Mechanisms for novel multidimensional choices in prefrontal cortex

Behavioural, imaging and causal approaches to non-human primate decision-making

Alessandro Bongioanni

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Experimental Psychology at the University of Oxford

Wolfson College

Trinity term 2018

γνῶθι σαυτόν

know thyself

[from the temple of Apollo in Delphi]

Mechanisms for novel multidimensional choices in prefrontal cortex - Behavioural, imaging and causal approaches to non-human primate decision-making

Abstract

In this thesis the cognitive and neuro-computational bases of novel value-based decision-making in non-human primates are studied. Humans routinely perform choices among options that have not previously been experienced, using value expectations derived from their component attributes. Animal studies, however, are typically performed after extensive training ensuring familiarity with the task. Here we show how Rhesus macaques can readily perform value-guided choices among novel stimuli – if their components are known. An economic analysis of subjective biases is provided. Neuroimaging and causal techniques are then used to uncover a network centred in anterior medial frontal cortex (AMFC) underlying adaptive novel decisions. A first experiment, using fMRI with a binary decision task, reveals that orbitofrontal cortex is equally modulated by value in novel and familiar choices, but in contrast AMFC and dorsolateral prefrontal cortex (dlPFC) show a markedly different pattern of responses between conditions. A second experiment, using fMRI and single option trials, reveals how AMFC and dlPFC, but no other regions, encode representations of abstract value independent from its components (reward magnitude and probability), appearing across sessions with exposure to the novel options. Additionally, a pattern of hexagonal symmetry is found in AMFC activity, indicating a grid-like encoding of two-dimensional value space. Such a code provides a mechanistic explanation of how novel options' values may be computed. A last experiment confirms this by showing that transiently disrupting AMFC with focused transcranial ultrasound impairs the use of an integrated representation of value when novel decisions are made. Together, these experiments bridge neuroscientific studies in humans and monkeys and reveal a neural basis for novel decision-making that has not previously been investigated in either species.

Alessandro Bongioanni

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Experimental Psychology at University of Oxford
Word count: 50.000
Wolfson College
Trinity term 2018

Acknowledgements

I am first of all most grateful to my supervisor, Matthew Rushworth, who has been incredibly nice in all circumstances, since the first Skype meeting we had when Oxford was flooded and the first retreat in the Cotswolds four years ago. His unassuming guidance has helped me through the many difficulties of this project and academic life, and his brilliant scientific mind as well as personal friendliness and generosity have meant a lot to me during my DPhil. He is definitely the best supervisor I could have ever had.

All the members of my lab have been incredibly helpful and kind and I could not have completed the present work without such an amazing team. In particular, Urs Schuffelgen, as scanner manager, helped me endless times in training and testing monkeys in the MRI scanner. He trained me as scanner operator and shared with me all his pro tips any time I encountered problems. Jerome Sallet “le roi” of Biomedical Services Building (BSB) was always behind the smooth functioning of everything related to monkey testing. He performed the surgeries and the structural scans for my experiment, and I am very grateful for his help. He greatly improved the quality of my time in BSB with the boxes of chocolate he regularly brought in. Davide Folloni – the only person I could speak Italian with – has taught me all the secrets of TUS (that we still affectionately call FUN) and has painstakingly assisted me through all the neurostimulation experiment. Georgious Papageorgiou was the first person who trained me to perform macaque MRI testing, and I will remember his Spartan morale. Even if Bolton Chau left shortly after I arrived in Oxford, he has been the person – with Jerome – who set up most of the NHP testing tools, techniques and software in BSB, and it is not an exaggeration to say that my research could not have happened without them. Rhyanne

Dale, Andrew Emberton, Jessica Wood, Claire Pouillas and Kelly Simpson have all been wonderful research assistants. They all worked very hard with me in the training and caring of the animals and replaced me when I was busy. I think these experiments would have taken double the time without them. Also thanks again to all the people mentioned so far for the cheerful atmosphere they brought during the long hours spent together in the MRI control room. Many more people who work in BSB have contributed to the success of my experiments as well as the well-being of the monkeys, in particular Maria Martinez, Caroline Bergmann, Tamara Ramadan, April Shipton and all the BMS staff, as well as other lab members who helped me occasionally with testing like Nima Khalighinejad, Kentaro Miyamoto and Kevin Marche. Most of all, I wish to thank my four wonderful monkeys, Ulrich, Ultra, Vampire and Winky, who have done an amazing job in these and other experiments and who are truly beautiful animals. I always aimed to treat them with great respect.

Outside the monkey world, the person who has played the biggest part in my doctoral project has been Miriam Klein-Flügge; not only has she regularly given me very insightful ideas and trained me in the intricacies of fMRI designs, but she has always been so positive and enthusiastic about what I was doing that I couldn't help but share in some of her passion. Lennart Verhagen did an incredible job for our team by writing most of the code that we now use for behaving macaque MRI preprocessing. Marco Wittman, Nils Kolling Jacqueline Scholl, Lev Tankelevich, Léa Roumazeilles, Matthias Schurz, Alizée Lopez, Elsa Fouragnan have all discussed with me about this project and shared recommendations, scripts or ideas. I also remember a useful discussion with Alexa Constantinescu while sharing a traditional hotpot in Shanghai.

To conclude I would like to mention all my friends around the world, and especially here in Oxford at Wolfson college, who have been so nice to me, and with whom I

enjoyed living and studying these last four years. I have been very lucky to have met you. But above all, I would like to dedicate this work to the most important people in my life, who have always been beside me – even at distance – and without whom I am sure I would have never been able to complete this doctorate: my parents Tania and Carlo, and my partner, now fiancée, Jingyue Xing. Thanks!

Table of Contents

ABSTRACT **3**

ACKNOWLEDGEMENTS **4**

CHAPTER 1 GENERAL INTRODUCTION: NOVEL AND FAMILIAR VALUE-GUIDED DECISION-MAKING IN THE BRAIN **10**

1.1 STAGE SETTING 10

1.1.1 What are novelty and familiarity? A philosophical primer *10*

1.1.2 Birth of the scientific understanding of decision-making *12*

1.1.3 The rationale of the present experimental work *15*

1.2 THE BEHAVIOURAL STUDY OF DECISION-MAKING 18

1.2.1 Economic and ecologic approaches to decision-making *18*

1.2.2 Decision-making under risk and behavioural biases: theory *21*

1.2.3 Decision-making under risk and behavioural biases: experiments in primates *23*

1.3 THE COGNITIVE NEUROSCIENCE OF DECISION-MAKING 27

1.3.1 The structure of a decision: conceptual and computational aspects *27*

1.3.2 An overview of the anatomy of prefrontal cortex *35*

1.3.3 Distributed neural circuits for value-guided decision-making in prefrontal cortex 40

1.3.4 Habitual and goal-directed behaviour *52*

1.4 FUNCTIONAL MAGNETIC RESONANCE IMAGING FOR COGNITIVE NEUROSCIENCE 57

1.5 THE PRESENT THESIS 65

CHAPTER 2 MACAQUE MONKEYS FACING NOVEL RISKY CHOICES. A COGNITIVE AND ECONOMIC APPROACH **68**

2.1 INTRODUCTION 68

2.2 METHODS 71

2.2.1 Overview *71*

2.2.2 Experimental setup *72*

2.2.3 Task design: the value space, stimuli, timing and conditions *74*

2.2.4 Training and test stages *81*

2.2.5 Model-free behavioural analysis *83*

2.2.6 Subjective value modelling *86*

2.3 RESULTS 94

2.3.1 Model-free behavioural analysis *94*

2.3.2 Subjective value modelling *99*

2.4 DISCUSSION 104

CHAPTER 3 NEURAL CORRELATES OF NOVEL MULTIDIMENSIONAL CHOICES IN MACAQUE MONKEYS 111

3.1 INTRODUCTION 111

3.2 METHODS 114

3.2.1 Task design *114*

3.2.2 Head-post surgery *118*

3.2.3 MRI setup and data acquisition *120*

3.2.4 fMRI data pre-processing *122*

3.2.5 fMRI whole-brain data analysis *124*

3.2.6 fMRI ROI-based data analysis and timecourse analysis *131*

3.3 RESULTS 132

3.4 DISCUSSION 150

CHAPTER 4 NEURAL REPRESENTATIONS OF REWARD-PREDICTIVE STIMULI IN VALUE SPACE WITHIN PREFRONTAL CORTEX 159

4.1 INTRODUCTION 159

4.2 METHODS 163

4.2.1 Experimental design and task *163*

4.2.2 Repetition suppression design and analysis *166*

4.2.3 Grid-like code analysis *172*

4.3 RESULTS 179

4.3.1 Repetition suppression *179*

4.3.2 Grid-like code *182*

4.4 DISCUSSION 185

CHAPTER 5 EFFECTS OF MEDIAL PREFRONTAL ULTRASOUND STIMULATION ON NOVEL MULTIDIMENSIONAL VALUE INTEGRATION 192

5.1 INTRODUCTION 192

5.2 METHODS 197

5.2.1 Experimental design and task *197*

5.2.2 Focused transcranial ultrasound neurostimulation procedure *200*

5.2.3 Data analysis and regression simulation *203*

5.3 RESULTS 206

5.4 DISCUSSION 208

CHAPTER 6 GENERAL DISCUSSION 211

REFERENCES 217

Chapter 1
General introduction: Novel and familiar value-guided decision-making in the brain

The sun is novel each day

Heraclitus

1.1 Stage setting

1.1.1 What are novelty and familiarity? A philosophical primer

The ancient philosopher Heraclitus from Ephesus (circa 535-475 BC), renowned for his deep and obscure thinking, challenged a principle upon which our common understanding of the world is based, namely that things (substances) are stable (subsist). For him, "no man ever steps in the same river twice, for it is not the same river, and he is not the same man" (Kirk, 1954). The idea here is that the second time when the man goes to the river, the water that was present the first time has since flown downstream and new water is now present; the man himself has changed as well, because his body and his mind have also undergone some transformation of sorts. Therefore, strictly speaking, it should be said that a novel man has stepped into a novel river.

What is striking, though, is that even if philosophers are already known to be quite remote from common sense, Heraclitus, and the major modern philosopher influenced by him, namely Georg W.F. Hegel, have both acquired the reputation of being particularly obscure and counterintuitive. So strong is our spontaneous tendency to identify patterns

and connect them across time, that we cannot prevent ourselves from identifying today's sun with yesterday's sun. No philosopher will be able to convince the lay man that today's sun may be a novel one: it is obviously the good old familiar sun.

But this leads to a more general observation about cognition: if we are able to survive and thrive in our environment it is because we recognize and learn patterns and we use them to predict the future and the consequences of our potential actions. Learning patterns always requires some degree of abstraction: the river I am stepping into now is not entirely the same as yesterday, but it is similar enough to let me say that it is the "same" and more concretely to let me predict the temperature and speed of the water, the presence of fish etcetera. If I travel to a new land and I encounter a different river, it will be in some sense novel, because it differs from the ones I have seen in the past, but at the same time it will look like an aggregate of properties (size, colour, speed), each of which is in some sense familiar, because it may remind me of one or another of my past experiences; therefore, I will still be able to make some predictions about this new river, but possibly fewer compared to the previous one, and with a lesser degree of certainty.

In sum, even if according to Heraclitus everything flows ("*panta rhei*"), our minds are naturally tuned to the stable regularities in the world and this allows us to make predictions and take meaningful actions by two means: firstly, we identify many things that we encounter as being "the same" as those we have already encountered in the past; secondly, we evaluate novel things, that differ from those that we have already encountered, by recognising partial aspects of them that may be analogous to one or another familiar category. It's by this double ability that we can both thrive in our familiar world and successfully explore new worlds.

1.1.2 Birth of the scientific understanding of decision-making

For many centuries, science was admirably applied to the study of most aspects of the natural world, but not human behaviour. Renée Descartes played a major role in the foundation of modern science, including physiology, but at the same time radically limited it by stating the widely influential assumption that matter and mind are two essentially different substances, and while all matter is subject to the laws of physics, the mind is independent from them (Descartes, 1641).

It was not until the mid-19th century that the study of human societies was initiated by philosophers such as Auguste Comte (1853) and Karl Marx (1867), and human psychology was first approached experimentally by physiologists like Wilhelm M. Wundt (1873-4) and Ernst H. Weber (1851), and theoretically by the philosopher William James (1890), while links between the nervous system and behaviour were explored by physicians like Pierre Paul Broca (1861) and Carl Wernicke (1874) and the physicist and physiologist Hermann von Helmoltz (Meulders, 2010). Still, the biggest questions related to the will, why we do what we do, and how decision-making works, could not be investigated because of both technological and conceptual limitations (Mandler, 2011; Benjamin, 2007; Hergenhahn, Henley, 2013).

An early insight into the link between the brain and decision-making came from the case of Phineas Gage (1823 -1860), an American railroad worker whose frontal lobe was damaged by an iron rod following an explosion. Having survived the accident, Gage did not display any straightforward sensory or motor impairment, nor any linguistic or mnemonic deficit; at first sight, he seemed to have fully recovered from the brain damage, however, his character was changed. He could not stick to plans, his behaviour was disrespectful towards others, he lost multiple jobs and could never marry, suggesting that "higher-level" cognitive abilities, previously thought to depend exclusively on an

immaterial soul, were in fact bound to a specific part of the brain (Macmillan, 2002; Damasio *et al.* 1994).

Notwithstanding the spread of a materialistic worldview from the mid-19th century, for many decades even prominent scientists, such as, for example, Charles Sherrington, would still reject the possibility that physiology may investigate phenomena like thoughts and feelings (Smith 2014). In the early and mid 20th century behaviourism sponsored a scientific approach to behaviour that would not rely on any mentalistic concept (Watson, 1913; Skinner, 1938, 1974), but the cognitive turn in psychology (Miller, 1956; Chomsky, 1959) brought the mind back under the spotlight, and the appearance of the computer provided a powerful metaphor: the mind is a process performing information computations like a software, implemented in the nervous system as the hardware (McCulloch, Pitts, 1943).

But still, research traditions in psychology and neuroscience, in particular with respect to decision-making, were largely ignoring each other. It was only towards the end of the 20th century that cognitive neuroscience was born as a recognised discipline (Gazzaniga, 1984); this was partly driven by the scientific bond allowed by the development of functional neuroimaging in healthy humans (Belliveau *et al.* 1991), but more generally it was the effect of a coordinated effort to investigate the neural bases of cognition using a variety of neuroscientific techniques in both humans and animals, jointly with sophisticated psychological theories and experimental paradigms.

Within cognitive neuroscience, the subfield of decision-making presents additional specificities and challenges, as well as promises. Because of the particularly elusive and subjective nature of decision-making, no single approach can provide a definitive answer, but a large range of disciplines have some relevance for the topic, and the synthesis of different approaches may hold the key to success. Mathematical psychology provided

numerous models of decision-making, but to help discriminating among them, neural data may be necessary beyond behavioural data. Some researchers identify the birth of the cognitive neuroscience of decision-making with the foundation of the field, and the society, of neuroeconomics (Glimcher, 2004); their key insight is that economists have already delved with the conceptual intricacies of subjective preferences and choices since the 19th century and they have eventually developed normative tools that can now be used to constrain the interpretation of biological and behavioural data. Other researchers argue that the rationalistic axiomatic approach taken in economics may have merits within its field, but is too abstract to account for the actual functioning of an organism (DeMartino *et al.* 2006; Gigerenzer, Gaissmaier, 2011; Santos, Platt, 2014). Brains are organs shaped by evolution under a large number of constraints, and disciplines like behavioural economics (Kahneman, Tversky, 1979) and behavioural ecology (Dukas, 1998; Krebs, Davies, 2009) are more relevant for its understanding. Finally, the astounding complexity of the brain, not only because of the sheer number of neurons and synapses, but also because of their variety and rate of change, means that empirical investigations will always stop far from providing a complete and detailed mechanistic picture; but computational neuroscience and machine learning can contribute to our understanding of empirical data by simulating how large networks behave and how they may self-organise in non-trivial ways to perform a given task (Rolls, 2007; O'Reilly, Franck, 2006; Yamins, DiCarlo, 2016).

The challenge of the cognitive neuroscience of decision-making for the past few decades has been, and is still, to bring together the knowledge, the practical expertise, the conceptual tools and the mindsets of all these disciplines into a single coordinated effort; the promise is that this enterprise will reveal more and more of the secrets of the human mind.

1.1.3 The rationale of the present experimental work

This introduction is organised as follows: part 1.2 summarises behavioural research on decision-making, with reference to behavioural economics and behavioural ecology, and with a particular focus on non-human primates and decision-making under risk, which are central features of the present doctoral work. Next, part 1.3 discusses the main theories and findings in the field of cognitive neuroscience of decision-making, in particular with respect to value-guided decision-making. The stages of valuation and comparison are described, classic computational models are presented, including dynamical models of decisions. The anatomical subdivisions of the prefrontal cortex are illustrated and empirical results are described with a particular attention to distinguishing the role of different brain areas in the overall process of decision-making. The distinction between actions that are habitual or goal-directed is also addressed. Subsequently, part 1.4 discusses one of the most commonly used methods in cognitive neuroscience, which is at the core of the present investigation: functional magnetic resonance imaging (fMRI). Finally, part 1.5 introduces the research we conducted to uncover the neurocognitive mechanisms of novel and familiar choices in the macaque. Three experiments were run with a similar behavioural paradigm; the first one employed fMRI with a binary choice task, the second one employed fMRI with a single option task, the third experiment assessed the impact of neurostimulation on behaviour.

As this introduction aims to illustrate, the cognitive neuroscience of value-guided decision-making is a very multi-faceted field, with a rich and complex history; its fleeting object was deemed for a long time to be outside the reach of science, and it is not surprising that its systematic study requires the interaction of a large number of methods and theoretical insights. However, even today individual experimental approaches remain partial, either because of intrinsic limitations, or because historical idiosyncrasies of

particular sub-fields limit their ability to provide a fully integrated picture of the mechanisms of decision making. Here, we identify three dichotomies and suggest that bridging them may prove instrumental for major advances in the field: with the present research we intend to provide an example of such an integrative approach to the study of the brain.

Firstly, the study of animal and human decision-making has been traditionally largely independent, because of differences in ideas about animal and human cognition, because of different neuroscientific techniques, different behavioural paradigms, different expertise and traditions. Nevertheless, results obtained in animals have had a strong impact on our understanding of human cognition (Miller, 1985; Nestler, Hyman, 2010). It is therefore of primary importance to further investigate the similarities in neurocognitive mechanisms across species.

Secondly, it is essential to better assess the impact of the different techniques and paradigms on the results observed. For example, whilst electrophysiology is almost exclusively limited to animal research, fMRI is mostly limited to human research. Obvious further differences follow from initial differences in signal measurements and analysis techniques. What is perhaps harder to discern, however, is that differences in task structure, task instruction or training, and subjects' motivational states may also impact the results in subtle but essential ways.

Finally, a central theme of this thesis is the level of familiarity with a given type of decision. Both humans and other animals engage in their natural environments via well-rehearsed familiar patterns of decision-making, as well as through on-the-fly, first-time choices, when facing new situations. However, human and monkey experiments are designed in such a way that the former almost always require participants to perform novel choices whereas the latter are based on very familiar, long-trained decisions.

So far, differing traditions and technical difficulties have not allowed to distinguish how these three factors – species, measurement techniques, and task familiarity – contribute to the results reported in the scientific literature. We believe that the present work may decisively contribute to taking apart some of these factors and in such a way to shed new light on the neural mechanisms of value-guided decision-making.

1.2 The behavioural study of decision-making

1.2.1 Economic and ecologic approaches to decision-making

A Rhesus macaque is foraging in the jungle in Vietnam. During a longer than usual trip, in an unexplored territory, it suddenly notices a tree full of strange, brightly-coloured fruits. The fruits look very attractive, but the monkey is cautious because they're unfamiliar and they may well be poisonous. If they turn out to be good, this tree would be a very rich source of food indeed; but is it worth taking the risk?

Humans and other animals encounter such problems on a daily basis. The outcomes of our actions are never fully predictable, yet we must act. This requires to estimate possible outcomes and their relative likelihood, and take decisions based on these expectations. Sometimes choices can be framed as being among a limited number of available options, such as between dishes at a restaurant, in other cases the choice is between taking an opportunity or going on, as for the monkey in the example above. Cognitive neuroscientists have generally focused on binary choices either from a perceptual or an economical point of view (Summerfield, Tsetsos, 2012) but there is a growing interest in foraging decisions (Hayden, Pearson, Platt, 2011; Rushworth *et al.* 2012; Kolling *et al.* 2012, 2014, 2016).

Historically, human and animal decision-making were largely independent fields. Human decision-making was studied by economists who focused on rationality and by cognitive psychologists who focused on departures from it (Simon, 1986; Plous, 1993; Ariely, 2008). Animal decision-making was the realm of ethology, where choices are read in terms of contribution to fitness (Krebs, Davies, 2009).

Biological data are often very rich but hard to interpret without a theoretical framework and the relatively recent field of the cognitive neuroscience of decision-

making has relied on models from the abovementioned disciplines (Glimcher, 2004), as well as from engineering (Dayan, Daw, 2008), to study the relation between the brain and decision-making. A common assumption in the field is that cognitive processes in humans and animal models are homologous (de Visser *et al.* 2011; Paglieri *et al.* 2014; Pearson, Watson, Platt, 2014; Heilbronner 2017).

Rationality is a core concept in modelling decision-making (Kacelnik, 2006). In neo-classical economics, an agent is rational if her choices respect some fundamental assumptions such as consistency and transitivity; such choices can be described as maximising a class of subjective utility functions (Samuelson, 1938). However, both animals and humans have empirically been found to contradict such neoclassical assumptions (Tversky, Kahneman, 1974).

From an evolutionary perspective, the normative prediction is that decision-making should maximise the agent's fitness (Krebs, Davies, 2009). Embodied rationality must respect a number of constraints: first the complexity of the computations required to make an optimal choice is traded off with the energy and size requirements of a brain able to perform them. Secondly, in real life the speed of the choice is often as important as its accuracy, but the brain processing speed is limited, so compromises must be made. Thirdly, in the relatively stable natural environment in which a species lives, cognitive biases often help, instead of undermine, decision-making; more generally, given limited computational power, informative priors acquired over phylogenetic and ontogenetic experience help in taking fast and accurate decisions. However, this reliance on priors backfires in artificial conditions, such as the ones tested in the laboratory (Pearson *et al.* 2014).

Since the birth of cognitive neuroscience, a direct comparison of human and animal decision-making has become more relevant. Questions derived from psychological

science can be applied across species: are human biases inherited from our non-human ancestors, or do they derive from the artificial conditions of modern society? Are biases variable across species, or are they consistent? Are they influenced by lifestyle? Moreover, questions arising from ethology can be applied to human decision making: if choices are not rational in the neoclassical sense, can they be described as rational by taking into account cognitive limitations (Simon, 1972), spatial and temporal limitations, environment (Gigerenzer, Todd, 1999) and fitness maximisation in pre-historical times?

All primates stand out in the animal world for their adaptability to very different environments, which is, in turn, a consequence of their cognitive flexibility (Tomasello, Call, 1997), and also by their ability to extrapolate information from past experience to novel situations. Non-human primates, therefore, offer a particularly well-suited opportunity not only to use neuroscientific techniques not available in humans, but also to address questions about our own cognitive evolution.

The experimental study of decision-making has often focused on a few essential factors affecting choice: those compatible with the rational economical approach, namely preferences among different goods (oranges vs apples), quantity of each item, probability of reward and uncertainty about it, outcome delay, effort required. And non-classical factors such as context-dependency, order of presentation, irrelevant alternatives, attentional biases, frame of reference (Trepel, Fox, Poldrack, 2005; Glimcher, 2014; Rangel, Camerer, Montague, 2008).

In the next two sections, we will focus on risky decisions. A situation where the choice outcomes are not known but the probability of each outcome is known, is called decision-making under risk; this differs from a situation of uncertainty where the probabilities themselves are unknown (Platt, Huettel, 2008).

1.2.2 Decision-making under risk and behavioural biases: theory

Risk is the most ancient field in the study of decision-making (Hald, 2003); the 17-th century French philosopher and mathematician Blaise Pascal introduced the concept of expected value (defined as the average of the value of each possible outcome weighted by their relative probability), as a quantity that should be maximised in order to maximise long-term reward (Pascal, 1670). However, this was soon found to be unsatisfactory, as highlighted by the St. Petersburg paradox, where a bet with an infinite expected value was in practice comparatively unattractive for gamblers (Shafer, 2004). Bernouilli suggested that utility may be a logarithmic function of value; while the shape of the function was arbitrary, it recognized that the marginal utility of an extra unit of value decreases as the total wealth increases (Bernoulli, 1738). As mentioned above, neoclassical economics assumes that agents maximise expected utility, where the utility function is any of a class of increasing monotonic functions (Samuelson, 1938), and in the case of risky decisions, the utility of each possible outcome is multiplied by its probability (Morgenstern, von Neumann, 1953).

A non-linear utility function is compatible with the classical axioms of rationality, but many other empirically observed effects are not. Systematising previously observed biases (Allais, 1953; Ellsberg, 1961) while adding many novel experimental observations, Kahneman and Tversky (1979) built a non-normative framework to understand decision-making. In this framework, the reference of value is not absolute zero, but a prospect that depends on the agent and which can even be affected by the wording of the choice. The function mapping objective reward magnitude onto subjective reward magnitude can have a "kink" at the reference point (the *prospect*), and its derivative may have different signs above and below it. For example, humans are often found to be risk-averse in the

domain of gains (concave function) and risk-seeking in the domain of losses (convex function) as well as loss-averse (steeper slope for losses than for gains).

Additionally, Kahneman and Tversky described a function mapping objective reward probability onto subjective reward probability. It involves a weighting of probability such that small probabilities are perceived as larger than they actually are, and large probabilities as smaller (inverse-S shape). Figure 1.1 illustrates a subjective utility function and a probability weighting function, both fitted to typical human values from Tversky and Kahneman (1992).

