynamics in countermanding action: a spiking neural circuit model. J Neurosci 29:9059-9071.
Lo CC, Wang XJ (2006) Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. Nat Neurosci 9:956-963.
Logothetis NK (2008) What we can do and what we cannot do with fMRI. Nature 453:869-878.

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. Nature 412:150-157.
Lorente de No R (1947) Action potential of the motoneurons of the hypoglossus nucleus. J Cell Physiology 29:207-287.
Louie K, Glimcher PW (2010) Separating value from choice: delay discounting activity in the lateral intraparietal area. J Neurosci 30:5498-5507.
Luck SJ (2005) An introduction to the event-related potential technique. Boston: MIT Press.
Mainen ZF, Sejnowski TJ (1996) Influence of dendritic structure on firing pattern in model neocortical neurons. Nature 382:363-366.
Malmivuo J, Plonsey R (1995) Bioelectromagentism - principles and applications of bioelectric and biomagentic fields. New York: Oxford University Press.
Marinkovic K, Cox B, Reid K, Halgren E (2004) Head position in the MEG helmet affects the sensitivity to anterior sources. Neurol Clin Neurophysiol 2004:30.
Marr D (1969) A theory of cerebellar cortex. J Physiol 202:437-470.
Mars RB, Jbabdi S, Sallet J, O'Relly JX, Croxson PL, Olivier E, Noonan MP, Bergmann C, Mitchell AS, Baxter MG, Behrens TEJ, Johansen-Berg H, Tomassini V, Miller KL, Rushworth MFS (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.
Marshall A (1920) Principles of economics: an introductory volume. London: Macmillan.
Matsumoto K, Suzuki W, Tanaka K (2003) Neuronal correlates of goal-based motor selection in the prefrontal cortex. Science 301:229-232.
Matsumoto M, Hikosaka O (2009) Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature 459:837-841.
Matsumoto M, Matsumoto K, Abe H, Tanaka K (2007) Medial prefrontal cell activity signaling prediction errors of action values. Nat Neurosci 10:647-656.
Mattout J, Henson RN, Friston K (2007) Canonical source reconstruction for MEG. Comput Intelligence Neurosci doi: 10.1155/2007/67613.
Maynard Smith J (1982) Evolution and the theory of games. Cambridge: Cambridge University Press.
McClure SM, Berns GS, Montague PR (2003) Temporal prediction errors in a passive learning task activate human striatum. Neuron 38:339-346.
McCoy AN, Platt ML (2005) Risk-sensitive neurons in macaque posterior cingulate cortex. Nat Neurosci 8:1220-1227.
McGlothin WH (1956) Stability of choices among uncertain alternatives. Am J Psychol 69:604-615.
Meltzer JA, Zaveri HP, Goncharova, II, Distasio MM, Papademetris X, Spencer SS, Spencer DD, Constable RT (2008) Effects of working memory load on oscillatory power in human intracranial EEG. Cereb Cortex 18:1843-1855.
Meunier M, Bachevalier J, Mishkin M (1997) Effects of orbital frontal and anterior cingulate lesions on object and spatial memory in rhesus monkeys. Neuropsychologia 35:999-1015.
Milosavljevic M, Malmaud J, Huth A, Koch C, Rangel A (2010) The drift diffusion model can account for the accuracy and reaction time of value-based choices under high and low time pressure. Judgement and decision making 5:437-449.
Milstein JN, Koch C (2008) Dynamic moment analysis of the extracellular electric field of a biologically realistic neuron. Neural Comput 2070.
Miltner WHR, Braun CH, Coles MGH (1997) Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "generic" neural system for error detection. In: Journal of cognitive neuroscience, vol. 9, pp 788-798.
Mirenowicz J, Schultz W (1994) Importance of unpredictability for reward responses in primate dopamine neurons. J Neurophysiol 72:1024-1027.
Mishkin M (1964) Perseveration of central sets after frontal lesions in monkeys. In: The frontal granular cortex and behavior(Warren, J. M. and Akert, K., eds) New York: McGraw-Hill.
Mitzdorf U (1985) Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. Physiol Rev 65:37-100.
Mitzdorf U, Singer W (1979) Excitatory synaptic ensemble properties in the visual cortex of the macaque monkey: a current source density analysis of electrically evoked potentials. J Comp Neurol 187:71-83.