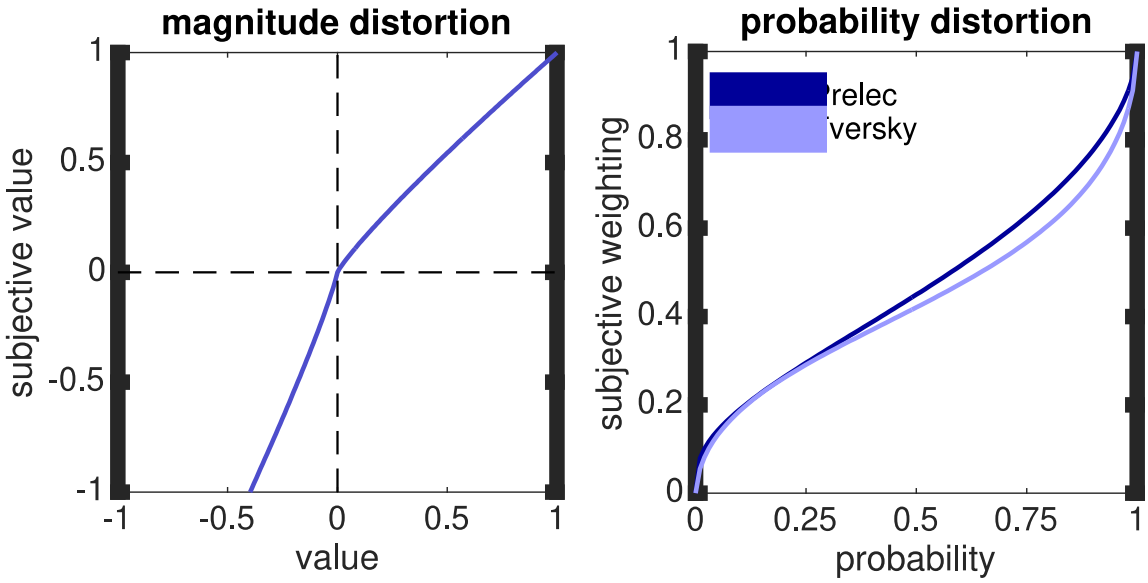


Figure 1.1 Illustration of magnitude **(left)** and probability **(right)** distortions in prospect theory, using typical human values of magnitude distortion = 0.88, probability distortion = 0.61 and loss aversion = 2.25. The dark blue line in the right panel ("Prelec") is an alternative function for subjective weighting (see chapter 2, section 2.2.6).

While it is often reported that humans are risk-averse for gains, we do also buy lottery tickets. Kahneman and Tversky described more precisely a *fourfold pattern of risk attitudes*: risk-averse behaviour when gains have moderate probabilities or losses have

small probabilities; risk-seeking behaviour when losses have moderate probabilities or gains have small probabilities (Kahneman, Tversky, 1979; Rabin, Thaler, 2001).

Notwithstanding the broadness of prospect theory, there are other important biases that go beyond its scope. Attentional effects (Hare, Malmaud and Rangel, 2011) may induce an overweighting of the attended option. Magnitude and probability may be summed, instead of being multiplied, as a shortcut to ease value computations (see chapter 2, section 2.2.6, chapter 5, section 5.4, and Scholl *et al.* 2014; Scholl, Kolling *et al.* 2015).

1.2.3 Decision-making under risk and behavioural biases: experiments in primates

Numerous studies have addressed decision-making under risk in primates: most researches have focused on the question whether monkeys show the same kind of economic behaviour as humans. In particular, humans are generally risk-averse, even if this depends on the context, and multiple studies have looked into animal risk-sensitivity. A meta-analysis by Kacelnik and Bateson (1996) across 59 animal studies in different species and contexts showed that many animals are risk-averse when the amount of reward is variable, although some are risk-neutral or even risk-seeking.

In particular, Yamada *et al.* (2013) argued that macaque monkeys after training are a good model of human risky decision-making because their choices are transitive and mildly risk-averse. Many other studies, however, have found macaques to be risk-seeking. Heilbronner and Hayden (2013) suggested that this is due to the specific design of most macaque experiments, conceived for electrophysiological recordings, that involve very small stakes, a large number of trials per day at a fast pace, and task parameters learned by experience.

Smith, Beran and Young (2017) found that macaques choose the risky option more often if the outcome of the lottery is signalled earlier, even if the time until reward is kept constant.

Another question often addressed relates to the cause of risk preferences. Hayden and Platt (2007) suggest risk sensitivity in macaques is a consequence of temporal discounting in repeated-choice tasks, and they derive this behavioural prediction under the assumption that the monkeys focus their attention to the time left until the next big reward. However, they acknowledge that other models may explain the same data. In any case, they argue (Hayden *et al.* 2008), the convexity of the utility function alone cannot explain monkeys' risk-seeking behaviour, because they prefer stochastic rather than deterministic lotteries with same expected value and are more sensitive to variations of the big reward than of the small reward.

The question of the shape of the probability distortion has been investigated by other groups as well. It's been reported that monkeys, like humans, overweight low probabilities and underweight high probabilities when they are described (Stauffer *et al.* 2015), which may lead to risk-averse behaviour[1], but follow the opposite pattern for probabilities learned by experience (McCoy, Platt, 2005). Additionally, even for gambles with equally probable outcomes (50%-50%), where no distortions are expected, both species are more risk-seeking for experienced than described gambles. (Barron, Erev, 2003; Ludvig, Spetch, 2011; Heilbronner, Hayden, 2016). This is called the description-experience gap (Camilleri Newell, 2013).

[1] This happens in common designs but is not necessary. If the choice is between a safe (100%) gain of £1 or a risky (20%) bet to win £5, if someone over-weights small probabilities, she will assign a higher value to the risky option. If the choice was between a sure gain of £4 or a 80% chance of winning £5, the same subjective distortion, down-weighting 80%, would lead to a preference for the safe option.

Farashahi and colleagues (2018) compared macaque's economic biases in two tasks, one of them including both gains and losses thanks to a token system. They found that the monkeys were (1) risk-seeking for both gains and losses, (2) more so for losses, and (3) overweighted large probabilities and underweighted small ones (these behaviours are all contrary to what is most commonly observed in humans). In the same study, they also report that a model including a differential weighting for the most salient stimulus in terms of magnitude accounted better for the data and reduced the other distortions.

Economic biases have been studied in primate species other than macaques, using more ecological designs. Chen, Lakshminarayanan and Santos (2006) supported the idea that deviations from the economic axioms are evolutionarily old showing that capuchins are subject to similar biases as humans when trading in a token market.

It seems now accepted that the behavioural ecology of a species may affect its risk-preferences, even for animals bred in captivity. Chimpanzees, compared to bonobos, have a riskier lifestyle, because the fruits they eat are less constantly available across the year, and they complement this diet with hunting. When tested in a binary lottery task, chimpanzees where found to take more risks compared to bonobos (Heilbronner *et al.* 2008). Lemurs, who tend to live in a harsh environment, are also risk-averse (MacLean, Mandalaywala, Brannon, 2012), while orang-utans, whose living habits are similar to chimpanzees, are risk-seeking (Haun, Nawroth, Call, 2011).

Overall, it is clear that human behaviour, both in terms of risk-sensitivity, and in terms of departures from economic rationality, is, in some ways, consistent with the range observed in other primate species, and is similarly affected by context.

Why then is it the case that such cognitively sophisticated animals still display numerous biases that in abstract economic scenarios do not maximise long-term reward? This is definitely an open question, but researchers have started to investigate the

cognitive constraints that may have led evolution to select these types of solutions. Tsetsos, Chater and Usher (2012) used a task in which rapidly presented information had to be integrated across time, and found that a mechanism weighted by salience accounts for the data and explains at the same time numerous well-known choice biases. Even in the wild, salience may be a useful shortcut in situations where the decision time is limited. Further research suggests that economically irrational behaviour is an adaptive consequence of noisy neural computations (Tsetsos *et al.* 2016; Wyart, Koechlin, 2016).

In conclusion, numerous studies have investigated the presence of some specific features of economical behaviour in primates, such as risk-preference and probability weighting, and how they are affected by context. Other features, such as whether the subjective value of risky options is composed in a multiplicative or additive way, have seldom been discussed. The length of the training, impacting the degree of familiarity with the task, was investigated in relation to the description-experience gap (Camilleri, Newell, 2013), but only as far as it affected risk sensitivity, not other aspects of decision-making. We will have the opportunity to shed further light on these questions thanks to the behavioural results of experiment 1, described in chapter 2.

1.3 The cognitive neuroscience of decision-making

1.3.1 The structure of a decision: conceptual and computational aspects

Contrary to economists and ethologists, a cognitive neuroscientist is interested in understanding how the decision process unfolds through time in the brain. Even though science requires simplified models of reality, cognitive neuroscience models of decision-making are generally more articulated than those used in the previously mentioned disciplines, reflecting a stronger interest for cognitive mechanisms and their components.

This section will focus on the principal conceptual and computational contributions to the field of decision-making; section 1.3.2 will present an overview of the anatomy of the prefrontal cortex; section 1.3.3 will discuss some of the major empirical results in the cognitive neuroscience of decision-making; finally, section 1.3.4 will focus on different action control systems.

The field of decision-making is traditionally divided in perceptual and value-guided decision-making (Summerfield, Tsetsos, 2012). Both types of choice are meant to be driven by a decision variable (DV) representing the amount of evidence in favour of one of the available options. However, whilst in perceptual choice the DV depends on perceptual information (Gold, Shadlen, 2007), in value-guided choices the DV is supposed to represent the subjective preference for one of the possible outcomes and can be described as the value difference between the favoured option and the other one(s) (Rangel, Hare 2010; Rushworth, 2011). In this introduction we will mainly focus on value-guided choice.

It has been suggested that basic computations involved in making a value-guided choice are five: representation, valuation, action selection, outcome evaluation and learning (figure 1.2, Rangel *et al.* 2008). The representation stage consists in the

identification of the available possibilities; this step received little attention (Kaiser *et al.* 2013), but has nevertheless a remarkable importance because in real life situations, the range of possible actions is essentially infinite, although an individual only focuses on a few at any given time. It was suggested that dopamine modulates option generation for behaviour (Ang *et al.* 2018). For example, apathy in Parkinson disease may be connected to an impairment in option generation (Sinha *et al.* 2013) whereas an aspect of obsessive-compulsive disorder and addiction is the repeated and intrusive representation of potential actions that would not even be considered by healthy people in the same circumstances.

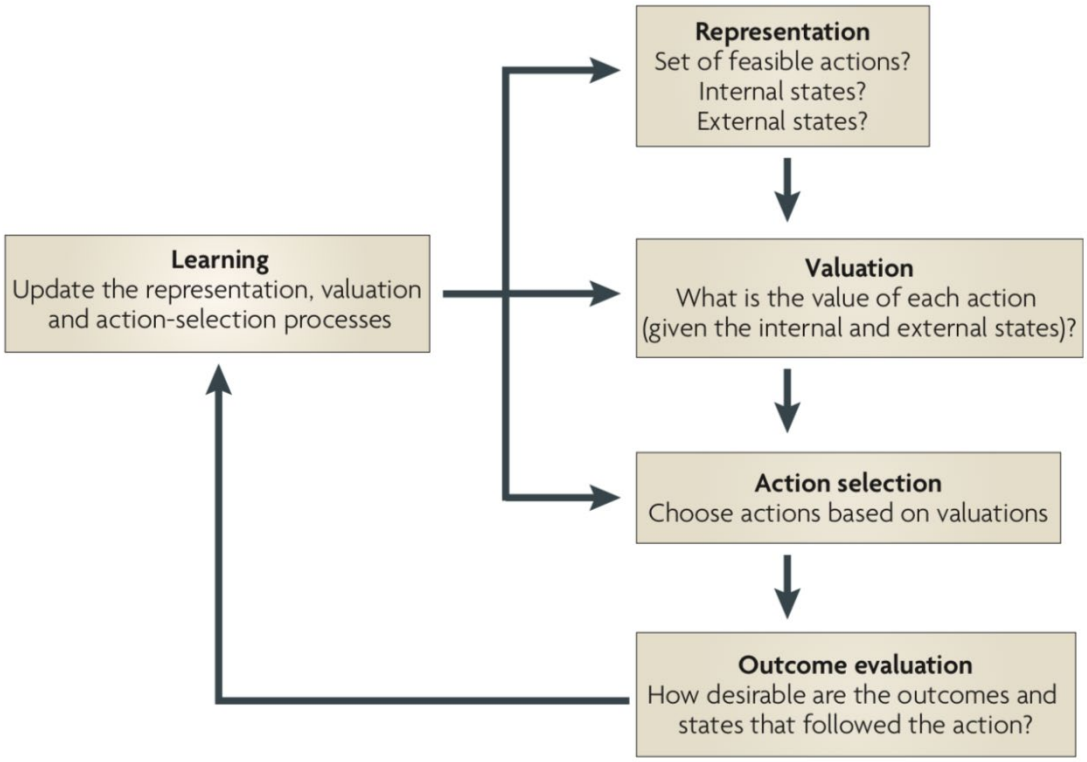


Figure 1.2 The five major step in the value-based learning and decision-making process. (reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, *Nature Reviews Neuroscience*, A framework for studying the neurobiology of value-based decision making, Rangel, A., Camerer, C., & Montague, P. R. (2008)).

The key steps at the time of a decision are valuation and action selection; contrasting with the other two steps happening *after* the action performance (outcome evaluation and learning). Valuation is the process of estimating the value of each possibility under consideration; action selection is the process of choosing the preferred option given the value of each possibility. Although conceptually these steps are distinct, some suggest that they may happen simultaneously and dynamically in the brain (Thura, Cisek, 2014; Hunt, Hayden, 2017), which means that the action selection process would start while evaluation is still ongoing and would be influenced by the fluctuations in the latter.

The theoretical concept of value is remarkably abstract: it is a single, quantitative attribute that can describe any possible outcome, therefore allowing comparison even when their other features are heterogeneous. Adult humans are familiar with this concept, being exposed to money on a daily basis; monetary value is a single continuous scale that can be used to quantitatively compare very different kinds of things like objects, services, workforce, debts. The intuitive concept of value is often tied to the one of pleasure, but many other goals influence value estimates, such as pain and fatigue avoidance, social acceptance and power.

The number of properties that affect the value of an outcome is very large. First, outcomes may be items or events or absence of these; they may follow the selection of a stimulus (choosing a dish in a restaurant menu) or they may follow actions in the absence of discriminating stimuli (driving to see your parents). Important outcome attributes affecting value are the probability distribution of possible outcomes, the temporal delay, and the amount of effort required. In humans, there may be monetary costs associated to outcomes. Outcomes themselves most often have multiple aspects: a car's value may be characterised by its brand, its shape and colour, its speed and acceleration, its fuel

consumption. Each of these aspects may have independent positive or negative value, and they all need to be integrated.

The abstract computational account of decision-making requires all these aspects of an option to be integrated into a single quantity, its subjective value, so to allow comparison and action selection. It is therefore remarkable that the neuroeconomical approach has found evidence in the brain, including in non-human animals, of a value signal respecting such economic requirements as independence from alternative options (Padoa-Schioppa, Assad, 2008). This has been an old dream of 19th century economists, who wished to be able to measure the subjective utility afforded by any possible good with an ideal *hedonimeter*. However, these ambitions were abandoned by the beginning of the 20th century, when a neopositivistic turn, akin to behaviourism in psychology, restricted economics to be the science of actual choices, not subjective value (Niehans, 1990). Still today, microeconomics models are "as if" models: agents' behaviour is described as if their empirical choices were maximising a subjective utility function (Kreps, 1990). Cognitive neuroscience, instead, aims to actually uncover what representations are encoded in the brain and drive our choices.

Other nuances need to receive consideration. Assuming the brain can keep track of the different aspects of value, an important theoretical question is whether a final common integration process really constructs a single value representation for each available option, that is derived from its component parts, and then use it to guide action selection. An alternative hypothesis is that the component aspects are encoded separately and they influence action selection directly rather than via a unitary value representation.

Secondly, whilst the idea that abstract, subjective value signals are measurable in prefrontal cortex is no longer controversial, different views are held about how the

mechanism works at the algorithmic level[2]. Value representations may be explicitly encoded in the brain and used to drive action selection (Kable, Glimcher, 2009; Padoa-Schioppa, 2011), or alternatively the signal measured may be a by-product of a process that is influenced by value components but does not encode value per se (Hunt, Hayden, 2017).

The action selection stage is less straightforward than what might be thought at first sight. In rational economical accounts, a subject should always choose the best available option; however, in psychology it is well established that subjects generally pick each available option with some frequency as a function of their value (Herrnstein, 2000); this phenomenon can take different forms depending on the reinforcement schedules used and may or may not be optimal. A common form of computational accounts of value-based decision-making derived from machine-learning (Kaebling, 1993) is that the likelihood of choosing any option is a softmax function of its relative value. For options 1 to N, having value V_1 to V_N, the probability p(i) of choosing option i will be:

$$p(i) = \frac{e^{(\alpha * V_i)}}{\sum_1^N e^{(\alpha * V_n)}}$$

where α modulates the precision with which the most valuable option is chosen, also called inverse temperature.

The fact that the best option is not chosen deterministically is partly explained by neural noise, however other reasons have been suggested. The most important is the necessity to explore (Thrun, 1992): because new options may appear with time, and

[2] David Marr famously distinguished three levels in cognitive science, according to the metaphor that the mind is like a computer: the computational level is concerned with the goals of a process and the broad lines of a computation; the algorithmic level is concerned with how this computation is performed in terms of specific information processing and the implementational level is concerned with how the computation is actually carried out by neurons (Marr 1982).

because the value of previously sampled options may change, it is essential to keep at least an approximate running estimate of the value of all available options; in order to do this, it is often necessary to choose suboptimal options (options that we expect, on average, to be less valuable than our favourite one) in order to re-assess their value from time to time (Cohen *et al.* 2007).

Following the decision, it is assumed that the subject assesses the outcome and updates the internal representations and the value estimates. The most widely used model of learning in neuroscience is by simple reinforcement (Rescorla, Wagner, 1972). The model assumes that each option is associated with a numerical internal representation of value, corresponding to the average expected reward for that option; each time the option is chosen, if the outcome does not match the prediction, the latter is updated as a function of two variables: a learning parameter and the reward prediction error, which is the difference between the actual reward and the expectation. This simple model does not assume any knowledge of the structure of the environment and it updates its estimation on a trial-by-trial basis. The learning parameter determines whether the estimate is based on the long-term experience or only on the most recent outcomes. A low learning rate is useful if the outcomes change probabilistically from trial to trial but the underlying statistics are stable. A fast learning rate is preferable in fast-changing environments. It has also been suggested that the environmental volatility itself should be tracked, because if the environment is relatively stable, there is no need to explore too much, and one can focus on exploiting the best option, but in a volatile environment the utility of exploration is larger (Behrens *et al.* 2007). The other variable that determines the size of the update following the feedback is the reward prediction error; if this is zero (the prediction was correct) no update is necessary, but the larger is the error, positive or negative, the larger the estimate update needs to be.

A temporally continuous version of this model is called temporal-difference reinforcement learning (Sutton, Barto, 1998) and instead of updating estimates only at the end of each trial, it does it at each time step. This allows for a continuous modelling of behaviour, which is especially useful in situations where multiple decisions need to be taken before any reward is obtained (as in a maze, or at a poker game). A more sophisticated approach to value-guided decision-making is represented by so-called model-based reinforcement learning, which implies some knowledge of the structure of the environment and the state of the agent within it, in order to enable efficient learning also in nonstationary environments with hidden states (Tolman, 1948; Doya *et al.* 2002).

Many theoretical accounts of decision-making, as we have described so far, assume that some quantities, namely the values associated with each option, are computed and then a selection is made based on these quantities. Whilst very popular, these approaches remain at the computational level and do not, to use Marr's terminology, address the algorithmic level (see footnote 2). On the other hand, algorithmic accounts tend to have different characteristics compared to the previous family of models. First, mechanistic models explicitly simulate the temporal dynamics of decision making and can better account for both the speed and the accuracy of decisions. Second, these models, while still abstract, are closer to the implementational level and are therefore more biologically plausible. Third, these models are more easily comparable to empirical data at the micro- (single-cell) and meso- (e.g. local field potential) scale, while offering a potential mechanistic explanation for signal measured at the macro-scale (e.g. fMRI) (Wang, 2002; Ratcliff, McKoon 2008; Hunt *et al.* 2015).

A very well established family of models in mathematical psychology are drift-diffusion models (Ratcliff, McKoon, 2008), used in both perceptual and value-guided decision-making (Gold, Shadlen, 2001; Krajbich *et al.* 2010). They assume that a random

variable, representing evidence in favour of either of two alternative options, drifts with an average determined by the objective difference in evidence; this will be either perceptual evidence or value difference. The decision is assumed to be triggered when the variable reaches either of two thresholds, with the starting point in between them. The drift tends on average to the best option, but is noisy, and therefore can make erroneous choices depending on the amount of noise and the distance between the thresholds. Overall, the model can make predictions about both reaction times and accuracy, it allows to formalise their trade-off, and it provides a good match with a large range of empirical choice datasets (Bogacz *et al.* 2006). This model can also be used to make neural predictions (Turner *et al.* 2015).

Here, we will focus on one model in particular that has been described as a biophysically realistic cortical network model for binary discrimination (Wang, 2002). It is characterised by three assemblies of neurons, two populations of excitatory neurons, each representing the evidence in favour of one possible choice, and a population of inhibitory neurons. The two choice populations receive an input proportional to the evidence in favour of the option they represent. Their recurrent excitation is mediated by NMDA receptors with a slow time integration, while their reciprocal inhibition is mediated by the third population of neurons.

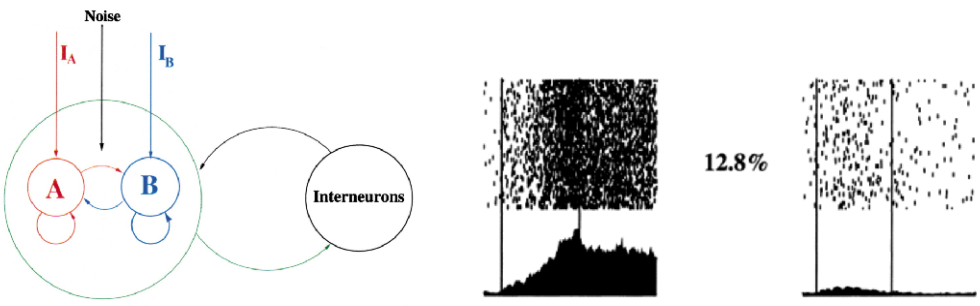


Figure 1.3 Attractor network model by Xiao-jing Wang (2002), with permission by Elsevier. **Left:** network structure. A, B: populations of excitatory neurons driving the

choice towards options A and B respectively. IA and IB represent the size of the respective input. Interneurons are inhibitory. **Right:** the network can simulate the firing rate of individual neurons during a perceptual task. In the example, the stimulus is a random dot kinematogram with a 12.8% coherence towards the left. The raster plots represent spikes of simulated neurons coding respectively the left and right options.

This model has been successfully extended to simulate, among other tasks, value guided (rather than perceptual) decision-making (Hunt *et al.* 2012) and three-option decisions (Chau *et al.* 2014).

After a brief description of the anatomy of the prefrontal cortex, we will discuss how the accounts described in this section can help explain the neural activity measured in the primate brain and what may be the contributions of the different regions of prefrontal cortex to value-guided decision-making.

1.3.2 An overview of the anatomy of prefrontal cortex

The prefrontal cortex is the most anterior part of the frontal lobe. It includes most of the frontal lobe excluding the motor, premotor and supplementary motor areas (areas 4 and 6). Cytoarchitectonically (in terms of its neural cell types and their distribution) it is distinguished by the presence of granule cells in its fourth layer (layer IV). Phylogenetically it is most prominent, or perhaps even only present, in primates (Wise, 2008; Passingham, Wise, 2012).

Within frontal cortex in both human and non-human primates, the regions that have been frequently discussed for their roles in guiding behaviour are: the anterior cingulate cortex (ACC), the orbital prefrontal cortex (OFC), the anterior medial frontal cortex (AMFC), the dorsolateral prefrontal cortex (dlPFC), and the frontal pole.

The frontal area perhaps least implicated in value-guided behaviour is the ventral part of the lateral prefrontal cortex (vlPFC); therefore, it will not be discussed below. There is evidence that vlPFC enables the association of arbitrary cues and motor responses (Wise, Murray, 2005) and the processing of sequences with complex rules (Wilson *et al.* 2015) in both humans and other primates. Because in humans part of the left vlPFC is specialised for language production (Broca, 1861; Binder *et al.* 1997), it has been suggested that sequence processing is an evolutionary preserved function underlying the development of linguistic skills in humans (Wilson *et al.* 2017, Kikuchi *et al.* 2018).

In this short overview, we will describe the key prefrontal areas, their subdivisions, their connectivity and their global functional specialisation. Importantly, we will try to make some order in the occasionally inconsistent usage of labels in previous publications. Because non-human primates, especially the species *Macaca mulatta*, have been extensively used in neuroscientific research, we are interested in the comparative neuroanatomy of this species and *Homo sapiens*. It is generally recognised that the macaque is a good model for the human brain, including for some higher faculties such as social cognition, but differences are also present and must not be underestimated (Passingham, 2009). The following neuroanatomical descriptions are sufficiently general to fit both the human and macaque brain, but exceptions will be signalled.

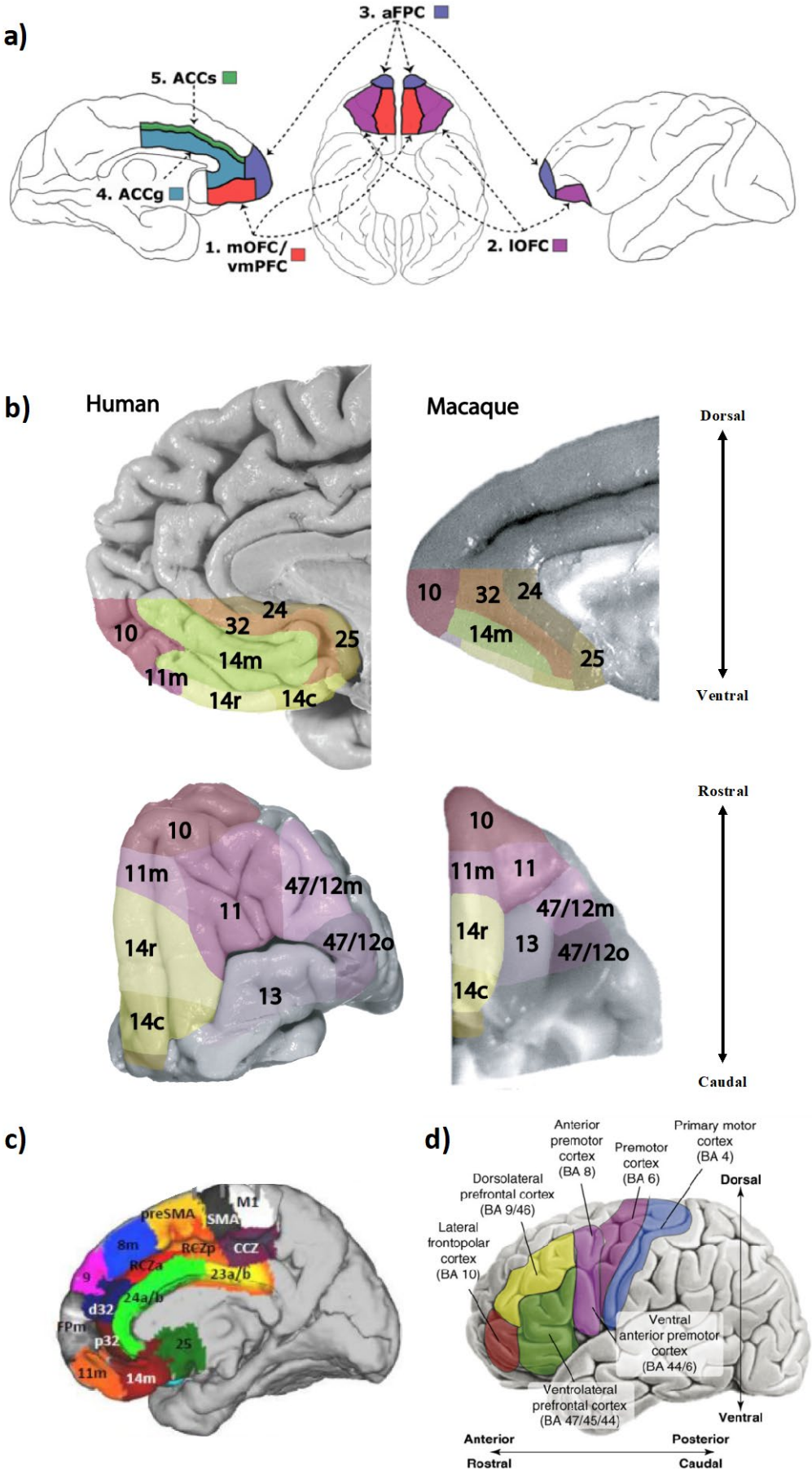

a)
3. aFPC
5. ACCs
4. ACCg
1. mOFC/
vmPFC
2. lOFC
b)
Human
Macaque
Dorsal
Ventral
10
32
24
25
14m
11m
14r
14c
11
47/12m
47/12o
13
Rostral
Caudal
c)
M1
preSMA
SMA
8m
RCZp
CCZ
RCZa
9
23a/b
24a/b
d32
FPm
p32
25
11m
14m
d)
Anterior premotor cortex (BA 8)
Premotor cortex (BA 6)
Primary motor cortex (BA 4)
Dorsolateral prefrontal cortex (BA 9/46)
Lateral frontopolar cortex (BA 10)
Ventral anterior premotor cortex (BA 44/6)
Ventrolateral prefrontal cortex (BA 47/45/44)
Dorsal
Ventral
Anterior
Rostral
Posterior
Caudal

Figure 1.4 Subdivisions of the prefrontal cortex: **a)** Overview of regions involved in value-guided decision-making in the macaque brain (from Rushworth *et al.* 2011, with permission by Elsevier) **b)** Cytoarchitechtonic parcellation of ventral and orbital PFC and comparison between human and macaque cortex (from Mackey, Petrides, 2010, with permission by Wiley). Top: medial view. Bottom: ventral view **c)** Connectivity-based parcellation of human medial frontal cortex (from Neubert *et al.* 2015) **d)** Regions of the lateral surface of the human prefrontal cortex (from Badre, 2008, with permission by Elsevier).

The ACC is composed of areas 24, 25, 32. Based on its gross anatomy, it can be coarsely distinguished into subdivisions in the sulcus and the gyrus of the cingulate cortex (ACCs and ACCg). One of the most studied subdivisions, the dorsal anterior cingulate cortex (dACC), lies in ACCs. More anteriorly and ventrally can be found perigenual ACC (pgACC) anterior to the genu of the corpus callosum and subgenual ACC (sgACC), area 25, in the subcallosal gyrus. It can be parcellated based on functional connectivity consistently across subjects (Beckmann *et al.* 2009). ACC is mainly connected to cortical motor, premotor and parietal areas, but also deep structures like the amygdala, hippocampus and the basal ganglia. This connectivity profile is largely maintained across human and non-human primates (Neubert *et al.* 2015).

OFC represents the ventral surface of the frontal lobe, adjacent to the orbital bones; it includes principally areas 11, 12, 13, 14. Medially, it is sometimes taken to include only areas 14r and 14c but according to other usage it includes vmPFC (area 14m). On the lateral side, some studies, but not all, include in the definition of OFC the lateral orbital gyrus (areas 12/47). Connectivity-based parcellations of OFC reliably distinguish two clusters: medial OFC (areas 11m, 14) and lateral OFC (areas 11o, 13 and 12/47) in both humans and non-human primates (Averback, Seo, 2008; Kahnt *et al.* 2012). Saleem and colleagues (2008) have shown that the connections of the medial and lateral parts of OFC

form largely distinct circuits. Lateral OFC, often simply called OFC when referring to areas 11 and 13, receives input from higher-level sensory areas of all modalities, has reciprocal connections with the temporal cortex, amygdala, hippocampus and the basal ganglia and sends its output to the lateral prefrontal cortex but not directly to premotor areas (Carmichael, Price, 1995a,b; Miyachi *et al.* 2005; Saleem *et al.* 2014).

The medial surface of the prefrontal cortex has been described very heterogeneously. From a cytoarchitectonic point of view, it includes areas 24, 25, 32, that are part of ACC, area 10 that is part of the frontal pole, and area 14m more ventrally, which as we said, can be described as part of medial OFC. Finally, the medial surface more dorsally includes areas 8 and 9, also called the superior frontal gyrus (SFG). A large number of human imaging studies have found decision value-related activations in various locations probably clustering in the vicinity of area 32 (part of ACC), area 14m (part of mOFC), and possibly area 10 (part of the frontal pole) across the anterior medial frontal cortex. In the absence of more detailed anatomical investigation it is difficult to identify the key cytoarchitectonic area precisely and so here the region will be referred simply as anterior medial frontal cortex (AMFC). This functionally-defined region has also sometimes been referred to as MPFC. By contrast, another rather general and functionally-based term, vmPFC, has been used to refer to activity in a slightly more ventral region, however, such terms are not always employed consistently. For that reason, in this thesis vmPFC is used to indicate the most ventral part of the medial prefrontal cortex, in and around area 14m, and by contrast *anterior medial frontal cortex* (AMFC) is used to refer to the anterior parts of both area 32 and 14m, which is the region most frequently activated by studies of decision value in humans. It should be noted that the same AMFC region is also a major node of the default-mode network. On the other hand, following general

usage, we will often adopt the label *OFC* for the central part of OFC, comprising areas 11 and 13.

Another region often involved in decision-making studies is dorsolateral prefrontal cortex (dlPFC), areas 9 and 46, lying at the midsection of the superior and middle frontal gyri (Petrides, Pandya, 2012). There is a broad correspondence between the regional anatomy in humans and macaques (Sallet *et al.* 2013), but Petrides and Pandya (1999) suggested a change in terminology, introducing a third area, 9/46, that was previously given inconsistent labelling across human and non-human primates.

Finally, the frontal pole can be essentially identified with area 10, the most rostral location of the frontal lobe. It is interconnected with other parts of the frontal cortex and only sparsely with the rest of the brain, including links with the posterior cingulate cortex, insula and superior temporal sulcus and gyrus (Petrides, Pandya, 1999). Although there are many similarities between several prefrontal areas in humans and other primates, area 10 is relatively more distinct in humans and other primates (Semendeferi *et al.* 2001).

1.3.3 Distributed neural circuits for value-guided decision-making in prefrontal cortex

Orbitofrontal cortex

Orbitofrontal cortex has consistently been found to be associated with value estimation and value-guided choice (Rudebeck, Murray, 2014). Lesion studies in rodents (Rudebeck *et al.* 2006), macaques (Noonan *et al.* 2010), and humans (Fellows, Farah, 2007) have shown that lesions to OFC are linked to disruptions specific to value-guided decision-making. Additionally, electrophysiological recordings in macaques consistently found OFC neurons coding value, both option value and choice value (Padoa-Schioppa, Assad, 2006).

However, value-related signal is found in many prefrontal areas, as well as lateral intraparietal cortex (Platt, Glimcher, 1999). In fact, because subjects in neuroscientific experiments work to obtain reward, it is not surprising that many neural events are modulated by value. For example, attention is enhanced (Kiss *et al.* 2009) and motor responses are stronger and faster when high value is expected (Niv *et al.* 2007; Shadmehr *et al.* 2010) compared to low value. Moreover, depending on the design, value can be correlated with many other factors (O'Doherty, 2014); for example, parietal neurons were found to respond better to saliency than value, when the two are decorrelated (Leathers, Olson, 2012).

It is therefore essential to understand in which manner different regions encode value representations and how each contribute to value-based decision-making. OFC activity is both correlated with expected value at the time of the decision and when the outcome is received (Tremblay, Schultz, 1999). Activity integrates and encodes different properties of primary rewards, such as taste, texture, smell and temperature, consistent with the OFC connectivity profile (Kringelbach, Rolls, 2004). Moreover, the value signal depends on the subjective state of the individual, decreasing with satiety (Rolls, 2008), and, accordingly, OFC lesions specifically impair monkeys in a devaluation task (Izquierdo *et al*. 2004). It represents both appetitive and aversive information (Morrison, Salzman, 2011) and its role is not restricted to choices among stimuli, because OFC lesions also impair choices among actions in a similar design (Rhodes, Murray, 2013).

It was found that some individual OFC neurons encode value of a specific option (whether it is chosen or not) and other neurons encode the value of the chosen option (whichever it is) and they do so independently from what options are available or not at the time of the choice (Padoa-Schioppa, Assad, 2006, 2008). However, other studies

have found an adapting coding, such that some neurons responded more strongly when the offer value was most different from the previous trial (Kennerley *et al.* 2011).

OFC is necessary for credit assignment, which means to associate a given outcome to the correct stimulus or action that led to it, and the subsequent use of this association to guide behaviour (Walton *et al.* 2010). This explains the somewhat surprising observation that animals with an OFC lesion are more impaired in learning easy as opposed to difficult tasks, compared to controls (Noonan *et al.* 2010). Overall, it can be concluded that OFC represents the identity and the motivational value of expected outcomes (Rudebeck, Murray, 2014).

The experiments described so far were performed on animal subjects and focusing on central and lateral OFC. Most human lesion studies do not have the specificity to distinguish between vmPFC and OFC contributions, because in most cases lesions encompass both areas. However, Noonan and colleagues (2017) confirmed the macaque results in human patients. Some human fMRI studies have reported an effect of the subjective value of expected reward in OFC. For example, Sescousse and colleagues (2003) reported a distinction between primary rewards activating posterior OFC (evolutionarily older) and secondary rewards activating anterior OFC. Using a devaluation task, similar to the one often performed in animal experiments, Gottfried and colleagues (2003) detected an fMRI signal related to expected value in human OFC. Additionally, Klein-Flügge and colleagues (2013) found an fMRI OFC signal for food rewards. Grabenhorst and Rolls (2009) reported a context-independent effect of smell pleasantness in both mOFC and central OFC. However, the large majority of fMRI reports have identified a role of human vmPFC/AMFC in value-guided decisions and no, or smaller, effects in central and lateral orbitofrontal cortex. This may be in part linked to

the use of sensorily salient rewards in animal studies and more abstract rewards in human studies.

Anterior medial frontal cortex

Plassmann and colleagues (2007) found an effect of willingness to pay for everyday goods in AMFC, as measured by an auction procedure. They also showed that this signal is present for both appetitive and aversive goal values (Plassmann *et al.* 2010). Congruently, AMFC encodes expectations of monetary loss (Tom *et al*. 2007). However, when hungry participants chose whether or not to buy food items for various prices, AMFC activity correlated with stimuli values but not their monetary cost, while central OFC correlated with net value (subjective value minus cost) and striatum activity was proportional to prediction error (Hare *et al.* 2008). Some types of cost are, however, included in the representation of the outcome. For example, Kable and Glimcher (2007) observed an effect of temporal discounting in AMFC when subjects compared monetary outcomes associated with different delays.

Lebreton and colleagues (2009) argue that AMFC/vmPFC encoding of subjective value is automatic and therefore task-independent; however, this contrasts with the reports of Plassmann *et al.* (2007, 2010), where the signal disappeared if the subjects are given no choice. Other studies have found a choice-dependent signal in AMFC, namely the value of the chosen option minus the value of the unchosen option (Boorman *et al.* 2009).

Overall, there is a very strong evidence coming from human fMRI studies that AMFC activity during value-guided choice encodes expected value at the time of the choice, but also outcome value, both in the gain and loss domains and for food as well as for money. As discussed in the previous section, many studies refer to this region as vmPFC, although in some cases the activations are more dorsal than area 14m while others refer

to this it as medial OFC. Two large meta-analyses of human fMRI studies, reported in figure 1.5 (Clithero, Rangel 2010; Bartra 2013) summarise these results. Both find a strong and focal effect in AMFC (labelled vmPFC) and ventral striatum. Only one finds an effect in lOFC, but smaller (vSTR ALE (activation likelihood estimate) = $69*10^{-3}$, AMFC ALE = $68*10^{-3}$, left OFC ALE = $29*10^{-3}$, Clithero, Rangel, 2010).

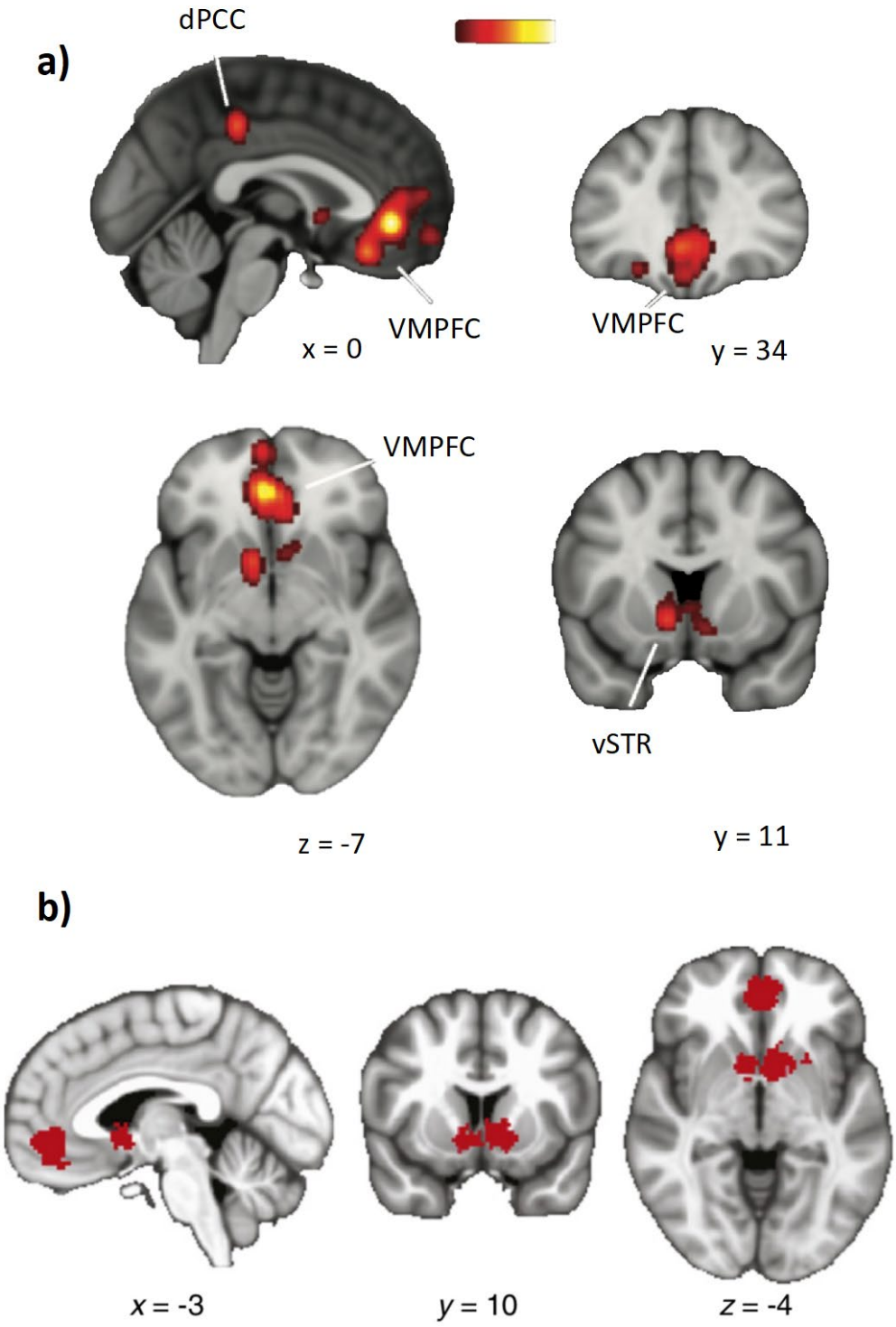


Figure 1.5 Two meta-analyses of value encoding in AMFC/vmPFC. **a)** Reliable neural features of value during the decision phase identified using coordinate-based meta-analysis (CBMA) of 69 studies. vmPFC peak at MNI coordinates [-2, 40, -4]. The colorbar spans activation likelihood estimation (ALE) values of $15x10^{-3}$ to $63x10^{-3}$ (Clithero, J. A., & Rangel, A. (2013). Informatic parcellation of the network involved in

the computation of subjective value. *Social cognitive and affective neuroscience*, *9*(9), 1289-1302, by permission of Oxford Univeristy Press). **b)** Subjective value (SV) representation identified by a five-way conjunction analysis (positive effect of SV, effect of SV for both decision and outcome, effect of SV for both monetary and primary incentives). Meta-analysis of 206 human fMRI studies. vmPFC centre-of-mass MNI coordinates [-1, 46, -7] (from Bartra *et al.* 2013, with permission from Elsevier).

It may seem confusing that most human neuroimaging studies do not find value correlates on the orbital surface of PFC (but see above for some exceptions) whilst studies in macaques, both using electrophysiology and lesions, reliably find significant effects there. Because the latter methods are focal, and macaque AMFC has seldom been studied independently from OFC, we have little information on AMFC's role in non-human primate decision-making.

Occasionally, in an attempt to reconcile this discrepancy, medial PFC and OFC have been described as a single functional region (Rangel *et al.* 2008; Padoa-Schioppa, 2011) overlooking the fact that the clusters of effects found in human and macaque studies have clearly distinct foci. This approach, however, is not satisfactory because, as was highlighted above, medial and lateral OFC have different connectivity profiles and comparatively little direct interaction (Carmichael, Price, 1996), and while the human AMFC cluster belongs to the medial network, most animal studies focused on area 13, in the lateral network. At the anatomical level, the evolutionary differences between species are small and the study of functional connectivity fingerprints confirms that human AMFC corresponds to macaque AMFC and human OFC corresponds to macaque OFC (Neubert *et al.* 2015). The fact that when tasks are carefully matched across species the results are similar (Gottfried *et al.* 2003) argues against major functional reorganization across regions.

Only a very limited number of published studies have recorded single-unit activity from macaque vmPFC; this is not only due to a lack of interest, in fact multiple laboratories reportedly found in pilot studies fewer value-encoding neurons in vmPFC, compared to lOFC (Wallis, 2012). Bouret and Richmond (2010) recorded single-unit activity from both vmPFC and lOFC and found that lOFC neurons respond to external motivational cues, such as stimuli predicting a certain amount of reward, whilst vmPFC neurons were tuned to the amount of reward expected from self-initiated movements. A further investigation revealed that in a task where monkeys had to choose whether to accept or reject offers of varying amounts of juice for varying amounts of physical effort, vmPFC neuron activity was strongly predictive of willingness to work, while at the same time only weakly correlated with objective option features (San-Galli *et al.* 2016). This suggests that the medial network deals more specifically with value when it is linked with internal processes.

Anterior cingulate cortex

Although some previous accounts have stressed a role of ACC in conflict detection or error expectation (Botvinick, 2001; Brown, Braver, 2005), ACC has been shown to be more generally involved in associating actions with their reinforcement consequences (Matsumoto *et al.* 2003), whether they are positive or negative; its activity reflects the value of search and behavioural change (Rushworth *et al.* 2004; Kolling *et al.* 2016). Binary value-guided choices, where motor costs are equated, generally show an fMRI effect of economic value in AMFC rather than more posterior ACC (Levy, Glimcher, 2012), but electrophysiological recordings in ACC found neurons responding to subjective chosen value (Cai and Padoa-Schioppa, 2012).

Arguably context-invariant economic value is an abstract construct, and binary choice tasks are not ecologically realistic models of all aspects of human and animal decision

making. Tasks modelled on animal foraging, where a subject needs to choose whether to engage with a present option or rather keep searching, do activate ACC also in human fMRI (Kolling *et al.* 2012). In particular, ACC encodes the search value, which depends on the average richness of the environment compared to the presently available option. Engaging in an option implies an opportunity cost, because the time and energy of engaging with the present option may not be spent with a potentially more valuable option elsewhere (Charnov, 1976; Stephens, Krebs, 1986).

There are different ways in which ACC estimates the environment value. First, its outcome-related activity depends on volatility (Behrens *et al.* 2007). When the situation is more volatile and the need for a change in behaviour is more likely to be indicated by feedback, then outcome-related activity is greater in ACC. Second, as already explained, its decision-time related activity reflects the value of an alternative choice or set of choices (Kolling *et al.* 2012; 2016). Thus, both at decision and outcome times, ACC activity is higher when a change in behaviour is more likely to occur. The estimates of the values of alternative behaviours that ACC holds, at the time of the decision, depend on whether reward rate is, on average, stable, increasing, or decreasing (Wittman *et al.* 2016). In addition, its multiple estimates of the average value of a choice are held at different timescales simultaneously (Meder *et al.* 2017).

ACC is involved when the choice is framed in terms of actions rather than stimuli (Rushworth *et al.* 2007) and is the location where costs and benefits of an effortful option are weighted in both rodents and primates (Walton *et al.* 2003; Rudebeck *et al.* 2006, Croxson *et al.* 2009) and ACC lesions cause a lack of self-initiated movements (Thaler, 1995). On the other hand, response conflict seems to be more associated with the superior frontal gyrus (SFG) than ACC (Rushworth *et al.* 2004). Finally, ACC gyrus (the

most ventral part) plays a specific role in social cognition (Amodio, Frith, 2006; Wittmann *et al.* 2018).

Dorsolateral prefrontal cortex

dlPFC has been often implicated in the cognitive control of choice and motor behaviour (Ridderinkhof *et al.* 2004; Lee *et al.* 2012; Cieslik *et al.* 2012); it is part of a multiple-demand network, including also the inferior parietal sulcus, the anterior insula and ACC (Duncan, Owen 2000; Duncan 2010). Such network is consistently activated across a range of complex cognitive tasks requiring for example sequential mental programming, focusing on relevant (as opposed to irrelevant) features, and working memory. Particularly important for the present research, dlPFC is recruited by the performance of novel, as compared to familiar, tasks such as the repetition of a new sequence (Jueptner *et al.* 1997). The multiple-demand network has been shown to be largely consistent between humans and macaques in a resting-state connectivity study (Mitchell *et al.* 2016). More specifically in the field of value-guided decision-making, it was found that dlPFC can modulate the subjective representation of value, as encoded in vmPFC, on the basis of factors such as the healthiness of food items (Hare *et al.* 2009). The relative activation of vmPFC and dlPFC is influenced by cognitive regulation and has an impact on behaviour (Hutcherson *et al.* 2012). dlPFC activity is enhanced when subjects choose delayed rewards compared to immediate ones (Hare *et al.* 2014), it correlates with the variability of the value predictions afforded by option attributes (Kahnt *et al.* 2011); dlPFC has also a specific role in moral judgment (Greene *et al.* 2004, Ruff *et al.* 2013) and metamemory (Miyamoto *et al.* 2017).

Frontopolar cortex

It was found that the human frontal pole, together with the inferior parietal sulcus, is activated during exploratory behaviour (Daw *et al.* 2006), it represents the value of

switching to a course of action alternative to the one presently implemented (Boorman *et al.* 2009) and underlies metacognitive appraisal in both perceptual and value-guided decision-making (Fleming *et al.* 2012; De Martino *et al.* 2012). It is believed to form the apex of the executive system underlying decision-making, allowing switches between plans or tasks (Koechlin, Hyafil, 2007; Badre, D'Esposito, 2007).

Modelling of the decision process

After a brief overview of some of the most relevant findings about the contribution of each area of prefrontal cortex to value-guided decision-making, we can now return to the question how these areas work jointly and how these empirical findings match the theoretical frameworks described above in section 1.3.1. The step of valuation is often associated to OFC (Rangel *et al.* 2010), because of the consistency with which it is activated by expected outcome value and outcome identity in animals, but also in humans when appropriate designs are used; however, as we said, mOFC and lOFC should not be treated as the same region. It is worth remembering that the OFC value representation is not the only determinant of decisions; in particular it includes most factors, but not certain costs. Costs that are perceived as intrinsically linked to the outcome, such as temporal delay, are integrated in this representation, but costs perceived as linked to the action itself, as opposed to the outcome, such as physical effort, may constitute a different representation, which is instead encoded in ACC (Rushworth *et al.* 2007); it is therefore difficult to argue that any single area constructs an absolute value signal upon which decision-making is based.

The location of the step for action selection is equally controversial; most human researchers believe that vmPFC or AMFC is the region that performs the key comparison and triggers the choice, at least if the choice does not involve the weighting of motor costs; not only a large number of correlational imaging studies supports this hypothesis,

but lesion studies as well. A specific hypothesis about how the selection is performed is based on the drift-diffusion model. Basten and colleagues (2010) found that compound stimuli predicting both monetary gains and losses elicited a difference signal in vmPFC that is consistent with the prediction of a drift-diffusion model of action selection. A lesion study in macaques (Noonan *et al.* 2010) found that a lesion of mOFC, but not lOFC, increased errors in trials where the value of the two options was close, and abolished the independency of the decision from the influence of irrelevant alternatives.