Mogenson GJ, Takigawa M, Robertson A, Wu M (1979) Self-stimulation of the nucleus accumbens and ventral tegmental area of Tsai attenuated by microinjections of spiroperidol into the nucleus accumbens. Brain Res 171:247-259.
Morecraft RJ, Van Hoesen GW (1992) Cingulate input to the primary and supplementary motor cortices in the rhesus monkey: evidence for somatotopy in areas 24c and 23c. J Comp Neurol 322:471-489.
Mountcastle VB, Steinmetz MA, Romo R (1990) Frequency discrimination in the sense of flutter: psychophysical measurements correlated with postcentral events in behaving monkeys. J Neurosci 10:3032-3044.
Murakami S, Okada Y (2006) Contributions of principal neocortical neurons to magnetoencephalography and electroencephalography signals. J Physiol 575:925-936.

Muthukumaraswamy SD, Edden RA, Jones DK, Swettenham JB, Singh KD (2009) Resting GABA concentration predicts peak gamma frequency and fMRI amplitude in response to visual stimulation in humans. Proc Natl Acad Sci U S A 106:8356-8361.
Nazir TA, Jacobs AM (1991) The effects of target discriminability and retinal eccentricity on saccade latencies: an analysis in terms of variable-criterion theory. Psychol Res 53:281-289.
Newell BR, Rakow T (2007) The role of experience in decisions from description. Psychon Bull Rev 14:1133-1139.
Neyman J, Pearson E (1933) On the problem of the most efficient tests of statistical hypotheses. Phil Trans Roy Soc A 231:289-337.
Nolte G (2003) The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. Phys Med Biol 48:3637-3652.
Noonan MP, Sallet J, Rudebeck PH, Buckley MJ, Rushworth MF (2010a) Does the medial orbitofrontal cortex have a role in social valuation? Eur J Neurosci 31:2341-2351.
Noonan MP, Walton ME, Behrens TE, Sallet J, Buckley MJ, Rushworth MF (2010b) Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. Proc Natl Acad Sci U S A 107:20547-20552.
Nowak M, Sigmund K (1993) A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. Nature 364:56-58.
O'Doherty J, Rolls ET, Francis S, Bowtell R, McGlone F, Kobal G, Renner B, Ahne G (2000) Sensory-specific satiety-related olfactory activation of the human orbitofrontal cortex. Neuroreport 11:399-403.
O'Doherty JP (2004) Reward representations and reward-related learning in the human brain: insights from neuroimaging. Curr Opin Neurobiol 14:769-776.
O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ (2003a) Temporal difference models and reward-related learning in the human brain. Neuron 38:329-337.
O'Doherty JP, Dayan P, Friston KJ, Critchley H, Dolan RJ (2003b) Temporal difference models and reward-related learning in the human brain. Neuron 38:329-337.
O'Doherty JP, Dayan P, Schultz J, Deichmann R, Friston KJ, Dolan RJ (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science 304:452-454.
O'Doherty JP, Hampton A, Kim H (2007) Model-based fMRI and its application to reward learning and decision making. Ann N Y Acad Sci 1104:35-53.
O'Doherty JP, Kringelbach ML, Rolls ET, Hornak J, Andrews C (2001) Abstract reward and punishment representations in the human orbitofrontal cortex. Nat Neurosci 4:95-102.
Ogawa S, Lee TM, Kay AR, Tank DW (1990) Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc Natl Acad Sci U S A 87:9868-9872.
Olds J, Milner P (1954) Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. J Comp Physiol Psychol 47:419-427.
Oliveira FT, McDonald JJ, Goodman D (2007) Performance monitoring in the anterior cingulate is not all error related: expectancy deviation and the representation of action-outcome associations. J Cogn Neurosci 19:1994-2004.
Ongür D, An X, Price JL (1998) Prefrontal cortical projections to the hypothalamus in macaque monkeys. J Comp Neurol 401:480-505.
Oostenveld R, Fries P, Maris E, Schloffen JM (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intelligence Neurosci doi:10.1155/2011/156869.
Ostlund SB, Balleine BW (2007) Orbitofrontal cortex mediates outcome encoding in Pavlovian but not instrumental conditioning. J Neurosci 27:4819-4825.