Regarding the localisation of the comparison process, single-unit data from non-human primates should be the technique best suited to provide an answer. Nevertheless, results have been discordant. Some researchers defend OFC as the key comparison area (Padoa-Schioppa, 2011); however, a rare study recording in vmPFC area 14 from macaques performing a binary gambling task gives reason to think that this is the key comparison area (Strait *et al.* 2014) because single-cell activity is consistent with a mutual inhibition model, such as the one described above by Wang (2002). This result matches MEG data in humans showing a similar pattern (Hunt *et al.* 2012). Another electrophysiological study in macaques provided a different insight (Kennerley *et al.* 2011): multi-site recordings show that during the decision process, firing-rate and local field potential (LFP) in many prefrontal areas follow a similar dynamical profile; however, a precise temporal analysis indicates that ACC explains dlPFC activity during effort-based choices and OFC explains dlPFC activity during delay-based choices (Kennerley *et al.* 2011; Hunt *et al.* 2015). Even if this study did not record in vmPFC, it gives an indication that at least one other region, namely dlPFC, shows an activity profile consistent with a mutual-inhibition network, a conclusion supported by a refined temporal analysis of human fMRI signal by Domenech and colleagues (2017).

At the time of the outcome, reward value is represented in large parts of prefrontal cortex, including AMFC and OFC, but also in the basal ganglia; in particular, a reward prediction error signal, proportional to the difference between actual and expected outcome, is provided by dopaminergic neurons in the midbrain (Schultz *et al*. 1997) projecting to the ventral striatum and to multiple cortical prefrontal areas. Outcome-based learning is mediated by specific areas: as described above, ACC is instrumental in the formation of action-outcome associations and their effect on subsequent behaviour. Lateral OFC has a role in reward-guided learning and credit assignment. Signal detectable after feedback in either of these regions predicts changes in the course of action at the next trial (Noonan *et al.* 2010; Chau *et al.* 2015; Fouragnan *et al.* 2018).

The overall picture of value-guided decision-making sketched by the studies review here is less one of a series of discrete, temporally ordered steps centred around an absolute representation of an option's value, and more akin to an interaction of many components working dynamically and in parallel, each one making slightly different contributions depending on the type of decision process, and where multiple forms of value modulate the neural activity profile. The existence of a unitary "action selection module" and its neural localisation are still open questions.

1.3.4 Habitual and goal-directed behaviour

The present review of the neuroscience of choice will be completed by this section, focused on the difference between habitual and goal-directed learning and action control systems. The question is related to many of the aforementioned results, and is important for the problem of novel versus familiar decision-making, although it comes from a specific tradition of instrumental conditioning studies in animals, and has generally been framed in different terms.

Complex actions (think of playing piano) performed for the first time require a higher degree of cognitive control, they are slower and more prone to errors, whereas well-rehearsed actions become habitual and are performed faster and more accurately, with a smaller involvement of prefrontal executive control. In some sense, all complex actions require choices as well; for example, a rat repeatedly exposed to the same large maze will initially need to choose a direction at every bifurcation: after some training its pattern of choices will be better than chance, but still not very accurate, and the animal will display hesitation at each crossing; after a much longer training, it will run seamlessly in the right direction every time. A downside of excessive training, however, is that it causes behaviour to become more inflexible and less sensitive to environmental changes (Dickinson, 1985; Graybiel, 2008)

In the field of neuroscience, a lot of attention has been paid to the differences between three learning systems: the Pavlovian system, the habitual system, and the goal-directed system (Balleine, O'Doherty, 2010; Rangel *et al.* 2008). We will focus in particular on the last two. More recently, computational accounts have explored in depth the difference between model-free and model-based behaviour, and parallels have been drawn between habitual and model-free behaviour as well as between goal-directed and model-based behaviour (Dolan, Dayan, 2013).

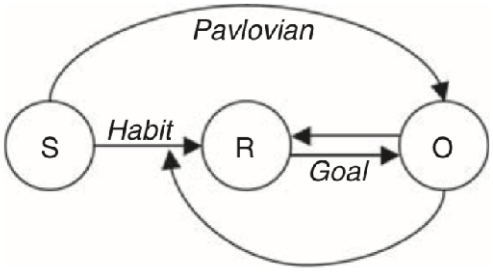
Notwithstanding the large interest brought to the difference between habitual and goal-directed behaviour, and the direct link between frequently repeated behaviour and habit formation, there is a surprising absence of research investigating the difference in neurocognitive mechanisms for first-time online choices versus well-rehearsed habitual choices in the classic framework of binary value-guided decision-making.

The Pavlovian system associates a reflexive stereotyped response with a stimulus, but it is characterised by the fact that the stimulus-outcome association is not contingent to

any action by the subject. The most classic example was provided by Pavlov (1927) who observed that a dog, used to hearing a tone before receiving food, would start salivating when hearing the tone, even in the absence of food.

The habitual system associates a stimulus with an action; although the association is learned because the action leads to a rewarding outcome in the presence of the stimulus, following over-training the stimulus-action association becomes involuntary, impulsive and therefore partly independent from the outcome (Dickinson, 1994). When someone is immersed in thoughts and realises he has been driving to the street where he used to live until the previous year, his behaviour is an example of a habitual stimulus-action association.

Finally, the goal-oriented system associates actions with outcomes in a flexible, purposeful way (Dickinson, Balleine, 1994). In this case an action may be initiated by a stimulus but only if the subject has an expectation that the action will help to obtain some outcome and if the outcome is desired. In both humans and animals, when a stimulus-action pair is reinforced by reward a large number of times, the action response may turn into a habit and lose its goal-directed character. For example, if rats are trained to press a lever to obtain pellets, and then pellets are devaluated by feeding the rats to satiety, rats who underwent moderate training would reduce the amount of lever pressing, but rats who underwent long training would not (Holland, 2004).



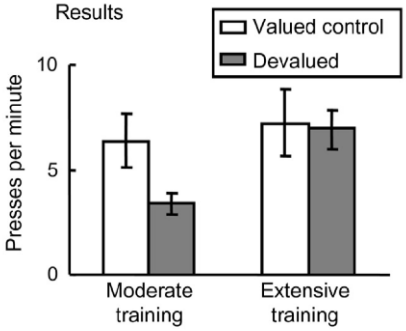


Figure 1.6 Left: Relation between the Pavlovian, the habitual and the goal-directed systems. S: stimulus, R: response, O: outcome, from Balleine *et al*. 2009. **Right:** effect of training on goal-directed versus habitual responding in rats. Image from Daw, O'Doherty (2009), with permission from Elsevier.

The striatum, which is a deep structure, part of the basal ganglia, and composed of the putamen and the caudate nucleus, has a central role in reinforcement learning. Three functional subregions have been distinguished based on both animal studies and human fMRI studies (Daw, O'Doherty, 2009): dorsolateral striatum is linked to habitual behaviour (Jueptner *et al.* 1997), dorsomedial striatum to goal-directed behaviour (Yin *et al.* 2005). It has been argued that ventral striatum primarily encodes a Pavlovian response (Parkinson *et al.* 2002) but others have argued for its causal role in model-based reinforcement learning (McDannald *et al.* 2011). While non-declarative learning is enabled by the striatum, it is the medial temporal lobe that mediates declarative learning (Poldrack, Packard, 2003). Prefrontal regions have also been implicated in goal-directed learning and action control, in particular vmPFC, which is consistent with its role in decision making discussed above (Balleine, O'Doherty, 2010). vmPFC-lesioned monkeys keep responding to a previously rewarded item even when the contingency is removed and OFC lesioned monkeys are insensitive to outcome devaluation (Rudebeck, Murray, 2011), which are the two criteria for lack of goal-directed behaviour (Yin, Knowlton, 2006). But at the same time, OFC, with amygdala, also represents a Pavlovian value signal (Ostlund, Balleine, 2007; Hatfield *et al.* 1997).

Interestingly, goal-directed model-based behaviour does not depend on the classic reward prediction error found in ventral striatum, but is associated to a "state prediction error" encoded in dlPFC and intraparietal sulcus (Gläscher *et al.* 2010). Antzoulatos and Miller (2011) recorded in dorsal striatum and dlPFC while monkeys were learning to

associate arbitrary stimuli with left or right saccades. When the number of associations to learn was small, the striatum encoded the response before dlPFC, but when the number of stimuli to learn became very large, subjects needed to group them into categories, and dlPFC was a better predictor of saccade than striatum. This confirms the role of dlPFC as a multiple-demand system and its relevance for decisions where responses to stimuli are not learned individually but are derived from a more abstract understanding of the task structure.

In sum, a specific literature has addressed the neural underpinnings of habitual actions, defined by a stimulus-response association, compared to goal-directed actions. This is distinct from the question of value-guided decision-making, but it impacts it in a broader sense, because any choice, and in particular over-trained choices, may elicit a mixture of Pavlovian, habitual and goal-directed responses (O'Doherty, 2014). The familiar and novel choices studied in this thesis cannot be simply identified with habitual and goal-directed responses, because they are in some sense both model-based, but nevertheless this research field constitutes the foundation upon which we may build a new account of novel and familiar value-guided decision-making in primates.

1.4 Functional magnetic resonance imaging for cognitive neuroscience

A very large range of techniques have been used to study the brain and link its activity to cognition and behaviour; some of these techniques are non-invasive and can be used in healthy human volunteers, others are restricted to human patients or non-human animals. The two main features that differentiate neuroscientific techniques are the direction of causality and their resolution, in both space and time (Huettel *et al.* 2008).

The direction of causality determines whether it is a measurement technique or an intervention technique; in measurements causality can be thought as going from the brain to the instrument: the instrument detects some form of brain activity during cognition, without modifying it. Electrophysiology, electroencephalography (EEG), and fMRI are all measuring techniques. In intervention techniques the causality goes from the instrument to the brain: the instrument interferes with the brain or its activity and the effects on behaviour are assessed. Transcranial magnetic stimulation (TMS), transcranial focused ultrasound stimulation (TUS), and lesions are examples of intervention techniques.

Not all techniques have the same resolution. Spatially, they go from a single-neuron resolution with electrophysiology, through millimetres with fMRI and TUS, to centimetres with EEG, MEG, TMS and lesions. Temporally, the resolution goes from millisecond-scale with electrophysiology, EEG, MEG, to seconds with fMRI to the permanent effects of lesions (Huettel *et al.* 2008).

In this section, we will discuss the foundations of the main technique used for the present research, namely functional magnetic resonance imaging (fMRI), including model-based fMRI. TUS will be described in chapter 5 (sections 5.1 and 5.2.2).

Magnetic resonance imaging of a static object is made possible by the resonance effect at the atomic scale, combined with a sophisticated system of magnetic gradients varying in time and space that enables the spatial reconstruction of the object. Brain functional imaging with blood-oxygen-level dependent (BOLD) contrast is a specific use of MRI that acquires whole-brain images very quickly, every few seconds, and the contrast is sensitive to changes in blood oxygenation; this allows tracking local neural activity through time albeit in a somewhat indirect manner (Ogawa *et al.* 1990).

All atomic nuclei spin around themselves; additionally, some nuclei like hydrogen nuclei (single protons) are said to possess the nuclear magnetic resonance (NMR) property if they have both a magnetic and an angular momentum; a collection of such nuclei forms a spin system. When a spin system is placed in a magnetic field, each nucleus precesses (like a spinning top) around the main field axis and the precession frequency is a linear function of the magnetic field strength. If an electromagnetic pulse is delivered by a transducer coil to a spin system, its energy will be absorbed only if the pulse frequency matches the nuclei precession frequency, which constitutes the phenomenon of magnetic resonance (Purcell, 1946; Bloch, 1946). Generally, MRI techniques focus on single protons (very abundant in organic samples because water contains hydrogen atoms), but other elements possess the NMR property as well. After a sample has absorbed the energy pulse, its net magnetisation is flipped to the plane orthogonal to the main magnetic field and this effect is detected by a receiver coil. The whole sample produces only a single output (an electromagnetic wave); however, a mechanism of gradients allows deconstruction of the temporal profile of the signal into a 3-D image (Lauterbur, 1973). Most modern MRI scanners use a 3 tesla static magnetic field; on top of that, a z-gradient slightly modifies the strength of the field along the z-axis; in this manner, because the proton frequency depends on the field, the transducer

pulse will only be absorbed by a thin slice of the sample, that is determined by the exact pulse frequency. One by one, slices are excited and their signal is measured, until the whole 3-D volume is reconstructed. Within each slice, additional x- and y- gradients modify the overall magnetic field once the excitatory pulse has happened. By quickly changing the field in the x- and y- directions, even if the signal from the whole slice is acquired simultaneously, the gradients continuously modify the spatial frequency and phase of the detected signal. Therefore, by knowing the exact timing of the pulses and the temporal evolution of the magnetic gradients at each spatial location, from the temporal profile of the measured signal, it is possible to reconstruct the whole 3-D image. The spatial frequencies of an image are represented in k-space, which is the inverse Fourier transform of the image in physical space. Standard structural imaging acquires one line of k-space after each excitation. This approach has been widely used to acquire static images of the body (Damadian *et al.* 1977), mainly for clinical purposes, but also for research.

Functional MRI needs two additional elements: firstly, the volume acquisition has to be much faster, going from many minutes for a structural scan down to few seconds. Specific acquisition sequences, such as echo-planar imaging (EPI) have been developed (Stehling *et al.* 1991). With EPI, the whole k-space, and therefore a whole slice image, is acquired after a single pulse. However, in order to allow the increased speed, the spatial resolution and the signal-to-noise ratio must necessarily decrease. In particular, inhomogeneity in the static magnetic field induced by density changes, for example at the boundaries of the nasal and aural cavities, or by bodily movements, can cause signal dropouts or incorrect reconstruction of the spatial source of the signal (Pusey *et al.* 1986).

Secondly, it is necessary to measure some correlate of ongoing neural activity and the BOLD signal is such a correlate (Ogawa *et al.* 1990). Oxygenated haemoglobin (Hb) is

diamagnetic, whilst deoxyhaemoglobin (dHb) is paramagnetic, which means that the second one distorts the surrounding magnetic field (Pauling, Coryell, 1936). By exploiting the same phenomenon that causes artefactual signal dropout, BOLD fMRI can measure the fluctuations in signal intensity due to the relative concentration of dHb in the blood (T2* contrast). The relation between neural activity and blood oxygenation is modelled by the haemodynamic response function (HRF). Initially, areas where activity is higher will consume more oxygen and dHb will increase, however the vascular system quickly responds by increasing the blood flow and delivering more oxygenated Hb. This overcompensation means that precisely in areas where the neural activity is higher, after a few seconds the ratio of dHb to Hb concentration decreases (Malonek, Grinvald, 1996) and the BOLD MRI signal increases.

The HRF in humans peaks after approximately 6 seconds and returns to baseline after approximately 12 seconds (which is followed by a small negative effect for the following 10 seconds). This implies that even very brief neural events may have long-lasting impact on BOLD signal. This phenomenon should be taken into account in the experimental design of the task (Mandeville *et al.* 1999).

Once the image has been reconstructed based on the procedure described above, a raw dataset is obtained, consisting in a 3-D movie of the head, with one measurement taken every pulse (repetition time, TR), usually every few seconds. Image intensity is based on the T2* contrast, that, as it was described, is sensitive to the relative deoxygenation of the blood. Preprocessing, which is the image treatment preceding the statistical analysis, is then performed. One important step is motion correction, to compensate potential movements of the subject in the scanner bore during data acquisition. For each session and each participant, images are registered to a high-resolution structural image of the subject's brain, acquired separately, and then all subjects are registered into a common,

standard space to allow group-level statistics. The step of brain extraction sets non-brain voxels to zero; then spatial smoothing and high-pass temporal filtering are routinely applied.

fMRI spatial and temporal resolution is limited by two factors: one is physiological and depends on the coarseness and slowness of the vascular response to neural activity; the other factor is physical: the resonance signal to noise ratio (SNR) is proportional to the size of a voxel and the duration of a TR, so that increasing resolution increases noise at the same time (Huettel *et al.* 2008).

Classic fMRI designs were by blocks, whereby a participant engaged in a given cognitive task for, for example, 30 seconds, followed by 30 seconds of either rest or a different cognitive task, and so on. In this way, the BOLD signal during the condition of interest may be subtracted from the signal in a control condition. A brain region is said to be activated by a task if the BOLD signal in that region is significantly higher in a given condition compared to the control.

In event-related fMRI, every event is modelled separately. Events are not necessarily equivalent to trials, and there may be more or less than one event of interest per trial. But because the HRF is slow, it is of the utmost importance to decorrelate the signal belonging to separate events. Typically, such a design will include long intervals between events of interest to allow the BOLD effect to return to baseline before the next event.

A more sophisticated approach is afforded by parametric designs. In such cases, a general linear model (GLM) is built with numerous regressors. One regressor for every event of interest will be constant, for example one regressor may be equal to 1 during stimulus presentation and 0 otherwise, another regressor may represent reward delivery. Another set of regressors will be parametric, but still time-locked to an event of interest: for example, a regressor representing the size of the reward may have a range of values

between -3 and 3 during the reward delivery time and be 0 otherwise; parametric regressors are generally normalised and take different values at each trial.

Parametric regressors allow estimation of how, on top of an average event-related neural effect, a trial-by-trial modulation occurs, depending on the level of a specific parameter such as the value of the chosen option. All regressors are convolved with the hemodynamic response function (HRF) to give a profile representing the expected temporal evolution of the BOLD signal if the brain region is indeed responsive to the variable represented by the regressor. All these regressors are then used simultaneously in a GLM to predict the BOLD timecourse of each voxel. A parametric design enables a better decorrelation in faster tasks, because even if the BOLD signal has not returned to baseline before the following event, consecutive events may have different parametric values and therefore their contributions to the overall signal may be distinguished statistically.

When the beta coefficient of a constant regressor in a GLM has a positive value, it means that the area is more active during the represented event compared to baseline, if the beta is negative, it means that the BOLD signal is suppressed compared to baseline. The BOLD decrease suggests a decrease in neural activity compared to rest; although other hypotheses are possible, such as the "steal effect", the evidence for them is weak (Smith *et al.* 2004; Kastrup *et al.* 2008). In the case of parametric regressors, for example the value of the chosen option, because the variables are z-scored, the neural signal is enhanced in trials where the estimated beta coefficient is positive and the variable takes a higher value; it is suppressed if the variable takes a lower value. But when the estimated beta coefficient is negative, it means that the neural activity is enhanced when the variable has a lower value and is suppressed when it has a higher value. The GLM approach allows inclusion of constant and parametric regressors for effects other than the

one of interest: by modelling them separately, if the regressors are decorrelated, the contribution of other events may be removed from the overall BOLD signal, enabling a more precise estimate of the size of the effect of interest.

Once beta coefficients are estimated at a single session level, they can be contrasted. For example, the main effect of reading a meaningful word can be subtracted from the main effect of reading a nonsensical word. Parametric regressors can be contrasted as well. Once the size of each contrast of interest has been estimated in each session and subject separately, a higher-level group analysis may be used to reveal whether, for example, the effect is on average significant, or if there is a significant difference between one group of subjects and another.

A well-known problem of fMRI is posed by multiple comparisons (Bennett *et al.* 2009): the functional image is reconstructed in three-dimensional units called voxels. There are generally millions of voxel in the final standard space, which means that thousands of voxels will be activated just due to noise. However, a traditional Bonferroni correction would be too strict for any real neural effect to survive. The most common approach in fMRI is to use family-wise cluster correction, which takes advantage of the fact that usually neural activations appear in confined areas spanning multiple contiguous voxels; if a single voxel has a large effect but is not surrounded by other voxels with a similar effect, chances are high that this may be due to noise. This technique first thresholds all voxels at a pre-set level, such as for example $Z>2.3$, then clusters continuous voxels with an effect larger than threshold, then counts the size, or mass, of each cluster, and finally assesses what are the chances that a cluster of this size or mass would be produced by chance (Worsley, 2001).

Overall, notwithstanding the intrinsic limits that we described, fMRI has imposed itself in the last three decades as one of the most important techniques in cognitive

neuroscience, affording a unique combination of non-invasiveness, ability to measure whole-brain activity at relatively high resolution, and great flexibility in its usage with very different experimental designs.

1.5 The present thesis

The series of experiments presented in this thesis were driven by an interest in the mechanisms underlying novel, as opposed to familiar, value-based decision-making. As mentioned in section 1.3.4, this question has not received much attention in the neuroscientific field, compared to the habitual and goal-directed systems in cognitive control. We wanted therefore to study specifically how novel value-guided multiattribute choices engage the brain differently from well-rehearsed ones. We hypothesized that separate neural mechanisms may potentially underlie the abilities of performing well-rehearsed and first-time choices. If such a mechanism specific for novel choices exists, we expected it to be more active for novel compared to familiar trials, but also more active for novel trials at the beginning of the experiment compared to the end. However, our hypothesis predicted that the mechanism may be more active for trials with a smaller value difference, compared to those with a larger one, leading to a negative effect associated with the parametric regressor.

Moreover, we noted that the strikingly different results in value-guided tasks between human and non-human primates are confounded by differences in paradigms and measurement techniques. Humans generally are subjects of fMRI studies and do not receive an extensive training before the experiment, whereas monkeys are used in electrophysiology and lesion studies and do require extensive training.

In order to take apart these elements, we designed a series of experiments where macaque monkeys performed both familiar and novel choices during fMRI acquisition. This design allows testing different combinations of the three key factors (species, technique, novelty of the task). Monkeys have very rarely been tested with fMRI, in particular in the field of decision-making (but see Chau *et al.* 2015; Papageorgiou *et al.*

2017; Miyamoto *et al.* 2017; Fouragnan *et al.* 2018). The specificity of this experiment was a design that allowed an extensive training in a subset of the choice space, while the remaining options were left completely unexplored; later, at the time of testing, subjects were asked to also make choices among the previously unexplored combinations. Given the multi-attribute nature of the stimuli, the monkeys could readily make inferences about the value of novel attribute combinations and perform adaptive choices in the novel condition as well. This ability to make first-time online novel choices reflects a similar ability in humans; but notwithstanding this remarkable adaptation to the novel choice space, we could observe clear differences in both behaviour and neural signal between familiar and novel conditions. Furthermore, in order to go beyond the correlational nature of fMRI experiments, we performed an additional causal investigation of the role played by the network identified with neuroimaging and revealed the behavioural consequences of disrupting this network.

In sum, three experiments were conducted on macaques. Following an extensive training of approximately four months in the familiar version of the task, experiment 1 combined fMRI with a binary choice task where half of the trials were familiar and half were novel. Experiment 2, performed in interleaved sessions with experiment 1, used the very same novel stimuli but presented in fast single-option trials, to investigate neural representations of option values, independently from choice. After approximately six months of interval and another three months of familiar training, the monkeys were tested in experiment 3, with a very similar task, but combined with focused transcranial ultrasound stimulation.

This thesis describes the results obtained from this sequence of experiments. Chapter 2 describes the behaviour of the monkeys in the novel choice task of experiment 1, focusing on the notion of decision under risk and on the biases displayed by the subjects.

Chapter 3 describes the neuroimaging results of the same experiment 1, highlighting neural correlates of novel and familiar choice. Chapter 4 describes the results of experiment 2. fMRI data were analysed using two different approaches: firstly, we assessed the repetition suppression effect when comparing options of same versus different value; secondly we searched for a neural signal compatible with a hexagonal grid-like coding of the two-dimensional value space. Finally, chapter 5 describes experiment 3 and the impact of transcranial ultrasound stimulation on value dimension integration. Chapter 6 draws general conclusions relating to all the experimental chapters.

Chapter 2
Macaque monkeys facing novel risky choices. A cognitive and economic approach

2.1 Introduction

The difference in neurocognitive mechanisms between familiar and novel decision-making has not received much attention. It is known that habitual patterns of action selection occur in a more automatic way, that is faster, has less variance, but is also less sensitive to relevant environmental changes (Dickinson, 1994). However, most experiments investigating habitual versus goal-directed choice have used complex multi-stage tasks, such as escaping a maze, or the two-step Markov decision task designed by Daw and colleagues (2011), where the key for success is to learn the structure of the task. On the other hand, value-based decision-making is by nature subjective, it doesn't necessarily require rehearsal of a difficult motor skill nor figuring out complex hidden contingencies. Nevertheless, in the field of value-based decision-making, it is possible that familiarity with a task has an equally large impact on behaviour and the neural mechanisms underlying its performance. This impact, however, has not yet been explicitly investigated.

Humans and other primates do engage in both familiar and novel value-based decision-making in their everyday life. When we go to the usual cafeteria for lunch, the chances are high that we need to sample among a menu that we have already experienced many times; even if in such conditions our choices are automatized, we need to engage every day again in this familiar decision process, each time with a slightly different outcome. On the other hand, if we move to a new city and we need to buy a house, all

possible options will be new to us. Even if we can assign a value to the individual attributes of the houses, such as their price, their size, their location, and as a consequence can make a satisfactory choice, it is possible that this type of one-off novel decision differs in the way it engages our brain. In some way the difference is analogous to that between driving to a new hotel while on holiday abroad as opposed to driving the usual and familiar path from home to work day after day. The experiments and analyses in this chapter and in the following ones examine the hypothesis that a specific cognitive and neural mechanism is recruited when choices are performed for the first time.

The reason why we are able to make adaptive choices among novel options never experienced before is that, as mentioned above in section 1.1.1, no option is ever completely novel (nor completely familiar for that matter). Items and events have multiple attributes and we may have a varying degree of familiarity with the component attributes even if we have never experienced their combination in a given, specific fashion. And even when we do not know a specific attribute we can often identify an approximate link to a larger category. For example, I may never have tried dhansak lamb curry, but if I have experience of other types of lamb dishes and other types of curry dishes, I can form an approximate expectation of how much I will enjoy the dhansak.

The two choice attributes that are investigated in the following experiments are the amount and the probability of a reward, namely blackcurrant juice. The use of these classic economic dimensions allows us to create a continuous two-dimensional space (instead of a space composed of discrete attributes) and is consistent with common paradigms used in non-human primate research. This design also allows us to study risk sensitivity in macaques, their decision biases, and how these effects change between familiar and novel choices.

More generally the goal is to study decision-making in novel situations and how this is implemented in the brain. In this chapter, focusing on the subjects' behaviour, we will be able to replicate some findings about risky decision-making in primates, as well as identifying new effects not previously reported. Additionally, we will be able to shed a new light on the topic by contrasting familiar and novel decisions. As discussed in the general introduction (section 1.2), amount of prior experience with a choice is likely to affect decision-making in multiple ways, and not only in terms of risk sensitivity. For instance, if the value of a multidimensional option is learnt from experience, its representation can become more precise with exposure. However, if there are systematic cognitive biases in value representation, they may not disappear with practice.

2.2 Methods

2.2.1 Overview

In order to assess how monkeys make value-guided decisions when facing novel situations, we designed a multi-attribute value space in such a way that the monkeys would know the meaning of component perceptual features of each new stimulus they encountered, but still have to make an inference to compute their overall values. This led to a two-dimensional design where reward is characterized by the magnitude and the probability of fruit juice being delivered to the animal. These are the two aspects of reward most frequently manipulated in the study of risky decision-making. Subjects aim to receive maximum reward, but under some conditions, may need to choose between smaller safer or larger riskier options.

Perceptually, the amount of juice is cued by the colour of the stimulus and the probability of the juice being delivered is cued by the number of dots in the stimulus. We distinguished the space of all possible magnitude / probability combinations into a familiar training set and a novel testing set (see below). The training set was designed in such a way that the monkey would have extensive experience of the exact amount of juice associated with each shade of colour and the exact probability of reward associated with each number of dots. Still, most of the possible combinations of stimulus features would not be actually experienced by the animal until the key test.

We investigated at the same time how novel options were assessed, and also how novel binary choices were made. The training task included only one-dimensional trials, where the animals would compare two options differing in only one attribute at a time. In this way, during training it was never necessary to trade off amount for probability, and one option was always objectively better. The test choices, instead, included all types of

stimuli pairs, most of which would differ in both attributes at the same time, and no one may be objectively superior.

This design allowed us to study the behavioural features, and, as we will discuss in later chapters, the neural correlates, of novel decision making, which requires two types of inference: firstly, the evaluation of each option, which may include a new combination of magnitude and probability not experienced before; secondly, the choice between two options, which may include new pairs not compared before. Comparison may also require to arbitrate the conflict that might occur when options differ in two respects simultaneously so that one stimulus is better in one dimension and the other stimulus is better in the other dimension.

Two experiments were run at the time of the first exposure to new stimuli, in alternate sessions; the behavioural data discussed in this chapter come from experiment 1, but experiment 2 will be mentioned to clarify the training procedure and the order of the data collection.

2.2.2 Experimental setup

Four male Rhesus macaques, here referred as A, B, C and D, aged between 4 and 7, and weighting between 11kg and 13kg at the time of the experiments, were trained to perform binary choices in a computer-based task. Monkeys lived on a 12-hour light-dark cycle, were fed once per day after testing, and had ad-lib water access for at least 3 hours a day. In most cases they had water access for 13 to 17 hours a day.

During training and testing, monkeys were sitting in sphinx position in a purpose-built MRI-safe chair (figure 2.1) with their head fixed by a cranial implant (for the implant surgery, see methods in chapter 3 section 3.2.2). They were trained in a wooden custom-made mock scanner and then tested in a standard 3T MRI scanner (for details about

scanning, see methods in chapter 3, section 3.2.3). The monkeys could freely move their arms and made their choices by touching one of two infrared sensors placed in front of them, 15 centimetres apart, in line with the stimuli.

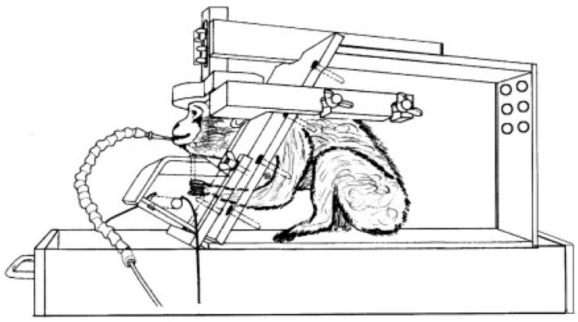


Figure 2.1 Monkey chair and reward delivery system. Drawing adapted from Vanduffel *et al.* (2001), with permission from Elsevier.

The stimuli and the feedback were showed on a display during training sessions and projected onto a custom-made plastic screen by a LX400 projector (Christie Digital Systems) with a custom-made lens during MRI sessions, with the screens positioned in both cases approximately 30cm away from the monkeys' head. The reward was delivered through a spout into their mouth and consisted in fruit juice (25% blackcurrant Ribena juice and 75% water); for two out of four monkeys, blended banana was added to the juice (half banana per litre). The room was dark except for the light coming from the screen, and even in training sessions pre-recorded MRI noise was played, with an intensity level matched to the real scanner (100dB). No earplugs were used, but the monkeys adapted to the level of noise gradually during an extensive familiarisation period. All procedures were conducted under a Home Office licence according to the Animal (Scientific Procedures) Act 1986 and the European Union guidelines (EU Directive 2010/63/EU).

2.2.3 Task design: the value space, stimuli, timing and conditions

The full space of possibilities was a 10 by 10 matrix of magnitude and probability of reward (figure 2.2), with magnitude varying from 1 drop to 10 drops of juice (0.5 ml each) in steps of 1 drop and probability varying between 10% and 100% of rewarded trials in steps of 10%. Magnitude was cued by colour and probability by number of dots. However, the monkeys never experienced the full space of possibilities until the testing phase in the MRI scan. Monkeys were trained in two sub-regions of this space: on even days they were presented only with stimuli with a magnitude of 5 but with a probability that was randomly drawn between 10% and 100% and varying in steps of 10% (figure 2.2, horizontal grey line); by contrast on odd days, they were only presented with stimuli with a probability of 60% and a magnitude drawn randomly from 1 to 10 drops varying in steps of 1 drop (figure 2.2, vertical grey line). Jointly, these stimuli will be defined as familiar stimuli. Training trials were either forced choices (a single option was presented on the screen) or binary choices of familiar stimuli (animals chose between two options).

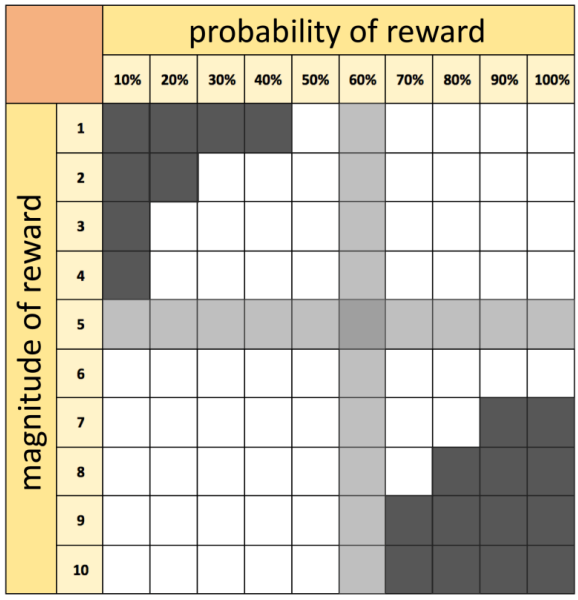


Figure 2.2 Value space. Familiar options in light grey, novel options in white, unused options in dark grey

This training procedure ensured that monkeys:

1) experienced only a subset of all possible combinations of magnitude and probability,
2) made only one-dimensional choices, comparing stimuli differing only in terms of one attribute,
3) learned the association of all the colour shades with magnitude levels and all dot numbers with probability levels.

At the time of testing, monkeys were presented for the first time with forced or binary choices of novel stimuli, requiring an integration of the attributes to determine the stimuli values and make choices. These trials were randomly interleaved with an equal number of familiar trials in experiment 1. The novel trials required the monkeys to:

1) Evaluate novel stimuli made of combinations of magnitude and probability not previously experienced together
2) Choose between pairs of stimuli differing in both dimensions at the same time

The expected value range of the stimuli, i.e. the magnitude times the probability, used during training was between 0.5 and 6 drops, and was preserved in the testing phase. Therefore, combinations of magnitude and probability that led to very low or very high expected value, outside the familiar range, were excluded (dark grey options in figure 2.2). The monkeys needed to either respond to a single option, or choose between two options presented on the left and right sides of the screen (figure 2.3).

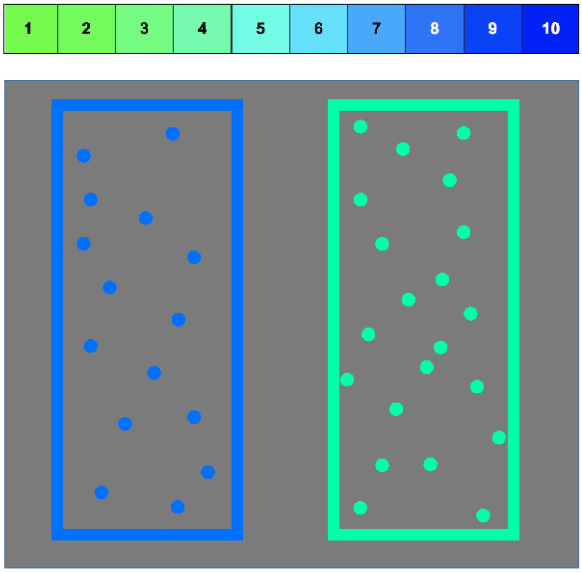


Figure 2.3 Top: colour coding used for the magnitude. **Bottom:** visual appearance of the task. In the example, a novel conflicting trial, the left option gives 8 drops with a 50% chance and the right option gives 4 drops with a 70% chance.

The background was grey throughout the task, in such a way that the overall luminosity was constant whether the stimuli were present or absent. Each option was represented by a 29 by 9 cm rectangle containing a number of dots varying between 3 and 30, in steps of 3. Each 3 dots represented a further 10% probability of reward, so that 30 dots represented 100% probability. The frame and the dots were coloured and the colour varied in ten intermediate shades between green and blue. Full green represented 1 drop of juice, full blue represented 10 drops, each drop corresponding to 0.5ml. Immediately after the response the option(s) disappeared and after a delay a feedback cue appeared on the chosen side, consisting of an upward triangle if the choice was rewarded and a downward triangle otherwise. This was simultaneous with the reward delivery (if any).

Experiment 1 was based on a choice task with the following timing (figure 2.4): intertrial interval (ITI) between 5s and 7s; decision between two options with free

response time, and a deadline of 60s; action-outcome delay between 3.5s and 4.5s; outcome time of 3s. If monkeys did not respond within the deadline the trial terminated and was repeated at the end of the session. The long action-outcome and intertrial delays were designed to allow disentangling of the blood-oxygen-level dependent (BOLD) signal related to the decision and the outcome. After response, the outcome time was fixed for all trials, regardless of whether the reward was delivered and how large it was.

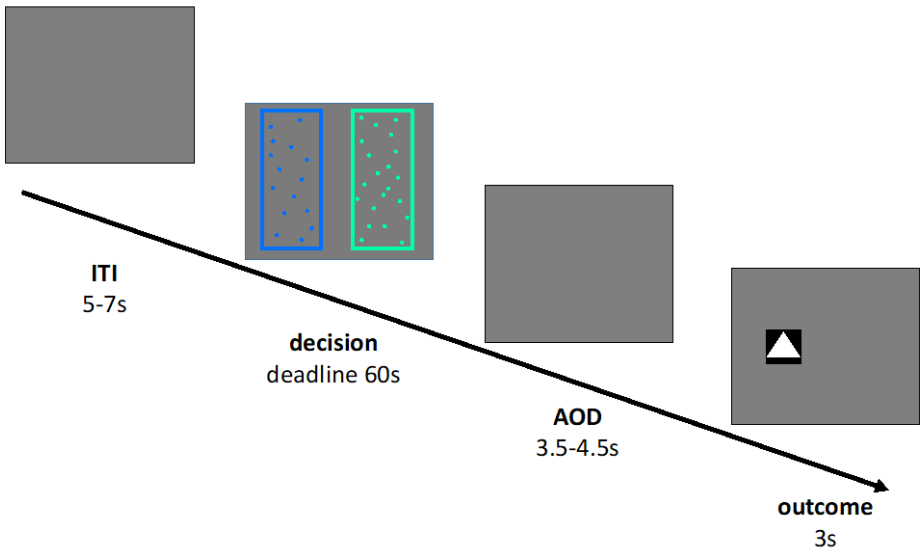


Figure 2.4 Choice task (experiment 1) timeline. ITI: inter-trial interval. AOD: action-outcome delay. The outcome event is either positive or negative depending on the chosen probability.

Each session consisted of 180 trials, trials with no response were repeated at the end of the session; monkeys consumed on average 400ml of juice per session, depending on performance; most sessions had a duration between 45 and 60 minutes but never above 120 minutes.

In experiment 1, all trials offered monkeys a binary choice, and trials were divided into six conditions comprising a 2 by 3 design in which the 30 trials were given per

condition, as illustrated in figure 2.5 and table 2.1. The first two-level factor was the *novelty* of the stimuli. The second three-level factor was the *dimensionality* of the choice.

Novelty of the stimuli:

- *familiar* (both options were familiar)
- *novel* (both options were novel)

Dimensionality of the choice:

- *coherent* (one option had both higher magnitude and higher probability than the other option)
- *one-dimensional* (the two options had an identical attribute and differed in the other attribute)
- *conflicting* (one option had higher magnitude and the other option had higher probability)

choice / stimuli	coherent	one dimensional	conflicting
familiar	1	2	3
novel	4	5	6

Table 2.1 Conditions in the choice task (experiment 1)

In both the coherent and one-dimensional conditions, for each choice one option is objectively superior, while in the conflicting condition, the best option depends on risk sensitivity.

In figure 2.5 one exemplar pair illustrates each condition. *Familiar* options are in the light grey area and *novel* options in the white area. *Coherent* comparisons are between pairs along the top-left / bottom-right diagonal, *1-dimensional* comparisons are between

horizontal or vertical pairs, *conflicting* comparisons are between pairs along the top-right / bottom-left diagonal.

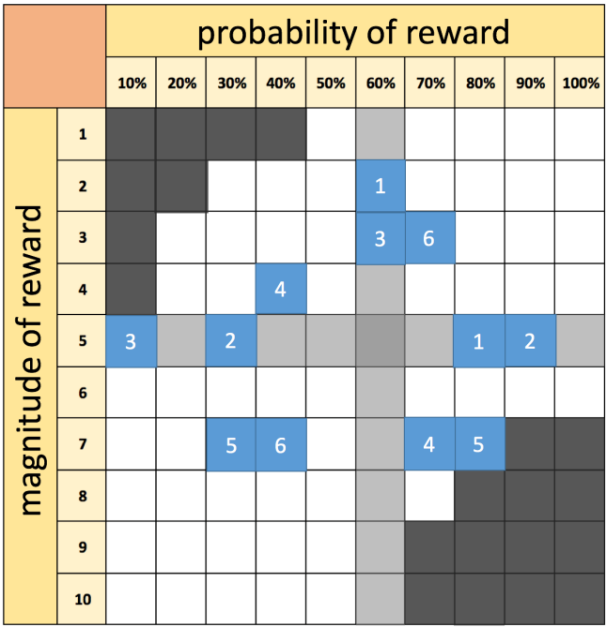


Figure 2.5 Exemplar choices.

1: familiar coherent;

2: familiar one-dimensional;

3: familiar conflicting;

4: novel coherent;

5: novel one-dimensional;

6: novel conflicting.

The choices sequences, different for every session and every animal, were designed in such a way as to respect a number of constraints, some of which were necessary for the fMRI analysis (described in chapter 3, section 3.2.1). The 180 choices of each sequence were divided in 30 choices per condition, the value sum of the two options was on average the same in each of the six conditions. The value difference could not, by definition, be matched across the three different levels of the dimensionality factor, but both the mean of the value difference and the variance of the value difference across trials were matched within each pair of conditions: familiar and novel coherent, familiar and novel one-dimensional, and familiar and novel conflicting. By matching it is intended that the absolute difference between the means in each condition would be smaller than 0.2 drops, and the absolute difference in variance would be smaller than 3.5 drops^2. Finally, we took care that the correlation between the best value and the worst

value offered in trials would not be stronger than r=0.33 when merging all familiar conditions and when merging all novel conditions. This was done to ensure that, using general linear model (GLM) analysis it was possible to examine whether the value of both options available, both the one ultimately chosen and the one ultimately rejected, affected decision accuracy and speed and, as shown in chapter 3, decision-related neural activity. It must be noted however that, within some conditions, and in particular the conflicting familiar condition, the correlation between best and worst option values had to be much higher.

In experiment 2, in order to assess repetition suppression effects and the gridlike code (see chapter 4), a faster version of the previous task was designed. In these sessions the setup was identical and the stimuli were also the same, but only one stimulus at a time was presented on the screen (on a random side) and they were all novel. The timing was as follows (figure 2.6): ITI between 0 and 0.5s, decision time fixed at 1s for monkey B, 1.1s for monkeys A and D and 1.5s for monkey C, no outcome delay, outcome time 3s.

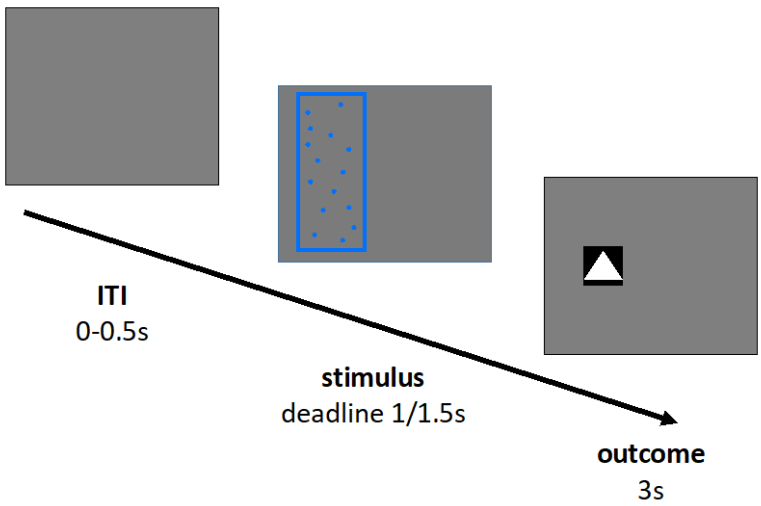


Figure 2.6 Sequence task (experiment 2) timeline. ITI: inter-trial interval. The outcome event is either positive or negative.

This fast timing was designed to minimise the interval between a stimulus and the next. If a monkey did not respond within the fixed decision time, the trial was unrewarded and repeated at the end. If he did respond, the outcome cue (positive or negative) appeared after the full decision time, regardless of the reaction time of the animal. And again the outcome time was independent from the actual occurrence of a reward.

2.2.4 Training and test stages

Monkeys were extensively trained with familiar stimuli only, for 65 (±10)[3] sessions, corresponding to 13300 (±500) trials and 12000 (±1200) responses, before testing. No-response trials were not analysed. Half of the sessions were fixed-probability and the other half were fixed-magnitude (figure 2.2; grey regions), all sessions included familiar stimuli only. During the training monkeys also learned to adapt to the timing of the two versions of the task, both of which were challenging: the slow one because of the long action-outcome delays, and the fast one because of the short response deadline.

Initially monkeys were trained in an adapted version of the choice task, in which all, half, or a third of the trials were forced choices (where only one stimulus was presented and the other side was not rewarded) in order to allow monkeys to sample repeatedly all possible stimuli, even the less favourite ones. In this initial phase the action-outcome delay (AOD) and the intertrial interval (ITI) were short and the number of trials was higher (300). Across days, the AOD and ITI were increased, and the number of trials was decreased, until when the monkeys were able to perform 180 trials with the desired AOD and ITI (4±0.5 and 5±1). All these sessions included at least a third of forced-choice

[3] Here and in what follows, ± indicates the range (from maximum to minimum), while σ represents the standard deviation.

trials. During the later stages of the training, monkeys alternated between this version of the slow task and the fast task with only one option per trial and a response deadline decreasing towards the target 1s, but for some monkeys we stopped short of this target, depending on behaviour.

Monkeys went through three stages of training, before the experiment start: first in the wooden mock scanner, then in the real MRI scanner but without data acquisition, and thirdly while neuroimaging data were acquired. A transient drop in performance was registered at each transition. The relative time spent by each monkey at each stage was variable, as illustrated in figure 2.7, and depended on its performance and previous exposure to MRI.

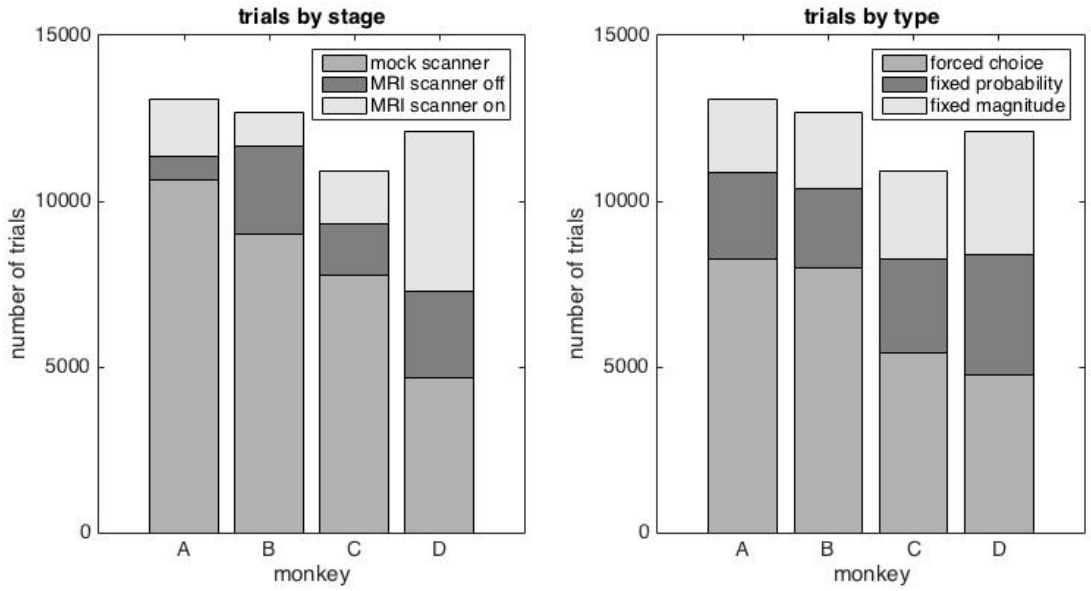


Figure 2.7 Number of completed training trials per animal per condition.

In the testing phase, each monkey performed 17 sessions in consecutive days (excluding weekends). They were organised as follows:

2	1	1	1	2	1	1	1	2	1	1	1	2	1	1	1	2

Table 2.2 Intermixed session order of experiments 1 and 2. The twelve (light grey) sessions of experiment 1 were alternating with the five (dark grey) sessions of experiment 2.

Choice sessions belong to experiment 1 and sequence sessions belong to experiment 2. Monkey C failed to complete the planned stimuli sequences in some sessions, but the trials that were performed were still used for analysis. The number of incomplete trials is substantial for two choice sessions (73 and 45 respectively) but still below 40%, and small in three further sessions (below 5% of the trials). In summary monkey C performed over 94% of the programmed trials in experiment 1. Monkey A missed only two planned choices in one session so that it performed over 99.9% of the programmed trials in experiment 1 and monkeys B and D performed all trials.

2.2.5 Model-free behavioural analysis

For the analysis of behaviour during training and testing, we generally defined *accuracy* as the β coefficient of the value difference parameter (left vs right) in a GLM predicting choice side with a logit link function. A higher coefficient indicates better discrimination between better and worse options. In some GLMs two regressors were included: absolute value difference (parametric) and condition (categorical), with optimal choice as dependent variable; this was done to test whether performance differed across conditions. Notice that when the dependent variable is the side, the regressor used is relative value difference (left value minus right value), but in order to compare conditions we had to

use correctness as dependent variable, and absolute value difference (best minus worst) and condition as regressors.

However, in some contexts performance was estimated as simply the proportion of trials in which the chosen option had an expected value equal or larger than the unchosen option, out of the total number of trials. The latter approach has the advantage of expressing performance in a fixed range between 0 and 1, but because it weighs all errors equally, it does not penalise more strongly errors that are made when the value difference is larger. This approach was used mainly for illustrative purposes, to report training and overall test performance, but also to compare the dimensionality condition (coherent, one-dimensional, conflicting) effect on performance (figure 2.12).

For training sessions, subjective value estimation was made unreliable by the fact that all choices were along a single dimension; therefore, the analysis of training data was based on the objective values of the stimuli only. A test was performed to assess the learning of the familiar stimuli by measuring performance improvement during the training phase; all sessions were first merged, and then the trials allowing a choice were separated in fixed-magnitude and fixed-probability groups and then, for each group, binned in 16 bins, each including 188 (±44) trials, the approximate size of a test session. The average performance during the last 5 bins was used as a benchmark; a one-tailed t-test at a significance level of $p=0.05$ assessed whether the performance in each initial training bin was lower than in the later-stage benchmark. Also, the effect of bin number on performance was tested with a general linear model (GLM). This, and all following GLMs, included a logit link function to predict binary variables.

Monkeys' learning of this task was made easier by their previous participation in a different experiment in which magnitude and probability were also connected to colours and number of dots respectively, but with a different structure.

Subsequently, we studied the monkeys' behaviour in the test phase, using the 12 choice sessions. Firstly, separately for each subject, we assessed the overall accuracy in terms of proportion of correct choices, then we ran a logistic regression to estimate the effect of value difference on the side of the choice, and also a linear regression to assess the effect of absolute value difference on the median reaction time measured in 40 bins of equal size per animal. The median was used because reaction time is strongly positively skewed. Additionally, a GLM with two regressors was run to measure the independent contribution of magnitude difference and probability difference to the choice. The ratio of the beta coefficients was calculated.

Later, trials in the overlearned condition of the test phase (familiar 1-dimensional trials, coded as "2" in table 2.1) were compared to the last 1000 training binary choices: a GLM was used to assess the effect of condition (training vs testing) on accuracy, above and beyond absolute value difference.

Further analyses distinguished test trials in familiar and novel. Here, familiar trials include "coherent" and "conflicting" choices that were not experienced during the training (conditions 1, 2 & 3 in table 2.1). Accuracy was compared first between the training and familiar testing conditions and then between the familiar and novel testing conditions; accuracy was tested with a GLM with two regressors, absolute value difference (parametric) and condition (categorical), and correct choice as dependent variable.

Within the testing phase, improvement in performance through the twelve sessions was tested, separately in familiar and novel trials, using a GLM where correctness was predicted by absolute value difference and session number.

Trials were then split by dimensionality regardless of novelty (1&4 vs 2&5 vs 3&6 in table 2.1) and performance was assessed separately in each condition by a GLM with correct choice as dependent variable and dimensionality condition as regressor.