Padoa-Schioppa C (2010) Neurobiology of Economic Choice: A Good-Based Model. Annu Rev Neurosci.
Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. Nature 441:223-226.
Pascual-Marqui RD, Michel CM, Lehmann D (1994) Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. Int J Psychophysiol 18:49-65.
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.
Pauling L, Coryell CD (1936) The Magnetic Properties and Structure of Hemoglobin, Oxyhemoglobin and Carbonmonoxyhemoglobin. Proc Natl Acad Sci U S A 22:210-216.
Pavlov I (1927) Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex. London: Oxford University Press.
Pearce JM, Bouton ME (2001) Theories of associative learning in animals. Annu Rev Psychol 52:111-139.
Pearce JM, Hall G (1980) A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychol Rev 87:532-552.
Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD (2006) Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature 442:1042-1045.
Pitt MA, Myung IJ (2002) When a good fit can be bad. Trends Cogn Sci (Regul Ed) 6:421-425.
Plassmann H, O'Doherty JP, 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.
Plonsey R (1977) Action potential sources and their volume conductor fields. Proc IEEE 65:601-611.
Preuschoff K, Bossaerts P (2007) Adding prediction risk to the theory of reward learning. Ann N Y Acad Sci 1104:135-146.
Preuschoff K, Quartz SR, Bossaerts P (2008) Human insula activation reflects risk prediction errors as well as risk. J Neurosci 28:2745-2752.
Price JL (2007) Definition of the orbital cortex in relation to specific connections with limbic and visceral structures and other cortical regions. Ann N Y Acad Sci 1121:54-71.
Ramnani N, Miall RC (2004) A system in the human brain for predicting the actions of others. Nat Neurosci 7:85-90.
Rangel A, Camerer CF, Montague PR (2008) A framework for studying the neurobiology of value-based decision making. Nature Reviews Neuroscience 9:545-556.
Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79-87.
Ratcliff R (1978) A theory of memory retrieval. Psy Rev 85:59-108.
Ratcliff R, Rouder JN (1998) Modeling response times for two-choice decisions. Psychological Science 9:347-356.
Ratcliff R, Van Zandt T, McKoon G (1999) Connectionist and diffusion models of reaction time. Psychol Rev 106:261-300.
Rendell L, Boyd R, Cownden D, Enquist M, Eriksson K, Feldman MW, Fogarty L, Ghirlanda S, Lillicrap T, Laland KN (2010) Why copy others? Insights from the social learning strategies tournament. Science 328:208-213.
Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Classical conditioning II: current research and theory(Black, A. H. and Prokasy, W. F., eds), pp 64-99 New York: Appleton-Century-Crofts.
Rilling JK, Sanfey AG, Aronson JA, Nystrom LE, Cohen JD (2004) The neural correlates of theory of mind within interpersonal interactions. Neuroimage 22:1694-1703.
Rizzolatti G, Fadiga L, Gallese V, Fogassi L (1996) Premotor cortex and the recognition of motor actions. Brain Res Cogn Brain Res 3:131-141.
Robbins TW, Everitt BJ (1992) Functions of dopamine in the dorsal and ventral striatum. Seminars in Neuroscience 4:119-127.
Roesch MR, Olson CR (2004) Neuronal activity related to reward value and motivation in primate frontal cortex. Science 304:307-310.
Roitman JD, Shadlen MN (2002) Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. J Neurosci 22:9475-9489.
Rolls ET, Critchley HD, Mason R, Wakeman EA (1996) Orbitofrontal cortex neurons: role in olfactory and visual association learning. J Neurophysiol 75:1970-1981.
Rolls ET, Grabenhorst F, Deco G (2010) Choice, difficulty, and confidence in the brain. Neuroimage 53:694-706.
Rolls ET, Hornak J, Wade D, McGrath J (1994) Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. J Neurol Neurosurg Psychiatry 57:1518-1524.
Rolls ET, McCabe C, Redoute J (2008) Expected value, reward outcome, and temporal difference error representations in a probabilistic decision task. Cereb Cortex 18:652-663.
Rolls ET, Sienkiewicz ZJ, Yaxley S (1989) Hunger Modulates the Responses to Gustatory Stimuli of Single Neurons in the Caudolateral Orbitofrontal Cortex of the Macaque Monkey. Eur J Neurosci 1:53-60.