Moreover, the effect of novelty on performance was assessed separately for the coherent and one-dimensional conditions (1&2 vs 4&5) and for the conflicting conditions (3 vs 6). This was tested with a GLM where correctness was predicted by absolute value difference.

Finally, in order to determine whether decisions took into account both dimensions of value, six separate GLMs per subject were run, one per condition, each one predicting choice from magnitude difference and probability difference. Then, focusing on the familiar and novel conflicting conditions, independent GLMs for each session were used and the ratio of the beta coefficients for magnitude and probability was estimated. If the ratio is close to 1, it is a sign that both dimensions were successfully taken into account; even if the session-by session estimate was noisy (any given session has only a limited number of data points), it was possible to use a paired t-test to assess whether the ratio was different for familiar and novel trials, across animals and sessions. The proportion of optimal choices was also measured in each condition.

Up to this point, correctness was always defined based on objective expected value. Below, we turn to the estimation of the subjective values that drove choice.

2.2.6 Subjective value modelling

When choosing among options varying continuously in both magnitude and probability of the reward offered, a normative approach to the task, maximising the total amount of juice obtained in the long run, requires that the Pascalian expected value of each option is assessed independently and then the option with the highest value is deterministically

selected at each trial. The Pascalian expected value is defined as the amount of each possible outcome multiplied by its probability. In this experiment the outcome is binary: either 0 or a defined number of drops of juice, so the expected value reduces to the number of drops multiplied by their probability. The formula for the normative choice is:

$$val_i = mag_i * prob_i \quad (1.1)$$

$$p_{left} = (val_{left} > val_{right}) \quad (2.1)$$

Where mag_i is the magnitude of reward associated with option i, $prob_i$ is the probability of reward associated with option i, val_i is the definition of the value attributed to the option i and p_i is the probability of choosing the option i (which in this case is 1 or 0).

However, we expected the behaviour of the animals to diverge from this normative approach to decision-making. We therefore designed a number of alternative decision models, differing in nine respects, each represented by a free parameter, and tested which one represented better the observed data. In the text below, equations numbered as 1.x will be alternative value definitions, and equations numbered 2.x will be alternative choice criteria.

One hypothesis (Scholl *et al.* 2014; Scholl, Kolling *et al.* 2015) is that instead of estimating the expected value of an option, the subject simplifies the problem by estimating what we call the additive value, which is the reward magnitude summed with the reward probability, after normalising their range between 0 and 1 and adding a weighting parameter (see figure 2.8).

$$val_i = \beta * mag_i + (1 - \beta) * prob_i \quad (1.2)$$

where β is the magnitude/probability weighting ratio. A different way of expressing the same heuristic, is to hypothesize that the subject is first comparing the left and right magnitudes, and computing a magnitude score, then comparing the left and right

probabilities, and computing a probability score, then using these two scores to make a decision. Averaging the two scores would be equivalent to using equation 1.2 with β = 0.5; while with β = 1 it would reduce to the magnitude difference only, and with β = 0 to the probability difference only.

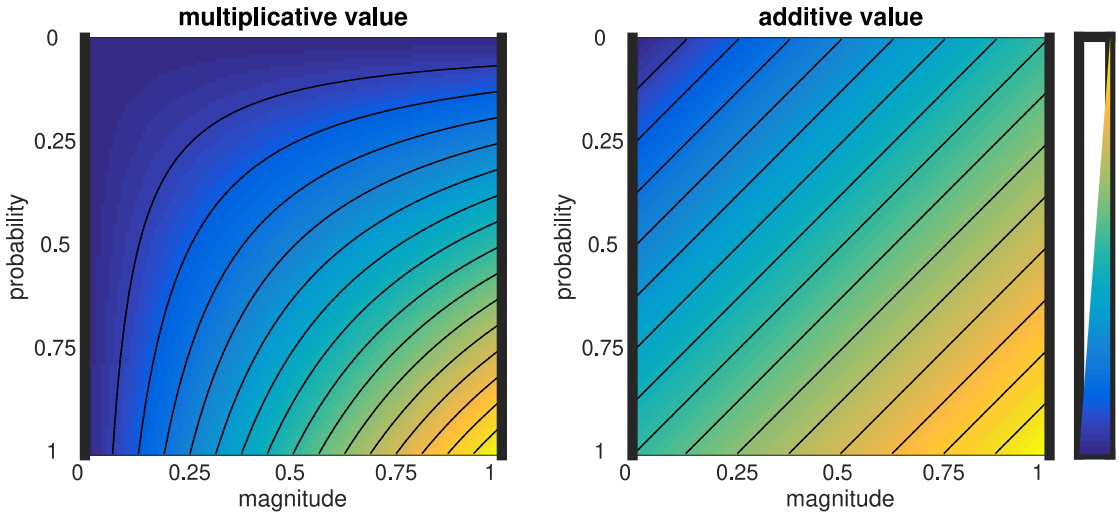


Figure 2.8 Illustration of multiplicative value (the normative definition) and additive value (with β = 0.5) as functions of magnitude and probability of reward

A widely used framework in behavioural economics is prospect theory (Kahneman, Tversky, 1979), described in the introduction (figure 1.1, section 1.2.2). The kink at the reference causes loss aversion, the concavity above causes risk aversion in gains and the convexity below causes risk seeking in losses.

In our experiment, we assume that the reference point will be the null outcome (no juice), and there are no possible losses. Therefore, concretely, a prospect theory model will be defined by the same two equations as the normative model (1.1 and 2.1), but magnitude and probability in equation 1.1 will be replaced by their weighted form, described by:

$$W_mag_i = mag_i^{\alpha} \quad (3)$$

$$W_prob_i = e^{-(-\log(prob_i))^{\gamma}} \quad (4)$$

where W_mag_i is the weighted magnitude of option i, and W_prob_i is the weighted probability[4] of option i; α is the magnitude distortion parameter (also called the risk aversion parameter, because values below 1, typical in humans, imply risk aversion) and γ is the probability distortion parameter. Notice that when α = 1, or γ = 1, there will be no magnitude or probability distortion.

Even if this is not common in the economics literature, it is possible to combine the prospect-like distortions of equations 3 and 4, with the additive definition of value of equation 1.2 to obtain a complex form of subjective value.

Contrary to the economics approach, it is common in psychology to assume that the subjects are noisy in their choices (Thrun, 1992), which means that they will not choose deterministically the best option each time, but their decisions will be a stochastic function of the value difference between the best option and the alternative one (or ones, if there are many). The so-called softmax replaces equation 2.1 by defining the probability of choosing left as a logistic function of the difference between the values of the left and right option:

$$p_{left} = \frac{1}{1+e^{-\vartheta*(val_{left} - val_{right})}} \quad (2.2)$$

[4]The formula for the probability distortion originally proposed by Kahneman and Tversky (1979) is $W_prob_i = \frac{prob_i^{\gamma}}{(prob_i^{\gamma}+(1-prob_i)^{\gamma})^{1/\gamma}}$. Later, Drazen Prelec (1998) introduced an alternative but similarly-shaped function defined as in equation 4. The model simulation reported below showed that, with an increasing number of free parameters, the generative coefficient of probability weighting became harder to recover with the original formulation, but was still correctly recovered with the Prelec formulation. For instance, for the winning 5-parameter model, the correlation of the generative and estimated values of the probability distortion coefficient was r=0.45 with the original formulation, but r=0.98 with the Prelec formulation.

where ϑ is the inverse temperature parameter, representing stochasticity, whose name comes from thermodynamics. Notice a lower ϑ means more noise, a higher ϑ means more precision and when it's infinite, the equation reduces to equation 2.1.

It is also possible to assume another type of noise, independent from the value difference between the options. This would be defined by a part of the choices being determined by a random variable, according to the formula:

$$p_{left} = (1-\delta) * \frac{1}{1+e^{-\vartheta*\left(val_{left}-val_{right}\right)}} + \delta * \frac{1}{2} \quad (2.3)$$

that combines the softmax with the random noise, where δ is the random noise parameter, as illustrated in figure 2.9. Notice that when δ = 0, this reduces to equation 2.2.

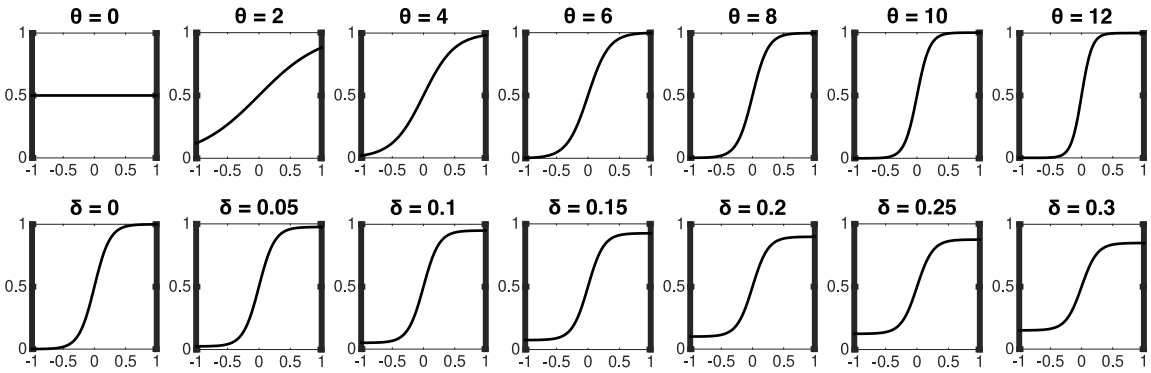


Figure 2.9 Noise parameters effects. X axis: value difference between left and right subjective value. Y axis: probability of choosing left. Top row: inverse temprature (θ) variations, with δ=0. Bottom row: random noise (δ) variations, with θ=8.

Additionally, it is possible that subjects' attention is shifted towards the most salient stimulus, increasing the likelihood of choosing it. Farashahi and colleagues (2018) modelled attentional bias (χ, called DW in their paper) that increases the strength of the

best item in terms of either magnitude, or probability, or expected value. In this way equation 2.3 is replaced with the following:

$$p_{left} = (1-\delta) * \frac{1}{1+e^{-\vartheta*(\chi*val_{left}-(1-\chi)*val_{right})}} + \delta * \frac{1}{2} \quad (2.4)$$

assuming that the left option is the salient one in the present trial, where χ is the attentional bias, between 0.5 and 1.

Finally, three types of spatial biases were tested. The first model, called stable side bias, assumes that, beyond the value difference between the two options, the subject would be more likely to choose *a priori* one side. The second model, called repetition side bias, assumes that the subject is more likely to choose again the same side as chosen in the previous trial. Last, the third model, called win-stay-lose-shift (WSLS) side bias, assumes that the subject is more likely to choose one side if the same side was rewarded in the previous trial and if the opposite side was unrewarded in the previous trial.

These three hypothesis were tested by replacing equation 2.4 with the following:

$$p_{left} = (1-\delta) * \frac{1}{1+e^{-\vartheta*\left(\chi*val_{left}-(1-\chi)*val_{right}+\zeta 1+prev*\zeta 2+wsls*\zeta 3\right)}} + \delta * \frac{1}{1+e^{\zeta 1+prev*\zeta 2+wsls*\zeta 3}} \quad (2.5)$$

where ζ_1 is the stable side bias parameter, ζ_2 is the repetition side bias parameter, which is multiplied by the factor *prev* that is either +1 or -1 depending on the previous choice, and ζ_3 is the win-stay-lose-shift side bias parameter, which is multiplied by the factor *wsls* that is either +1 or -1 depending on the previous choice and reward. Notice that when $\zeta_1 = \zeta_2 = \zeta_3 = 0$, this reduces to equation 2.4.

To determine what model best captured the subjective value and the behaviour of the monkeys, without overfitting the data, we ran a model comparison for each animal's dataset independently. The models tested were defined by the following system of equations, unifying all possible value modulations described in the previous section:

$$W_mag_i = mag_i^{\alpha}$$

$$W_prob_i = e^{-(-\log(prob_i))^{\gamma}}$$

$$val_i = \begin{cases} W_mag_i * W_prob_i\ (multiplicative) \\ (1-\beta) * W_{mag_i} + \beta * W_prob_i\ (additive) \end{cases}$$

$$p_{left} = (1-\delta) * \frac{1}{1+e^{-\vartheta * \left(\chi * val_{left} - (1-\chi) * val_{right} + \zeta 1 + prev * \zeta 2 + wsls * \zeta 3\right)}} + \delta * \frac{1}{1+e^{\zeta 1 + prev * \zeta 2 + wsls * \zeta 3}}$$

where nine parameters are included: β (magnitude/probability weighting ratio), α (magnitude distortion), γ (probability distortion), ϑ (choice inverse temperature), δ (random noise), χ (attentional bias), ζ_1, ζ_2, ζ_3 (stable side bias parameter, repetition side bias parameter and win-stay-lose-shift side bias parameter). The additive and multiplicative models of value were not nested, while all other models were nested and defined by fixing one or more parameters to their default values of β=0.5, α=1, γ=1, ϑ=∞, δ=0, χ=0.5, ζ_1=0, ζ_2=0, ζ_3=0 respectively. Fixing a parameter to its default value is equivalent to remove the bias or noise that it represents from the model. However, the two families of models (the multiplicative and the additive models) could not be nested into each other. Also, while the other parameters are either present or absent (two possible models), the possible models depending on attentional bias χ are four: no bias, bias for highest magnitude, bias for highest probability and bias for highest expected value. Given this space of possibilities, we ran a multiple comparison across 2^7x4 additive models and 2^7x4 multiplicative models. The normative model is the multiplicative one with all parameters fixed to their default value.

In order to compare the models, the free parameters of each model were fit for each subject separately but merging all 12 sessions for each subject, by maximising the log-likelihood of the choices given the model. The models were then compared by their

Bayesian Information Criterion (BIC) scores. The BIC test is more conservative than the AIC when the number of observations is very large.

We then ran simulations using the best model, to test whether it was able to retrieve the generative parameters of a large number of artificial datasets. In particular, we selected the model with 5 free parameters ($\beta,\alpha,\gamma,\vartheta,\zeta 1$) and δ, χ, ζ_2 and ζ_3 fixed to 0, and created 5^8 artificial agents, using all combinations of the 5 free parameters with 8 equidistant values each, in the following ranges: $0.395\leq\beta\leq0.605$, $0.7\leq\alpha\leq3.15$, $0.7\leq\gamma\leq3.15$, $0.8\leq\vartheta\leq32.5$, $-0.035\leq\zeta_1\leq0.035$. Each artificial agent produced stochastic choices based on the stimuli shown to one of the real agents. Then the same model was used to estimate the coefficients from the choices alone. The correlation coefficient of the generative and recovered parameters was assessed. Additionally, we measured the correlations in the signed errors of the recovered values for each pair of parameters.

Once the best model to describe subjective value was chosen, parameter fits for each subject were estimated, first for all trials, and then for familiar and novel trials separately. Subjective accuracy was also estimated in all data and in each condition, with GLMs analogous to those described above, but using the subjective estimate of value difference.

2.3 Results

2.3.1 Model-free behavioural analysis

During the training phase, all monkeys quickly learned the stimulus-outcome associations, and maintained a high level of performance across all 65 (±10) sessions, as illustrated in figure 2.10, where trials are binned in 16 bins per set per subject. Accuracy in the last 5 bins before the experiment onset was above 0.8 in all animals and conditions (fixed-probability trials, 0.82, σ=0.06, fixed-magnitude trials, 0.88, σ=0.04, forced choice trials, 0.98, σ=0.02). Taking the performance in the last 5 bins as benchmark, accuracy in the initial 11 bins was tested and was found significantly lower in only 1.1 bin on average and no more than 3 bins in any animal or condition (one-tailed binomial test, p=0.05). No positive trend in accuracy across all trials was found in any animal or condition (logistic regression, all df>2000, all t<1.8, p=0.05).

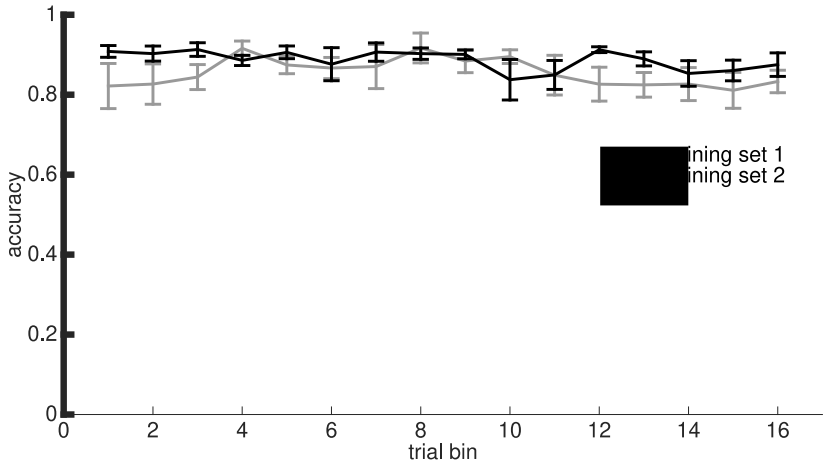


Figure 2.10 Training performance: choice accuracy across all training, divided by condition.

When exposed to novel stimuli and 2-dimensional choices in the testing phase, monkeys were able to integrate their knowledge of the value of the two visual attributes (colour and numerosity) and their relation to outcome, to drive choice. The monkeys' mean accuracy was 0.81, σ=0.07, based on the objective value (expected value) of the options available. Additionally, the proportion of errors scaled with the value difference between the options presented on each trial (figure 2.11, logistic regression, all df>2000, all t>18, all $p<10^{-70}$) and the median reaction time scaled with the absolute expected value difference between the options presented on each trial (figure 2.12, linear regression by bins, all $t_{38}>4.3$, $p<0.001$). Here and in the following analyses, where the dependent variable is the side of the choice, value difference is defined as left option minus right option. Where the dependent variable is instead the reaction time or the proportion of correct choices, then the independent value is the absolute value difference (best minus worst option).

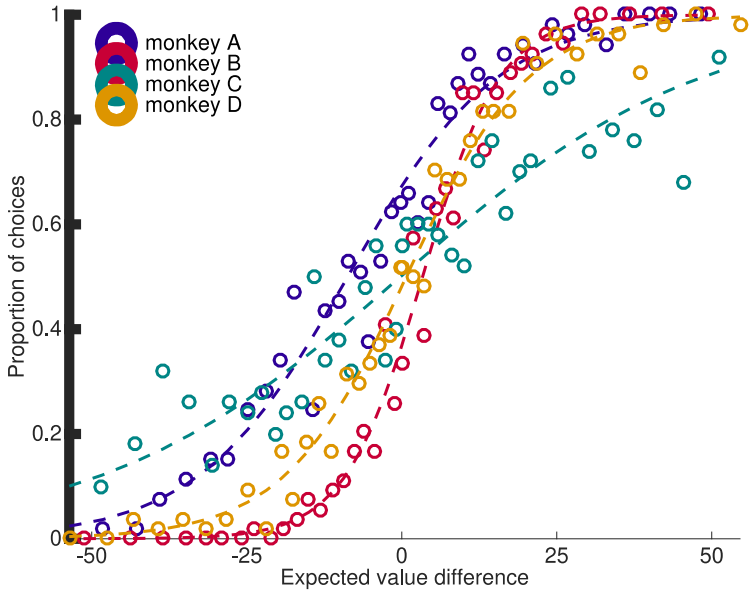


Figure 2.11 Performance as a function of value difference between the stimuli (left minus right). For each subject, choices were binned in 40 bins of equal size, each circle in the figure representing one.

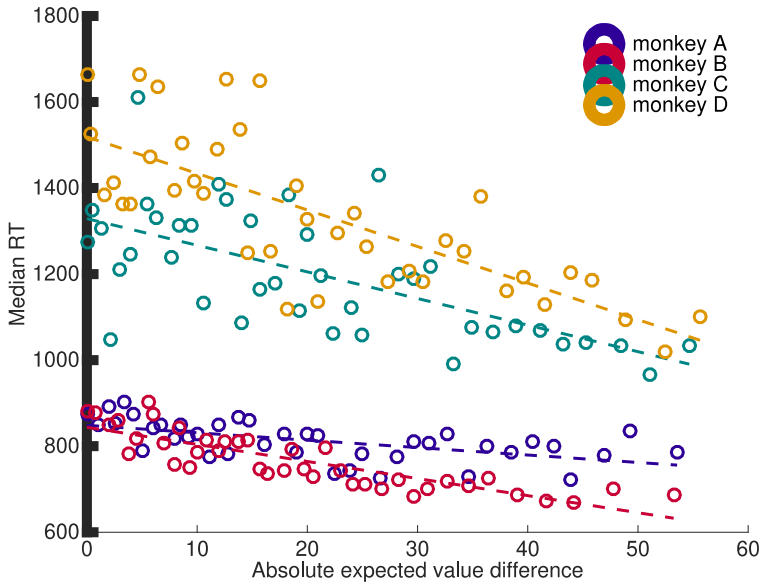


Figure 2.12 Reaction times become faster with larger value difference between the two options. Trials are divided in 40 bins per subject.

These results were not driven exclusively by one dimension of value, because the choices of each subject were significantly influenced by both probability and magnitude (general logistic regression, all df>2000, all t>14, all $p<10^{-45}$). Subjects varied in the degree by which the two dimensions influenced their choice, but were within the same order of magnitude (the ratio of the beta weights for magnitude difference and probability difference obtained with a GLM was 1.13, 1.27, 0.92, 1.08 for monkeys A, B, C and D respectively).

A comparison of accuracy between the last 1000 training choices and the test trials in the corresponding condition (one-dimensional familiar) using a GLM showed that choice accuracy was not significantly different in 3 subjects (p>0.05), and marginally lower in monkey C (t_{1335}=2.00, p=0.046). However, when comparing accuracy across the training, familiar test and novel test conditions, controlling for the dimensionality of the decision, different levels of performance were observed. Notice that the familiar test condition included the same familiar options as in the training, but in pairs of different

dimensionality, making the choice in some cases more difficult. All monkeys performed worse in the familiar test condition compared to the training (all df>3000, all t>3.3 all p<0.001); only two monkeys performed differently between the familiar and novel testing conditions, but with opposite signs (monkey A: t_{2156}=-2.84, p=0.005; monkey C t_{2031}=3.71 p<0.001; monkeys B, D n.s.).

Only one subject out of four showed a significant improvement across time during the testing phase, and this occurred in both familiar and novel conditions (monkey D, familiar t_{1077}=4.75, p<0.001, novel t_{1077}=3.02, p=0.002, all other subjects and conditions p>0.05).

When splitting trials by the dimensionality of the decision, regardless of novelty, the effect of value difference on the choice was modulated by the dimensionality condition, see figure 2.13. Accuracy was significantly higher in coherent vs one-dimensional conditions in three subjects (all df>680, monkeys A, B, D, t>2.86, p<0.004, monkey C n.s.); and significantly higher in one-dimensional vs conflicting dimensionality conditions in all subjects (all df>680, all t>5.38, all $p<10^{-7}$). However, even in the conflicting condition where its effect was smallest, value difference drove choice significantly (all df>667, all t>8.5, all $p<10^{-16}$).

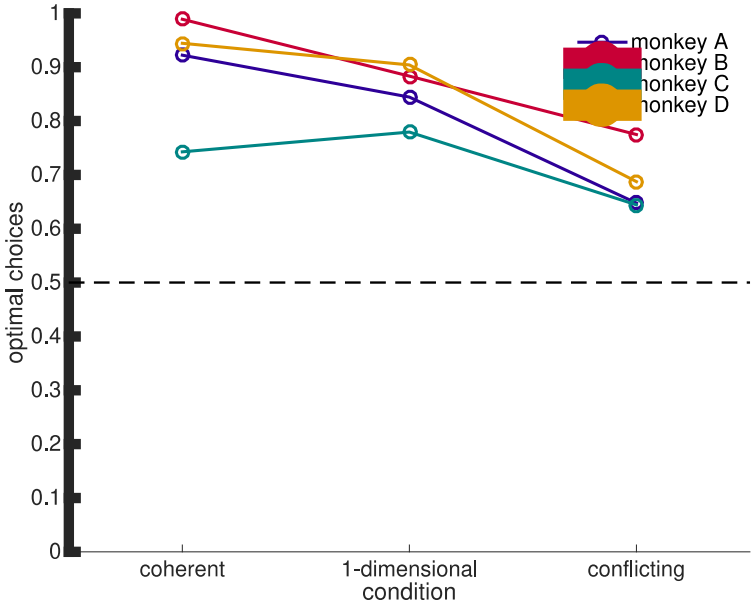


Figure 2.13 Proportion of optimal choices across three difficulty conditions: coherent (one stimulus is best according to both dimensions), one-dimensional (stimuli differ in only one dimension), conflicting (each stimulus is better according to one dimension). The dashed line represents the chance level.

It was then tested whether performance was similar between novel and familiar trials. In conditions where one option was objectively superior, namely the coherent and one-dimensional conditions, no monkey had a lower level of performance in novel compared to familiar trials (all $p>0.05$). When monkeys had to trade-off magnitude with probability (conflicting condition), one monkey (A) performed worse in novel compared to familiar trials and a second one (B) had a tendency to do the same (monkey A: t_{715} =-3.50, $p<0.001$, monkey B: t_{717}=-1.93, $p=0.054$, monkey C: t_{666}=0.66, n.s., monkey D: t_{717}=-1.57, n.s.). Still, all subjects in all conditions performed significantly better than chance (all $df>331$, all $t>4$, all $p<0.0001$), with mean accuracy in familiar conflicting trials: 0.69, σ=0.07; mean accuracy in novel conflicting trials: 0.68, σ=0.07.

The separate contribution of magnitude difference and probability difference was also tested in each of the four animals and each of the six conditions (48 tests). Both were

found to be significant in all animals and conditions except two combinations ($p<0.002$). In particular, in conflicting conditions where magnitude had to be traded off against probability, both quantities significantly drove the choice in both familiar and novel trials (all $df> 330$, all $t>5.2$, all $p<10^{-6}$). In those conditions, the effect of magnitude and probability was in the same order of magnitude, their ratio varying between 0.92 and 1.45, with an average of 1.16. Moreover, the relative effect of magnitude and probability was not different between novel and familiar trials across subjects (paired t-test $df=47$, $p>0.05$).