Romo R, Schultz W (1990) Dopamine neurons of the monkey midbrain: contingencies of responses to active touch during self-initiated arm movements. J Neurophysiol 63:592-606.
Rosenkilde CE, Bauer RH, Fuster JM (1981) Single cell activity in ventral prefrontal cortex of behaving monkeys. Brain Res 209:375-394.
Rudebeck PH, Behrens TEJ, Kennerley SW, Baxter MG, Buckley MJ, Walton ME, Rushworth MSF (2008) Frontal Cortex Subregions Play Distinct Roles in Choices between Actions and Stimuli. J Neurosci 28:13775-13785.
Rudebeck PH, Buckley MJ, Walton ME, Rushworth MSF (2006a) A role for the macaque anterior cingulate gyrus in social valuation. Science 313:1310-1312.
Rudebeck PH, Murray EA (2008) Amygdala and orbitofrontal cortex lesions differentially influence choices during object reversal learning. J Neurosci 28:8338-8343.
Rudebeck PH, Walton ME, Smyth AN, Bannerman DM, Rushworth MSF (2006b) Separate neural pathways process different decision costs. Nat Neurosci 9:1161-1168.
Rushworth MF, Hadland KA, Gaffan D, Passingham RE (2003) The effect of cingulate cortex lesions on task switching and working memory. J Cogn Neurosci 15:338-353.
Rushworth MF, Mars RB, Summerfield C (2009) General mechanisms for making decisions? Curr Opin Neurobiol.
Rushworth MSF, Behrens TEJ (2008) Choice, uncertainty and value in prefrontal and cingulate cortex. Nat Neurosci 11:389-397.

Rushworth MSF, Behrens TEJ, Rudebeck PH, Walton ME (2007a) Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. Trends Cogn Sci (Regul Ed) 11:168-176.
Rushworth MSF, Buckley MJ, Behrens TEJ, Walton ME, Bannerman DM (2007b) Functional organization of the medial frontal cortex. Curr Opin Neurobiol 17:220-227.
Rushworth MSF, Walton ME, Kennerley SW, Bannerman DM (2004) Action sets and decisions in the medial frontal cortex. Trends Cogn Sci (Regul Ed) 8:410-417.
Rutledge RB, Lazzaro SC, Lau B, Myers CE, Gluck MA, Glimcher PW (2009) Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. J Neurosci 29:15104-15114.
Salamone JD, Correa M (2002) Motivational views of reinforcement: implications for understanding the behavioral functions of nucleus accumbens dopamine. Behav Brain Res 137:3-25.
Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. Science 310:1337-1340.
Samuelson PA (1938) A note on the pure theory of consumer's behaviour. Economica 5:61-71.
Sarvas J (1987) Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. Phys Med Biol 32:11-22.
Saxe R (2005) Against simulation: the argument from error. Trends Cogn Sci 9:174-179.
Saxe R (2006) Uniquely human social cognition. Curr Opin Neurobiol 16:235-239.
Scheperjans F, Hermann K, Eickhoff SB, Amunts K, Schleicher A, Zilles K (2008) Observer-independent cytoarchitectonic mapping of the human superior parietal cortex. Cereb Cortex 18:846-867.
Scherg M (1990) Fundamentals of dipole source potential analysis. In: Auditory evoked magnetic fields and electric potentials, vol. 6 (Grandori, F. et al., eds), pp 40-69 Basel: Karger.
Schoenbaum G, Chiba AA, Gallagher M (1998) Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. Nat Neurosci 1:155-159.
Schoenbaum G, Roesch MR, Stalnaker TA, Takahashi YK (2009) A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. Nature Reviews Neuroscience 10:885-892.
Schultz W (1998) Predictive reward signal of dopamine neurons. J Neurophysiol 80:1-27.
Schultz W, Apicella P, Ljungberg T (1993) Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. J Neurosci 13:900-913.
Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593-1599.
Seo H, Barraclough DJ, Lee D (2009) Lateral intraparietal cortex and reinforcement learning during a mixed-strategy game. J Neurosci 29:7278-7289.
Seo H, Lee D (2007) Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. J Neurosci 27:8366-8377.