2.3.2 Subjective value modelling

So far all results used a normative definition of performance as choosing the option with highest expected value. The question of subjective value was then addressed using computational models simulating the choice process through a weighted estimate of the subjective value difference between the left and right stimulus, and a softmax function applied to subjective value difference. A large family of models were compared, including or excluding nine free parameters representing possible biases, listed in table 2.3.

Parameter	Symbol
Magnitude vs probability weight	β
Magnitude distortion	α
Probability distortion	γ
Softmax inverse temperature	ϑ
Random noise	δ
Attentional bias	χ
Stable side bias	ζ_1
Previous side bias	ζ_2
Win-stay-lose-shift bias	ζ_3

Table 2.3 List of possible biases tested by model comparison. Parameters shaded in grey did not improve overall fit.

Fitting all trials for each subject and then averaging, it was found that the model with the lowest BIC score included the following free parameters: magnitude vs probability weight β, magnitude distortion α, probability distortion γ, softmax inverse temperature ϑ and fixed side bias ζ_1; it did not include the following free parameters: random noise δ, attentional bias χ, repetition side bias ζ_2 and win-stay-lose-shift bias ζ_3. It will be subsequently referred to as the $[\beta,\alpha,\gamma,\vartheta,\zeta_1]$ model. Notice that, while the value of β represents the magnitude vs probability weight, the fact that it was included means that the additive model performed better than the multiplicative model, considering that the two are not nested. Hence, the additive model was retained. Figures 2.14 and 2.15 illustrate the relative fit of the 10 best models and all possible models.

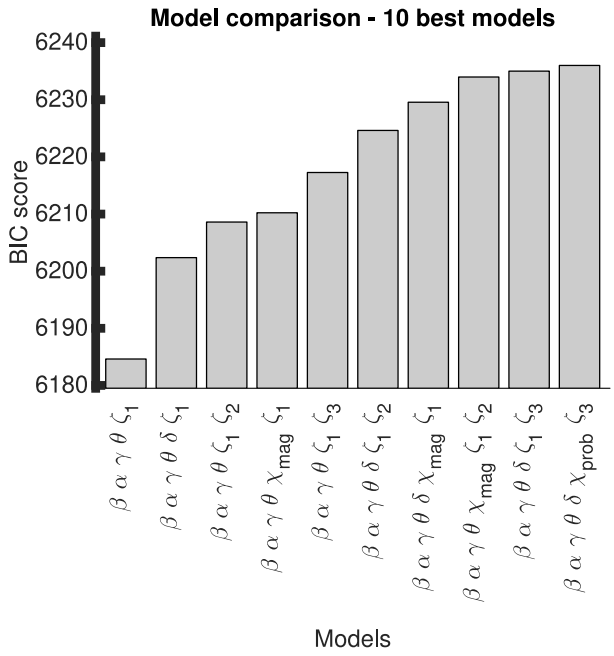


Figure 2.14 Model comparison, BIC score for the 10 best models.

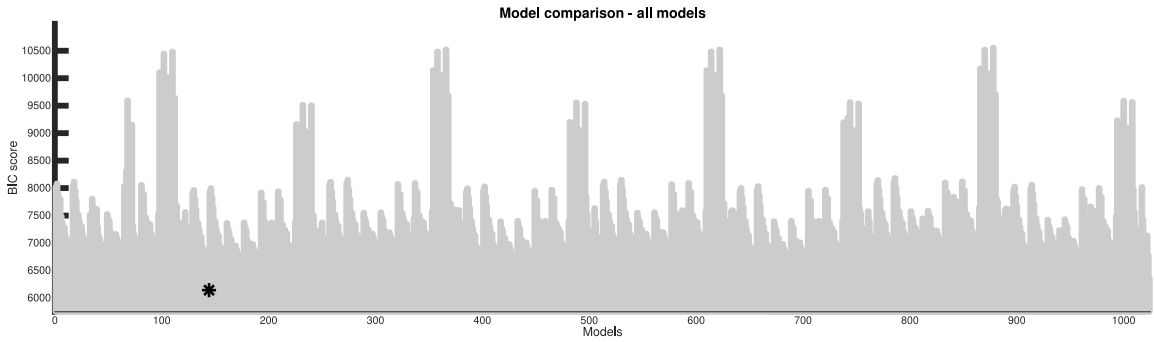


Figure 2.15 Model comparison**,** BIC score for all 2^8x4 models. The star highlights the winning model (a lower score represents a better fit).

When looking at the BIC scores for each subject independently, the result of the model comparison was found to be consistent in all of them, with the following two exceptions: side bias ζ_1 only improved the score of monkeys A and B but not monkeys C and D. Monkey B's behaviour was better explained by a probability-driven attentional bias χ_{prob} while the additive bias β was not retained. In the following analyses, thanks to

this generally high consistency, we will always use the single model that was found to be better on average across subjects.

Next, a simulation confirmed that it was possible to recover the generative parameters of artificial choices created using the winning [β,α,γ,ϑ,ζ1] model, in a plausible range of generative parameters. The correlation between generative and recovered parameters was above 0.98 for all parameters. Additionally, the correlation between estimate errors was below 0.3 for all pairs of parameters.

Finally, based on the [β,α,γ,ϑ,ζ1] model, the specific coefficients for each subject were estimated. They are reported in table 2.4. The distortions caused by α, γ and ϑ for each subject are also illustrated in figure 2.16.

Subject	Additive weight β	Magnitude distortion α	Probability distortion γ	Inverse temperature ϑ	Side bias ζ1
Monkey A	0.463	2.129	1.351	9.81	-0.085
Monkey B	0.416	1.358	1.604	17.32	0.037
Monkey C	0.473	3.262	4.79	4.94	-0.002
Monkey D	0.448	1.986	2.762	12.34	0.013

Table 2.4 Distortion coefficients for each subject based on all trials.

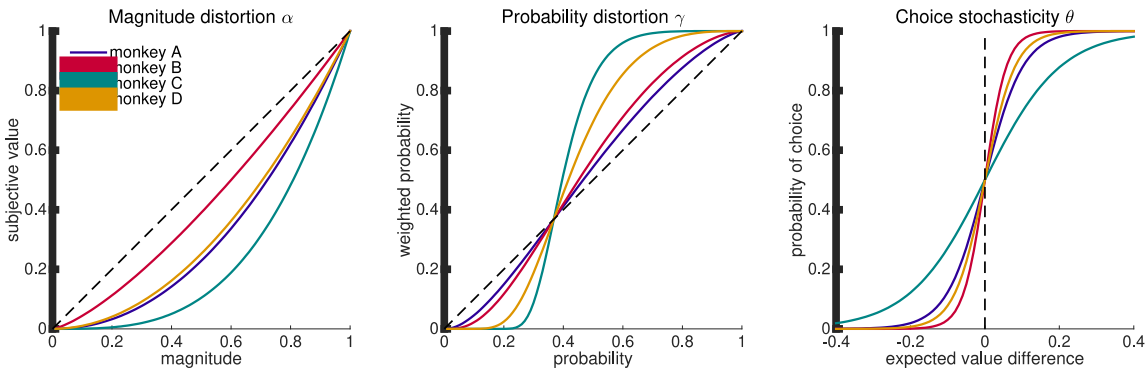


Figure 2.16 Subjective value distortions for subject-specific parameter fits.

It was found that, across subjects, the correlation between objective and estimated subjective value was r=0.90±0.05.

The subjective accuracy, that by definition is higher than the objective accuracy, given that the model was fitted on the choice data, is 0.84, σ=0.06 (compared to 0.81 σ=0.07). The improvement in fit was largely driven by the conflicting condition. The increase in variance explained by subjective compared to objective value difference was 0.05, 0.02 and 0.20 in the coherent, one-dimensional and conflicting conditions, respectively.

Fitting separately familiar and novel conditions revealed that all value biases (β, α, γ, but not side bias ζ_1) were stronger in familiar than novel trials – see figure 2.17. In particular, the additive bias in favour of magnitude went from β=0.36, σ=0.09 in familiar trials to β=0.47 σ=0.04 in novel trials (absence of bias is β=0.5). The magnitude distortion went from α=3.1, σ=1.4 to α=1.8, σ=0.4 (absence of bias is α=1). The probability weighting went from γ=2.9, σ=2.3 to γ=1.7, σ=0.4 (no bias γ=1).

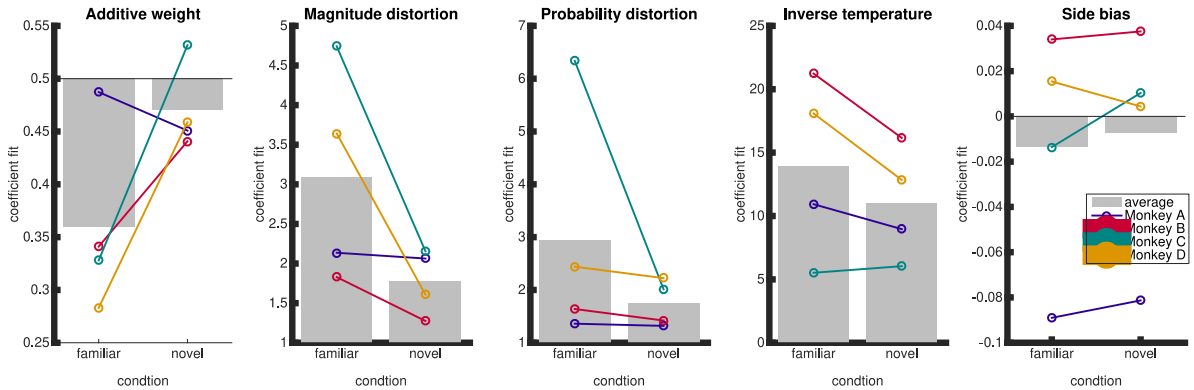


Figure 2.17 Distortion coefficients fitted separately on familiar and novel trials for each subject. Average in grey.

2.4 Discussion

The behavioural part of experiment 1 showed the effectiveness with which the Rhesus monkey can deal with novel tasks by inferring the value of risky options combining the previously independent features of magnitude and probability of reward for the first time. Our study observed similar deviations from neoclassical economic decision criteria, as observed in previous studies, in both familiar and novel decisions.

Monkeys were substantially over-trained in the familiar condition; their performance was at ceiling after just a couple of sessions, but they performed one-dimensional familiar choices for 65 sessions on average. In the testing phase, performance in condition 2, which was equivalent to the training dataset (in both cases each decision was either driven by just the probability difference or the magnitude difference), remained stable. However, their behaviour was also very accurate in the other conditions, not previously experienced. In the conflicting conditions, requiring arbitration between a smaller safer and a larger riskier option, objective performance was lower, and more so for novel than for familiar trials, but still strongly driven by value difference. Altogether, the accuracy and reaction time results, as well as the similar weight exerted by magnitude and probability on the choice, show that monkeys were indeed able to compute the overall value of stimuli and base their decision on it, including in familiar and novel conflicting trials.

The ability of non-human primates to achieve high performance from the outset in a novel decision making task, while perhaps not surprising to ethologists, is not generally acknowledged in decision neuroscience and has never been exploited before in an imaging experiment. Here, out of four different individuals, all were able to perform highly above chance from the first session, even if one of them (monkey C) showed more

noisy and biased behaviour. Interestingly, only one monkey (D) showed a slight improvement across sessions, whilst the others were virtually at ceiling performance levels from the start. In sum, this constitutes strong evidence that macaques have a reliable ability to make inference-based decisions correctly on the first occasion they encounter them. This ability was found to be present in all the monkeys we examined and does not appear to be the special preserve of an exceptional individual.

In terms of subjective estimations of value, we found that classic biases described in prospect theory (magnitude and probability distortions) were also present in macaques, but with an opposite direction to those seen in humans: our monkeys showed a convex (instead of concave) utility function in the domain of gains, meaning that for them ten drops of juice were more than twice as rewarding as five drops; moreover they behaved as if they under-weighted small probabilities and over-weighted large probabilities (S-shaped weighting function). These results may not, however, be due to inter-species differences but rather, as suggested by Heilbronner and Hayden (2013), they may be due to some task-specific features. Humans are equally risk-seeking for very small stakes (Weber, Chapman, 2005) and display an S-shaped probability weighting function when probabilities are not cued linguistically but learned with practice (Barron, Eren, 2003).

Furthermore, we showed that a model assuming additive value (magnitude + probability) explained behaviour significantly better than one that assumed multiplicative value (magnitude x probability), even though it implies suboptimal choices are sometimes made. A similar effect, however, has already been observed in two different experiments with human participants, that were run in our laboratory (Scholl *et al.* 2014; Scholl, Kolling *et al.* 2015).

Notice that in order to reduce the number of parameters to fit we selected the best out of the two models, additive and multiplicative, but an even more careful description of

the subjects' behaviour would be that they are using a mix of the two models (Scholl *et al.* 2014).

This additive as opposed to multiplicative approach has received comparatively little attention previously. From a cognitive point of view, it is plausible to assume that estimating and comparing additive values is much easier than for multiplicative values, but from an economical point of view, this leads to serious distortions of the decision pattern, and it can be complex to model. For example, the expected value difference between a 20% chance of winning £1 and a 20% chance of winning £2 is four times smaller than the difference between an 80% chance of £1 and an 80% chance of £2 (expected values of 20p vs 40p and 80p vs £1.60), but because the ratio is the same it may feel that the difference is equivalent. This matches the general tendency of the brain to focus more on relative rather than absolute quantities. On the other hand, using an additive formula the natural units of expected value are lost (£1 + 20% cannot have a meaningful unit) and this may be the reason why economists have not even tried to follow this path.

Since we used the additive formula of value,

$$(1-\beta) * W_{mag_i} + \beta * W_{prob_i}$$

we were able to adjust the parameter β, to estimate the differential contribution of magnitude and probability to the choice. Notice that, in the more classical multiplicative formulation, if the subject pays more attention to one dimension than the other, this can only be partially and indirectly accounted for by the magnitude and probability distortion parameters. When Farashahi and colleagues (2018) added an extra parameter representing attentional bias in a non-human primate risky decision-making experiment, they found that monkeys paid more attention to magnitude than probability, and also that by modelling this bias, the estimate of prospect theory biases was reduced. Similarly, we

found that our monkeys were more driven by magnitude difference than probability difference. This was illustrated by the fact that the β coefficient was below 0.5 for all subjects.

Furthermore, we explicitly compared models including additive value, attentional bias and both, and found that the Bayesian Information Criterion favoured a model including the additive value only. Notice that the two models make similar but distinguishable predictions. Stimuli can be represented by four quantities: magnitude left, probability left, magnitude right and probability right. Additive value with a preference for magnitude (for example) increases the weight of both magnitude scores (left and right) but not of the probability scores; on the other hand, the attentional model with a preference for magnitude increases the weight of both magnitude and probability on the side of the highest magnitude but not on the other side. They make therefore correlated but dissociable predictions. Farashahi and colleagues found that a model including attentional bias is better than one without, but the reason may be that they didn't test for additive value; at least for our data, when the two biases were put into a competition, attention to a given dimension (additive value bias) was a better explanation of the behaviour than attention to a given option. Also, we noted than once additive value with a higher weight for magnitude was included in the model, the addition of an extra parameter for attentional bias did not reduce any of the other biases further.

Finally, it is interesting that the three value biases described here (weighing for magnitude vs probability β, as well as prospect theory's magnitude and probability distortions α and γ), had the same direction in all four of our animals, even if these represented a wide spectrum in terms of ability in the task, revealed by the inverse temperature parameter ϑ ranging from very low (monkey C) to very high (monkey B).

In terms of the more traditional debate about risk sensitivity, it is important to remember that all biases can affect it: if the additive bias β favours magnitude, subjects will choose riskier options for the same expected value. Similarly, in most cases, when the magnitude and probability weightings α and γ are above 1 (convex and S-shaped), subjects will prefer riskier options of equal expected value. So all three parameters described here contribute to risky behaviour and are consistent with other findings in macaques; on the other hand, as highlighted above, the jury is still out on how much this relates to a cognitive difference with humans, and how much it is due to the experimental design. It is interesting to note that Scholl and colleagues (2014) found that their human participants weighted magnitude more than probability, which leads to risk-seeking behaviour and is consistent with what is shown in this chapter. This may hint at the fact that humans may have simultaneous risk-seeking and risk-averse tendencies driven by different factors.

The most specific feature of our design is nevertheless the comparison of familiar and novel decisions, which allowed us to fit the bias parameters in the two conditions separately. The fact that the bias directions were consistent across all four animals and two conditions indicates the reliability of these estimations. But most interestingly, we observed that all three biases were stronger in the familiar then in the novel condition; this excludes the possibility that these biases may be temporary and may go away with further experience. Even if systematic biases can cause suboptimal choices, they clearly have a different origin and function than mere noise in the choice process, which was captured by the inverse temperature parameter. The latter revealed, as expected, more uncertainty in decision making in the novel condition. Therefore, overall, the transition from first-time choices between new options to over-trained choices between familiar

options seems to imply both a decrease in stochasticity and a strengthening of systematic biases.

This pattern may be hypothetically explained by the recruitment of different cognitive and neural processes for habitual decision-making versus online inferences about the value of new complex objects, assuming that the first circuit is more affected by the kind of subjective value distortions that are revealed by our model.

A complementary view on this phenomenon may come from the description-experience gap literature, where researchers have found that both humans and other primates are more risk-seeking for objects whose value is learned from experience than for objects whose value is cued symbolically (Camilleri, Newell, 2013). In the present experiment all options are cued in the same way, but still animals are more risk seeking in one condition than in the other. It is possible that familiar stimuli are evaluated by a mechanism that is both more automatic and more risk-seeking, which is the one that is generally tapped by tasks learnt through experience. Novel multidimensional options, on the other hand, require online inference and therefore they could be evaluated by a mechanism that is more similar to the one engaged by abstract or linguistic cues in humans. If this was the case, this would explain why both values learned by description and novel values estimated by inference induce more risk-averse behaviour. Of course it is also possible that the experience-description gap and the novelty effect we see here are two independent phenomena.

In any case our data go against the hypothesis that non-normative behaviour in primates may be due to lack of experience, but instead is compatible with the idea that humans, just like other primates, may be risk-averse in the kind of situations described by behavioural economists also because they are engaging with relatively new choices. It can be argued, in an ethological perspective, that it is rational to be relatively more risk-

averse in novel contexts and more risk-seeking in familiar contexts. The work presented here indicates that, if this is the case, then this duality of approaches may not be due to flexibility within a single neurocognitive mechanism but to the interaction between two distinct mechanisms.

In the next chapters, we will go into further depth about how the brain performs the type of value computations postulated here, using neuroimaging and neurostimulation techniques.

Chapter 3
Neural correlates of novel multidimensional choices in macaque monkeys

3.1 Introduction

In this chapter, we are going to study the neural correlates of the multidimensional novel and familiar decision processes described in the previous chapter. We will ask in particular how binary risky choices are made at the neural level in macaque monkeys and how these neurocognitive mechanisms differ between novel and familiar decisions.

There is a striking difference in design between most extant human and monkey value-based decision-making experiments: while in human experiments participants are asked to make on the spot new choices for the first time, animal experiments, in particular those involving monkeys, require very extensive training and a small number of options, such that by the time the neural data are acquired, decision making has become very overtrained and might even be described as habitual. The degree to which neural mechanisms mediating habitual choices in monkeys are genuinely comparable to the mechanisms for novel decision-making in humans is unclear.

Novel choices, as already discussed, are enabled by the ability to combine the value of different attributes, and even if multi-attribute choices are complex processes involving many brain areas, in particular medial prefrontal cortex been shown to play a pivotal role. Patients with vmPFC lesions make multi-attribute choices with a very different strategy compared to both healthy controls and frontal patients with different lesions (Fellows, 2006).

One recent study has focused in particular on the neural system for decision-making among novel items. Barron and colleagues (2013) asked human participants to evaluate new items that they had never experienced before but that were combinations of two known ingredients, such as tea jelly. They found that activity in the medial frontal cortex, both ventral and dorsal, is modulated by the value of the chosen compound option. However, they did not explicitly contrast choices between novel items versus choices between familiar items.

Barron and colleagues also highlighted a potential correspondence between the value-related activation they had found find in the more dorsal part of the medial prefrontal cortex and a previous reported activation in what initially appears to be a very different type of task (Nicolle *et al.* 2012). Nicolle and colleagues investigated how we are able not just to make decisions on our own behalf but how we can model another person and make the decisions that they would want to make for them. They found that the "modelled" value for an agent who will not actually be the recipient of the choice outcome, is also represented in dmPFC, while a more familiar "executive" value for the recipient of the outcome was represented in vmPFC.

The present experiment addresses in particular the differences in neural mechanisms underlying novel and familiar choices. We use in our design binary choices between multi-attribute and risky options, varying in magnitude and probability at the same time. However, our focus is neither on attribute integration per se, nor on risk per se, but in the contrast between novel and familiar decision-making. For this reason, all other features are equated across conditions: both familiar and novel options are bi-dimensional, both involve risk. Comparisons are also matched in terms of difficulty (difference in value between options) and dimensionality (number of attributes to compare and their congruency). It is also important to point out that while familiar choices are, by design,

more habitual, they are not automatic. There were nineteen familiar stimuli in total and therefore 361 different decisions could be made between the different possible combinations of stimuli; it is, therefore, not likely that they were encoded as independent automatic stimulus-response associations. This means that while the value of each familiar option is readily available, the comparison between two options can still be described as a goal-directed (rather than habitual) process.

It was revealed in the behavioural results of experiment 1 (chapter 2) that subjective distortions of value representations are weaker in novel trials. This may indicate that existing value representations based on experience may be encoded in a more idiosyncratic way, affected by systematic biases. On the other hand, the value of novel stimuli, computed on-the-fly as the result of combining two attributes, is not systematically biased but it is much noisier, and therefore novel choices are still overall less accurate than familiar choices (see chapter 2, section 2.3.2). Given this difference between the two conditions, in order to avoid biases in favour of either condition, the initial fMRI analysis of experiment 1 was based on the objective stimulus values (the product of the stimulus' expected reward magnitude and probability); a later analysis compared the neural signatures of subjective and objective value.

3.2 Methods

3.2.1 Task design

As described in the previous chapter, section 2.2.3, the task consisted in choosing between pairs of options, each one represented by a coloured rectangle containing dots, where the colour indicated the amount of juice at stake, varying between 1 and 10 drops, and the dot number indicated the probability of receiving it, varying between 10% and 100%. Familiar stimuli were experienced in more than 12000 training trials, with either a single option or a choice between two stimuli differing in only one dimension for any given session (figure 3.1). At the beginning of the experiment, novel combinations were experienced for the first time and options could differ in both attributes simultaneously. Some combinations were not used in order to maintain the same range of expected values (from a minimum of 0.5 drops to a maximum of 6 drops) in both familiar and novel conditions.

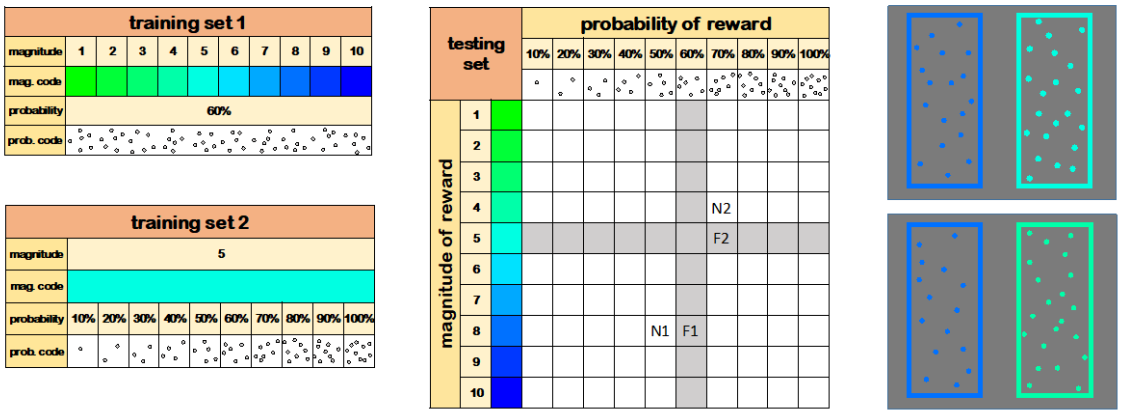


Figure 3.1 Left: Separate training sets for magnitude (top) and probability (bottom) and illustration of the outcome coding. **Centre:** Two-dimensional value space and illustration of the outcome coding. Familiar options in light grey, novel options in white, unused options in dark grey. **Right:** visual appearance of the task. In the top example, a familiar

conflicting trial where the left option (F1) gives 8 drops with a 60% chance and the right option (F2) gives 5 drops with a 70% chance. In the bottom example, a novel conflicting trial, where the left option (N1) gives 8 drops with a 50% chance and the right option (N2) gives 4 drops with a 70% chance.

The stimuli remained on the screen for up to 60s until the monkey responded by touching one of two infrared sensors. Then an action-outcome delay followed, the duration of which was drawn from a uniform distribution between 3.5 and 4.5 seconds, after which a visual cue indicated either a positive or negative outcome. In the first case juice was also delivered through a spout. The outcome cue lasted 3s regardless of the nature of the outcome (positive or negative), and then an inter-trial interval ensued, drawn from a uniform distribution between 5 and 7 seconds (figure 3.2). The inclusion of delays of this duration and variation made it possible to decorrelate the blood-oxygen-level dependent (BOLD) signal related to the decision and the motor response from the BOLD signal related to the subsequent feedback processing and juice consumption. This was possible because the hemodynamic response function (HRF) in the macaque is faster than in humans (Chau *et al.* 2015). The dissociation of neural correlates of decision- and motor-related processes, instead, was not temporal but relied on confound regressors included in the GLMs used for the neuroimaging analysis. Regressors indexing reaction time and choice side were convolved with the HRF, and additional choice side regressors were not convolved with the HRF (this captured changes in the BOLD signal related to field distortion at the actual moment the motor response was made).