Serences JT (2008) Value-based modulations in human visual cortex. Neuron 60:1169-1181.
Seymour B, O'Doherty JP, Dayan P, Koltzenburg M, Jones AK, Dolan RJ, Friston KJ, Frackowiak RS (2004) Temporal difference models describe higher-order learning in humans. Nature 429:664-667.
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.
Sheffield FD (1965) Relation between classical and instrumental conditioning. In: Classical conditioning(Prokasy, W. F., ed), pp 302-322 New York, NY: Appleton Century Crofts.
Shepherd SV, Deaner RO, Platt ML (2006) Social status gates social attention in monkeys. Curr Biol 16:R119-120.
Shidara M, Richmond BJ (2002) Anterior cingulate: single neuronal signals related to degree of reward expectancy. Science 296:1709-1711.
Shima K, Tanji J (1998) Role for cingulate motor area cells in voluntary movement selection based on reward. Science 282:1335-1338.
Shizgal P (1997) Neural basis of utility estimation. Curr Opin Neurobiol 7:198-208.
Shuler MG, Bear MF (2006) Reward timing in the primary visual cortex. Science 311:1606-1609.
Sigman M, Dehaene S (2005) Parsing a cognitive task: a characterization of the mind's bottleneck. PLoS Biol 3:e37.
Silk JB (2007) Social components of fitness in primate groups. Science 317:1347-1351.
Singer T, Seymour B, O'Doherty JP, Stephan KE, Dolan RJ, Frith CD (2006) Empathic neural responses are modulated by the perceived fairness of others. Nature 439:466-469.
Smith SM (2002) Fast robust automated brain extraction. Hum Brain Mapp 17:143-155.
Smith SM (2004) Overview of fMRI analysis. The British journal of radiology 77 Spec No 2:S167-175.
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 (2004a) Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 Suppl 1:S208-219.
Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, 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 (2004b) Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 Suppl 1:S208-219.
Snyder LH, Batista AP, Andersen RA (1997) Coding of intention in the posterior parietal cortex. Nature 386:167-170.
Soltani A, Wang XJ (2006) A biophysically based neural model of matching law behavior: melioration by stochastic synapses. J Neurosci 26:3731-3744.
Soltani A, Wang XJ (2010) Synaptic computation underlying probabilistic inference. Nat Neurosci 13:112-119.
Spruston N (2008) Pyramidal neurons: dendritic structure and synaptic integration. Nat Rev Neurosci 9:206-221.
Spruston N, Jonas P, Sakmann B (1995) Dendritic glutamate receptor channels in rat hippocampal CA3 and CA1 pyramidal neurons. J Physiol 482 (Pt 2):325-352.
Sugrue LP, Corrado GS, Newsome WT (2004) Matching behavior and the representation of value in the parietal cortex. Science 304:1782-1787.
Sul JH, Kim H, Huh N, Lee D, Jung MW (2010) Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. Neuron 66:449-460.
Summerfield C, Egner T, Greene M, Koechlin E, Mangels J, Hirsch J (2006) Predictive codes for forthcoming perception in the frontal cortex. Science 314:1311-1314.
Sutton R, Barto A (1998) Reinforcement learning: an introduction. Cambridge, MA: MIT Press.
Sutton RS, Barto AG (1990) Time-derivative models of Pavlovian reinforcement. In: Learning and computational neuroscience: foundations of adaptive network(Gabriel, M. and Moore, J., eds), pp 497-537 Boston, MA: MIT Press.
Swisher JD, Halko MA, Merabet LB, McMains SA, Somers DC (2007) Visual topography of human intraparietal sulcus. J Neurosci 27:5326-5337.
Takahashi YK, Roesch MR, Stalnaker TA, Haney RZ, Calu DJ, Taylor AR, Burke KA, Schoenbaum G (2009) The orbitofrontal cortex and ventral tegmental area are necessary for learning from unexpected outcomes. Neuron 62:269-280.
Tallon-Baudry C, Bertrand O (1999) Oscillatory gamma activity in humans and its role in object representation. Trends Cogn Sci 3:151-162.
Tallon-Baudry C, Bertrand O, Delpuech C, Permier J (1997) Oscillatory gamma-band (30-70 Hz) activity induced by a visual search task in humans. J Neurosci 17:722-734.
Tanaka SC, Doya K, Okada G, Ueda K, Okamoto Y, Yamawaki S (2004) Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. Nat Neurosci 7:887-893.
Taulu S, Kajola M, Simola J (2004) Suppression of interference and artifacts by the Signal Space Separation Method. Brain Topogr 16:269-275.
Taulu S, Simola J (2006) Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Phys Med Biol 51:1759-1768.
Thorndike EL (1911) Animal Intelligence. Bristol: Thoemmes Press.
Thorpe SJ, Rolls ET, Maddison S (1983) The orbitofrontal cortex: neuronal activity in the behaving monkey. Experimental brain research Experimentelle Hirnforschung Expérimentation cérébrale 49:93-115.
Tikhonov AN, Arsenin VY (1977) Solution of ill-posed problems. Washington: Winston & Sons.
Tobler PN, Fiorillo CD, Schultz W (2005) Adaptive coding of reward value by dopamine neurons. Science 307:1642-1645.
Tolman EC (1948) Cognitive maps in rats and men. Psychol Rev 55:189-208.
Tom SM, Fox CR, Trepel C, Poldrack RA (2007) The neural basis of loss aversion in decision-making under risk. Science 315:515-518.
Tomlin D, Kayali MA, King-Casas B, Anen C, Camerer CF, Quartz SR, Montague PR (2006) Agent-specific responses in the cingulate cortex during economic exchanges. Science 312:1047-1050.
Tremblay L, Schultz W (1999) Relative reward preference in primate orbitofrontal cortex. Nature 398:704-708.
Tsujimoto S, Genovesio A, Wise SP (2009) Monkey orbitofrontal cortex encodes response choices near feedback time. J Neurosci 29:2569-2574.
Tversky A, Kahneman D (1992) Advances in prospect theory: cumulative representation of uncertainty. J Risk Uncert 5:297-323.
Uchida N, Mainen ZF (2003) Speed and accuracy of olfactory discrimination in the rat. Nat Neurosci 6:1224-1229.
Ungless MA, Magill PJ, Bolam JP (2004) Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. Science 303:2040-2042.
Usher M, McClelland JL (2001) The time course of perceptual choice: the leaky, competing accumulator model. Psychological review 108:550-592.
Van Hoesen GW, Morecraft RJ, Vogt BA (1993) In: Neurobiology of cingulate cortex and limbic thalamus(Vogt, B. A. and Gabriel, M., eds).
Van Overwalle F (2009) Social cognition and the brain: a meta-analysis. Human brain mapping 30:829-858.
van Veen BD, Buckley KM (1988) Beamforming: a versatile approach to spatial filtering. ASSP Magazine, IEEE 5:4-24.

van Veen BD, W. vD, Yuchtman M, Suzuki A (1997) Localization of brain electrical activity via linearly constrained minumum variance beamforming. IEEE Trans Biomedical Engineering 44:867-880.
Vanduffel W, Fize D, Peuskens H, Denys K, Sunaert S, Todd JT, Orban GA (2002) Extracting 3D from motion: differences in human and monkey intraparietal cortex. Science 298:413-415.
VanVeen BD, vanDrongelen W, Yuchtman M, Suzuki A (1997) Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. Ieee Transactions on Biomedical Engineering 44:867-880.
Vickers D (1970) Evidence for an accumulator model of psychophysical discrimination. Ergonomics 13:37-58.
von Helmholtz H (1853) Über einige Gesetze der Vertheilung elektrischer Ströme in körperlichen Leitern, mit Anwendung auf die thierisch-elektrischen Versuche. Annals of Physics and Chemistry 89:211-233.
von Neumann J, Morgenstern O (1944) Theory of games and economic behavior. Princeton, NJ: Princeton University Press.
Vrba J (2002) Magnetoencephalography: the art of finding a needle in a haystack. In: Physica C, vol. 368, pp 1-9.
Vrba J, Robinson SE (2001) Signal processing in magnetoencephalography. Methods 25:249-271.
Waelti P, Dickinson A, Schultz W (2001) Dopamine responses comply with basic assumptions of formal learning theory. Nature 412:43-48.
Wald A (1947) Sequential analysis. New York: John Wiley and Sons.
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.
Walton ME, Croxson PL, Rushworth MF, Bannerman DM (2005) The mesocortical dopamine projection to anterior cingulate cortex plays no role in guiding effort-related decisions. Behav Neurosci 119:323-328.
Walton ME, Devlin JT, Rushworth MSF (2004) Interactions between decision making and performance monitoring within prefrontal cortex. Nat Neurosci 7:1259-1265.
Wang XJ (2002) Probabilistic decision making by slow reverberation in cortical circuits. Neuron 36:955-968.
Wang XJ (2008) Decision making in recurrent neuronal circuits. Neuron 60:215-234.
Watanabe M (1996) Reward expectancy in primate prefrontal neurons. Nature 382:629-632.
Watson JB (1913) Psychology as the behaviorist views it. Psychol Rev 20:158-177.
Weiss Y, Simoncelli EP, Adelson EH (2002) Motion illusions as optimal percepts. Nat Neurosci 5:598-604.
White NM (1989) Reward or reinforcement: what's the difference? Neurosci Biobehav Rev 13:181-186.
Williams SM, Goldman-Rakic PS (1993) Characterization of the dopaminergic innervation of the primate frontal cortex using a dopamine-specific antibody. Cereb Cortex 3:199-222.
Wipf D, Nagarajan S (2009) A unified Bayesian framework for MEG/EEG source imaging. Neuroimage 44:947-966.
Wise RA, Rompre PP (1989) Brain dopamine and reward. Annu Rev Psychol 40:191-225.
Wise RA, Spindler J, deWit H, Gerberg GJ (1978) Neuroleptic-induced "anhedonia" in rats: pimozide blocks reward quality of food. Science 201:262-264.
Wiskwo JP (1995) SQUID magnetometers for biomagnetism and nondestructive testing: important questions and initial answers. IEEE Transactions on Applied Superconductivity 5:74-120.
Wolpert DM, Doya K, Kawato M (2003) A unifying computational framework for motor control and social interaction. Philos Trans R Soc Lond B Biol Sci 358:593-602.
Wolters CH, Anwander A, Tricoche X, Weinstein D, Koch MA, MacLeod RS (2006) Influence of tissue conductivity anisotropy on EEG/MEG field and return current computation in a realistic head model: a simulation and visualization study using high-resolution finite element modeling. Neuroimage 30:813-826.
Wong KF, Huk AC, Shadlen MN, Wang XJ (2007) Neural circuit dynamics underlying accumulation of time-varying evidence during perceptual decision making. Front Comput Neurosci 1:6.
Wong KF, Wang XJ (2006) A recurrent network mechanism of time integration in perceptual decisions. J Neurosci 26:1314-1328.
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.
Woolrich MW, Ripley BD, Brady M, Smith SM (2001) Temporal autocorrelation in univariate linear modeling of FMRI data. Neuroimage 14:1370-1386.
Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC (1996) A unified statistical approach for determining significant signals in images of cerebral activation. Hum Brain Mapp 4:58-73.
Wunderlich K, Rangel A, O'Doherty JP (2009) Neural computations underlying action-based decision making in the human brain. Proc Natl Acad Sci USA.
Yang T, Shadlen MN (2007) Probabilistic reasoning by neurons. Nature 447:1075-1080.
Yin HH, Knowlton BJ, Balleine BW (2004) Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. Eur J Neurosci 19:181-189.
Yin HH, Ostlund SB, Knowlton BJ, Balleine BW (2005) The role of the dorsomedial striatum in instrumental conditioning. Eur J Neurosci 22:513-523.

Yoshida W, Seymour B, Friston KJ, Dolan RJ (2010) Neural mechanisms of belief inference during cooperative games. J Neurosci 30:10744-10751.
Yu AJ, Dayan P (2005) Uncertainty, neuromodulation, and attention. Neuron 46:681-692.
Yuval-Greenberg S, Tomer O, Keren AS, Nelken I, Deouell LY (2008) Transient induced gamma-band response in EEG as a manifestation of miniature saccades. Neuron 58:429-441.
Zald DH, Boileau I, El-Dearedy W, Gunn R, McGlone F, Dichter GS, Dagher A (2004) Dopamine transmission in the human striatum during monetary reward tasks. J Neurosci 24:4105-4112.