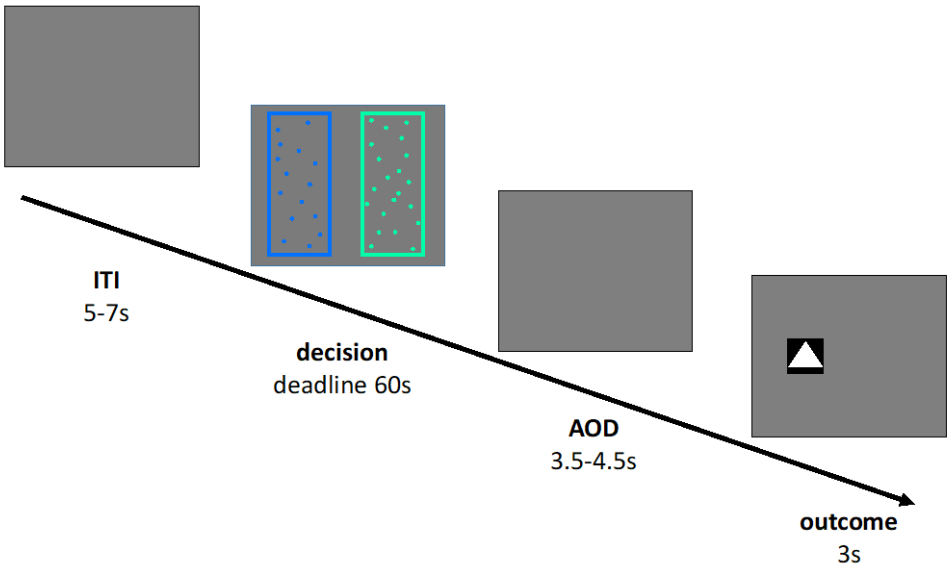


Figure 3.2 Choice task (experiment 1) timeline. ITI: inter-trial interval. AOD: action-outcome delay. The outcome event is either positive or negative depending on the chosen probability.

The stimuli were characterised by four parameters: reward magnitude of the option on the left hand side of the screen (magnitude left), reward probability of the option on the left hand side of the screen (probability left), reward magnitude of the option on the right hand side of the screen (magnitude right) and reward probability of the option on the right hand side of the screen (probability right).

As discussed in the general introduction (sections 1.3.1 and 1.3.3), the analysis is based on previous findings of neural activity that is correlated on a trial by trial basis with a key decision variable: the value of the option that will ultimately be chosen relative to the value of the alternative option that is rejected.

First, trials were classified as either familiar or novel, depending on whether the exact combinations of magnitude and probability in both options were the same as those experienced during the earlier extended training period, or if they were novel combinations. Secondly, both familiar and novel trials were divided into three conditions

(see figure 3.1, central panel, and table 3.1). In the first condition (coherent dimensions) one option was better than the other both in terms of probability and magnitude. In the second condition (identical dimension) the two options were identical in one dimension (they were associated with either the same reward magnitude or the same reward probability). In the third condition (conflicting dimensions) one option was the better one in one dimension but the other option was better in the other dimension. Using this 2 (novelty) x 3 (dimensionality) design, we created a different stimulus set for each test session for each subject, with 30 trials per condition, and therefore 180 trials in total per session.

choice / stimuli	coherent	one dimensional	conflicting
familiar	1	2	3
novel	4	5	6

Table 3.1 Conditions in the choice task (experiment 1)

In order to best decorrelate the signal related to the chosen option and the unchosen option, while avoiding confounding factors, we designed the stimulus set respecting a number of constraints. Before running the experiment, and because of the nature of the training, we could define value only in an objective sense, as the number of drops at stake multiplied by the probability of receiving them (Pascalian expected value). We chose the stimuli in such a way that the expected value sum (left option value plus right option value) was on average very similar for each of the six conditions (6±0.02 drops) and the correlation between value sum and absolute value difference was below 0.3 (Pearson's r) within each condition. It should, however, be noted that, by definition,

value difference cannot be kept the same across the three levels of dimensionality because coherent trials are associated with larger value difference than conflicting trials. However, within any given level of dimensionality it was possible to match value difference across the familiar and novel conditions: we designed our stimulus set in such a way that both the mean and the variance of the expected value difference would be very similar (<0.02 drops and <0.035 drops^2 respectively) for each pair of novel and familiar conditions (coherent; one-dimensional; conflicting). Additionally, we took care that the correlation between the value of the best option and the value of the worst option was smaller than 0.33 for all familiar trials and also all novel trials. Notice, however, that because of the nature of the value space, it was not possible to keep the correlation between best and worst option value equally low within each condition: for conflicting conditions, and in particular the familiar conflicting condition, the correlation was much higher. The trials belonging to the six conditions were randomly intermixed, with no correlation between condition and trial number.

The careful selection of the stimulus set allowed us to dissociate neural signals related to the values of the chosen and the unchosen options, as well as signals related to the decision variable value difference as opposed to the value sum. Because values and value differences were matched across novel and familiar conditions, it was possible to be confident that any differences in decision variable-related signals in the familiar and novel conditions were meaningful.

3.2.2 Head-post surgery

Four male Rhesus macaques were trained to perform choices between screen-presented stimuli, while sitting horizontally in a chair, as described in chapter 2, section 2.2.2. To

restrain their head movement during fMRI, a head post was surgically implanted prior to the training specific to this experiment.

Before any surgeries monkeys were treated with a steroidal anti-inflammatory [20mg/kg methylprednisolone injected intramuscularly (i.m.)] and an antibiotic (8.75 mg/kg amoxicillin, i.m.) a minimum of 12 hours prior to surgery so as to reduce the risk of postoperative infection, inflammation or oedema. During surgery, extra steroidal supplements were provided at 4 to 6 hour intervals if necessary. On the morning of the surgery, monkeys were sedated with ketamine (10 mg/kg, i.m.) and xylazine (0.5 mg/kg, i.m.) and were injected with non-steroidal anti-inflammatory agents (0.2 mg/kg meloxicam), atropine (0.05 mg/kg), and opioid (0.01 mg/kg buprenorphine). These were provided to reduce secretions and provide analgesia. They were further treated with a histamine H2 receptor antagonist (1 mg/kg ranitidine) for gastric ulceration protection due to the administration of both steroidal and non-steroidal anti-inflammatory treatments. Animals were then moved to the operating theatre where they were intubated, switched onto sevoflurane inhalational anaesthesia and placed in a head holder. Their head was shaved and cleaned using alcohol and antimicrobial scrub (chlorhexidine). Throughout surgeries, respiration rate, heart rate, body temperature, blood pressure and expired CO2 were continuously monitored. The surgeries were carried out under sterile conditions. The head posts implanted were MRI compatible head posts (Rogue Research). A midline incision was made, the tissue was retracted in anatomical layers and the head post fixed with dental cement and Thomas Recording ceramic screws. After a recovery period of at least six weeks (to allow bone strengthening around the implant), the animals were trained to perform the task inside the actual MRI scanner under head fixation. All surgeries were performed by two members of the laboratory with experience in monkey surgery.

3.2.3 MRI setup and data acquisition

All monkeys were initially trained in the familiar version of the task in a purpose-built wooden mock-scanner, reproducing most features of a real scanner, including the noise. Performance of all monkeys dropped when first transitioning to the real MRI scanner but it improved again after further practice. By the beginning of testing, all monkeys had had similar experience with the familiar task (13300±500 trials), but only monkey D had not previously participated in an fMRI experiment, therefore he was moved to the real scanner at an earlier stage of the training to allow for better familiarisation.

Structural and functional magnetic resonance images (MRIs) were acquired with a conventional 3T scanner, using a radial transmission coil and a four-channel phased-array receiver coil (Windmiller Kolster Scientific). The chair was positioned on the sliding bed of the scanner and transmitting and receiving coils were positioned closely around the head of the animal, while it was fixated using a head post holder. The MRI-compatible chair, head post, and head post holder were produced by Rogue Research. The same sensor box and juice delivery system as in the mock scanner were also used in the MRI scanner (described in chapter 2 section 2.2.2). A plastic screen (MRC Cambridge) was positioned 30 cm in front of the head of the animal, while the images were projected from the opposite side, through a hole, by a LX400 projector (Christie Digital Systems) with a custom-made lens located outside the MRI room.

For each monkey, structural images were acquired under general anaesthesia prior to the experiment, using a T1 MP-RAGE sequence with a resolution of 0.5x0.5x0.5mm^3, repetition time (TR) = 2.05s, echo time (TE) = 4.04ms, inversion pulse time (TI) = 1.1s, and flip angle of 8˚. Anaesthesia was induced using intramuscular injection of ketamine (10 mg/kg), xylazine (0.125–0.25 mg/kg), and midazolam (0.1 mg/kg). Macaques also received injections of atropine (0.05 mg/kg, i.m.), meloxicam (0.2 mg/kg, i.v.), and

ranitidine (0.05 mg/kg, i.v.). Local anaesthetic (5% lidocaine/prilocaine cream and 2.5% bupivacaine injected subcutaneously around the ears to block peripheral nerve stimulation) was also used at least 15 minutes before placing the macaque in the stereotaxic frame. The anesthetized animals were placed in an MRI-compatible stereotactic frame (Crist Instruments) in a sphinx position before being put into the bore of the MRI scanner. Scanning commenced ~2 h after induction. The depth of anaesthesia was assessed using physiological parameters (heart rate and blood pressure, as well as clinical checks before the scan for muscle relaxation). Macaques were maintained with intermittent positive pressure ventilation to ensure a constant respiration rate. Respiration rate, inspired and expired CO2, and inspired and expired isoflurane concentration were monitored and recorded using VitalMonitor software (Vetronic Services Ltd.). In addition to these parameters, core temperature and SpO2 were monitored throughout the scan. For monkeys A and B, three structural images were averaged whilst for monkeys C and D two structural images were averaged.

During the behavioural experiment, fMRI data were collected with a gradient-echo T2* echo planar imaging (EPI) sequence with a resolution of 1.5x1.5x1.5mm^3, repetition time TR=2.28s, echo time TE=30ms and flip angle of 90˚. At the end of each session, to help signal pre-processing, a proton-density-weighted image was acquired using a gradient-refocused-echo (GRE) sequence with a resolution of 1.5x1.5x1.5mm^3, repetition time TR=10ms, echo time TE=2.52ms and flip angle of 25˚.

As described in chapter 2, monkeys performed twelve choice sessions, interleaved by five single-option sessions belonging to experiment 2. Monkeys C and D occasionally stopped working during the experiment: if they restarted within approximately 15 minutes, the fMRI acquisition was continued, otherwise it was stopped and the cognitive task restarted from the beginning after a break. No separate runs were merged for data

analysis. In experiment 1, retained scans had an average duration of 43 min (σ=2 min) for monkeys A and B, and 88 min (σ=15 min) for monkeys C and D.

All procedures were conducted under a Home Office licence according to the Animal (Scientific Procedures) Act 1986 and the European Union guidelines (EU Directive 2010/63/EU).

3.2.4 fMRI data pre-processing

Preprocessing and analysis used the FMRIB Software Library (FSL) (Jenkinson *et al*. 2012). The T1-weighted images were processed in an iterative fashion cycling through brain-extraction (BET) (Smith, 2002), RF bias-field correction, and linear and non-linear registration (FLIRT and FNIRT) (Jenkinson, Smith, 2001; Jenkinson *et al.* 2002) to the *Macaca mulatta* F99 atlas (van Essen, 2002).

A template specific to the group of subjects was constructed with Advanced Normalisation Tools (ANTs) by registering each individual's structural to the F99 atlas, averaging them to obtain a first-order reference, then again registering each individual to this reference and averaging a second time. All coordinates reported in this thesis refer to our template in F99 space.

A script was developed to automatically select the best of two or three proton-density-weighted images acquired at the end of each session, based on minimisation of out-of-the-head ghosting artefacts. The selected images were used to help the reconstruction of the T2* EPI image by an offline-SENSE reconstruction method (Windmiller Kolster Scientific, Kolster *et al.* 2009).

Motion correction was performed by a custom-built script, written in the research group, to reduce as much as possible the artefacts caused by body movements (such as hand responses as well as juice drinking and other task-unrelated movements) while the

head itself remained fixed. Body movements distort the magnetic field that should ideally be homogenous in the field of view (FOV) and therefore they bias the reconstruction of the signal source, which is based on a quickly changing field gradient. The effect is that each slice of the image may appear as shifted in the phase-encoding (y) direction depending on the amount of field distortion present at the time of the slice acquisition. Volumes were acquired by interleaved slices, which often led to an apparent misalignment of odd versus even slices. To compensate for that, high-quality volumes with minimal distortion were selected from the timeseries and averaged to serve as a reference. Then each slice of each volume was registered to the reference, first linearly, then non-linearly, to correct for distortions in the phase-encoding direction induced by motion. This newly developed software (Lennart Verhagen, personal communication) gave excellent results compared to previous versions; however, one limitation that could not be overcome is that when an image slice is stretched in the y direction, the same amount of signal is spread over a larger number of voxels, decreasing the signal. A correct motion correction assigns the signal to its real source, by squeezing the voxels, but still cannot compensate for the intensity change.

Finally, the functional image was non-linearly registered to the high-resolution structural reference of each subject, then to the group-specific template using ANTs and the two transformation warps were concatenated. The image was then brain-extracted based on the high-resolution structural, spatially smoothed (3mm, full-width-half-maximum) and temporally high-pass filtered (cut-off 100s). After running the first-level whole-brain analysis (described below) on the motion-corrected low-resolution images in their original acquisition space, the statistical images obtained for each contrast were transformed from acquisition to standard space with a single resampling using the concatenated warps and ANTs.

3.2.5 fMRI whole-brain data analysis

A univariate general linear model (GLM) approach was taken for the statistical analysis of the whole brain functional data using FEAT (FMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl, Beckmann *et al.* 2003). Additional regressors were included beyond the task-related regressors of interest: the task-related confounds and the task-unrelated confounds. The task-related regressors of interest are:

- Six constant regressors representing the main effect of decision as a boxcar of amplitude 1, with onset 600ms before the response and a duration of 600ms. One regressor was specific to each of the three levels of dimensionality in both the familiar choice trials and the novel choice trials (i.e. six conditions in total).
- Twelve parametric regressors representing the value difference (chosen minus unchosen option value) and value sum (chosen plus unchosen option value) in each of the six conditions. Value difference is a positive number if the subject chooses the best option and negative otherwise. These two sets of six parametric regressors were aligned to the respective constant regressors, starting 600ms before the response and ending at the response time.

There were eleven task-related confound regressors:

- The logarithm of reaction time (logRT) and the time since the beginning of the session were parametric regressors temporally aligned with the other decision-related regressors of interest.
- The response side, left or right, coded as a binary variable (±1), was also temporally aligned to the other decision regressors.

- A constant regressors modelled the onset of trials where no response was given within 60 seconds. It was shaped as a boxcar of amplitude 1 and duration 600ms starting at the stimulus onset.
- Two constant regressors denoted trials with either a positive or negative outcome. These regressors started at the time of the outcome and had an amplitude of 1 and a duration of 1 second.
- Two more parametric regressors represented the amount of reward delivered (between 1 and 10 drops of juice) and the reward prediction error (equal to the actual reward minus the average expectation given the chosen option); they were temporally aligned with the constant outcome regressors (from the outcome onset for 1s).

All parametric regressors (namely value difference and value sum, as well as logRT, time in the task, side of the choice, reward amount and reward prediction error) were normalised. The GLM included additional regressors indexing the temporal derivative of all the regressors described so far. All these regressors were convolved with an HRF implemented as a gamma function with standard deviation of 1.5s and mean lag of 3s. The HRF models the latency of the change in the relative amounts of oxyhaemoglobin and deoxyhaemoglobin in the tissue in response to the change in neural activity. The HRF in macaque monkeys has been found to be faster than in humans (Chau *et al*., 2015) and therefore the HRF employed shorter temporal parameters that are typically used in investigations of human brain function. Three regressors modelled instant signal distortions not due to neural activity but to changes in the magnetic field caused by movement. These regressors were therefore not convolved with an HRF and their temporal derivatives were not added to the GLM.

- Two constant regressors indicated, separately, left choice and right choice, but importantly they started at the beginning of the TR when the response was recorded and had a duration of exactly one TR.
- A constant regressor modelled the juice consumption and started at the beginning of the TR when juice delivery started and terminated at the end of the TR when the juice delivery ended (for a total of one or two TRs).

These three regressors were not HRF-convolved.

As mentioned above, the regressors related to the decision process, were aligned to 600ms before the response. This allowed us to capture decision-related activity even though we allowed a response time (RT) up to 60s in order to increase the session completion rate; in practice 10% of trials had an RT above 3s and 5% of trials had an RT above 15s; this was almost entirely due to monkeys C and D. Therefore, had we aligned the regressors to the stimulus onset, we would not have captured the decision-related neural activity in some very slow RT trials. The possibility of excluding those trials was also undesirable, because slow RT was correlated with task parameters such as familiarity, value sum, and value difference, so exclusions would have unbalanced the task design. On the other hand, a preliminary test on the data from only animals A and B (less than 0.8% of RTs above 3s) showed that the neural signature of the regressors of interest when aligned to the stimulus onset and when aligned to 600ms before the response was highly similar. Henceforth, the latter approach was taken.

Task-unrelated confounds can be grouped in two sets:

- 13 motion regressors were added to the GLM, representing the motion components as estimated by a principal component analysis (PCA) during the preprocessing.
- Low quality volumes were excluded from the analysis; volume quality was assessed based on 1) slice-registration cost (the normalized correlation between the current

volume and the robust average after optimal registration), 2) linear scaling along the phase-encoding direction (directly related to signal intensity loss due to motion distortion), 3) non-linear deformation (penalizing volumes that require highly non-linear deformations). Volumes further than 2.5 standard deviations from the average of all non-sleeping volumes, in any of the three metrics, were flagged as low quality; "sleeping volumes" were defined by an absence of task-related behaviour. Overall, approximately 5% of the non-sleeping volumes were excluded. For each volume to be excluded, a specific regressor was added to the GLM, equal to zero at all timepoints bar the flagged volume. This procedure increased the size of the GLMs by on average 60 regressors and was considerably time-consuming, but preliminary testing showed that it increased the quality of the final statistical results, as assessed in task-related regressors other than the regressors of interest.

Figures 3.3 and 3.4 illustrate the correlations of the task-related and all confound regressors, respectively, averaged across sessions; correlations vary only slightly across sessions depending on stimuli and animals' behaviour. Decision, value difference, and value sum terms are represented by six regressors each, one per condition (cf. table 3.1, section 3.2.1). In the second figure, each task-related regressor is followed by its temporal derivative.

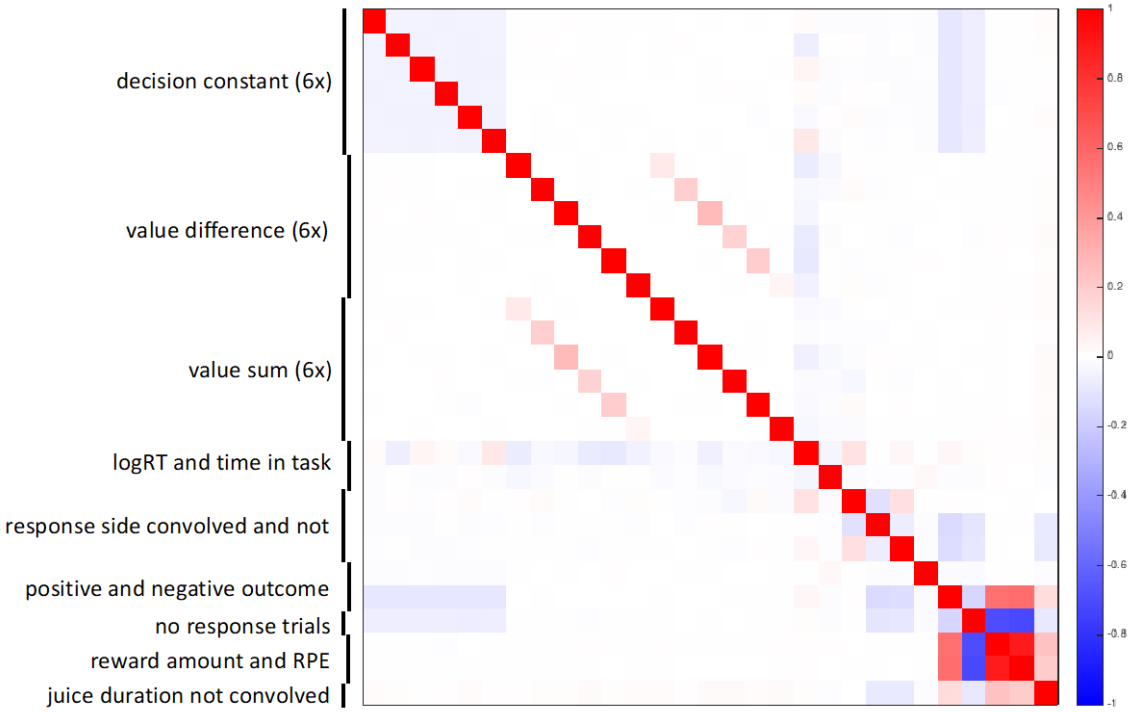


Figure 3.3 Average correlations among task-related regressors across all sessions, excluding time derivatives for visualisation purposes. For the value difference regressors all $r<0.14$.

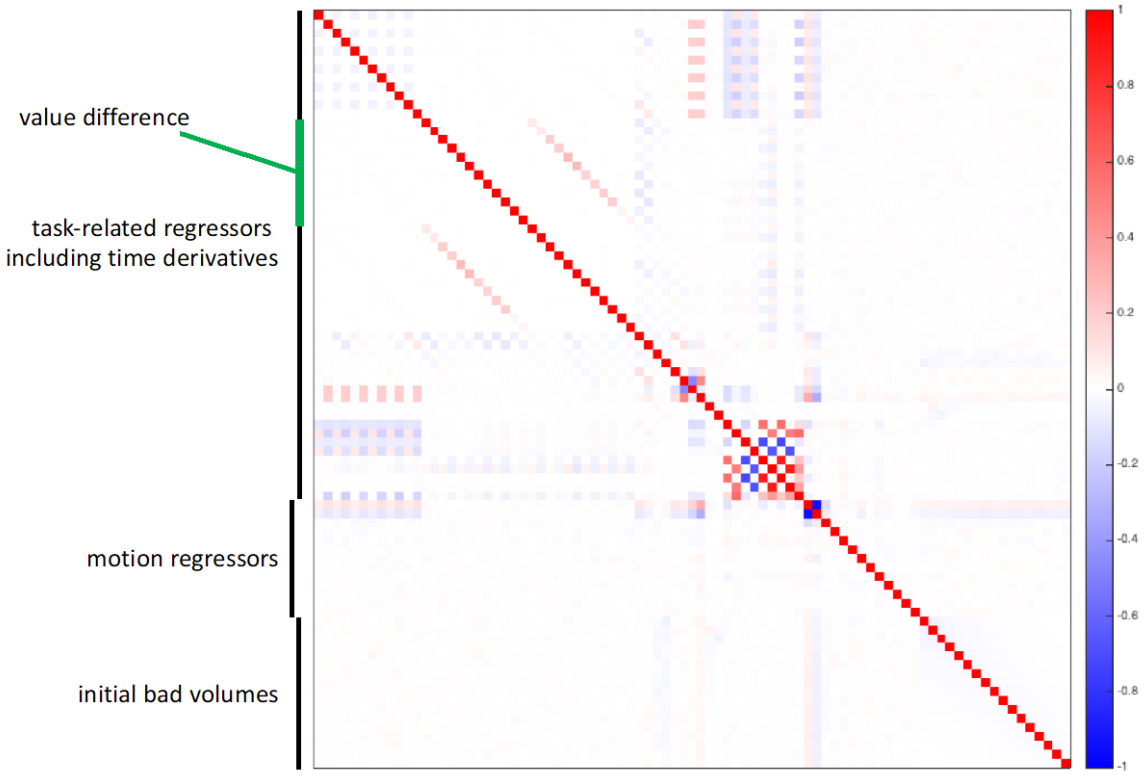


Figure 3.4 Average correlations among all regressors, including time derivatives and task-unrelated confounds. Only the first 17 low quality volume regressors are illustrated, but their number can vary between sessions.

Although there are correlations between confound regressors, the key regressors of interest, namely value sum and value difference, are largely decorrelated from all others. The average correlation between value difference and value sum regressors (averaging the absolute values) is r=0.14 and only in a few particular sessions and conditions reaches a maximum of r=0.54 (and so <0.30 of the variance between regressors is shared). There was some negative correlation between value difference and logRT, due to the fact that subjects responded faster to choices with a large value difference between options, but this was on average r=0.08 (averaging the absolute values) and never exceeded r=0.28 in any given session. The inclusion of the RT regressor may have reduced the size of the value difference effects we found; we took this conservative approach to make sure that the effects we report are entirely due to a key decision

variable (chosen minus unchosen stimulus value) and not to differences in motor processes related to RT. Correlations of value difference with all other regressors were $r<0.06$ on average, and $r<0.42$ in any given session for any combination of regressors.

Each of the 48 sessions was analysed at the first-level using the FEAT tool in FSL, which computed parameter estimates (PEs) representing the beta values for each regressor, and contrasts of parameter estimates (COPEs) for each contrast, as well as variance estimates (VARCOPEs). The COPEs and VARCOPEs were then transformed into standard space using ANTs and given as input to FEAT again for a higher-level mixed effects analysis with FLAME 1 (FMRIB's Local Analysis of Mixed Effects, stage 1, Woolrich *et al.* 2004; Woolrich 2008).

The higher-level analysis used two contrasts: the "mean" contrast weighted across all 48 sessions equally, effectively averaging them, and the "linear trend" contrast weighted negatively the early sessions of each subject and positively the late sessions (weights going from -5.5 to +5.5 in steps of 1) and was therefore able to detect whether any first-level contrast varied linearly with practice in the task. Sessions were not distinguished by subject at the second level of analysis; although in human fMRI protocols it is usual to adopt a three-level approach if subjects perform multiple sessions, the number of subjects typically used in primate research, limited for practical and ethical reasons, meant that a three-level approach would not have provided sufficient power to detect effects of interest.

Results are reported as Z (Gaussianised T) statistic images, and clusters were determined by a threshold of $Z>2.3$ and cluster-corrected significance threshold of $p=0.05$

3.2.6 fMRI ROI-based data analysis and timecourse analysis

Regions of interest (ROIs) were defined as spheres of 3mm of radius, centred at the peak of the activation for a given contrast. In the case of one OFC activity pattern found that had a bimodal shape, two adjacent spheres were combined. ROIs were also cut to follow the brain contour so as not to include any non-brain voxel.

For the individual subject ROI analysis, the beta values (PEs) for the regressor of interest were averaged across voxels and then across sessions for each subject. The mean is displayed in a bar plot; error bars represent the standard error across sessions.

For timeseries analysis, the ROIs defined in standard space were warped back into the acquisition space of each separate session. The filtered timeseries of each voxel within the ROI were averaged and normalised, then up-sampled (spline interpolation) by a factor of 10. Task-unrelated confounds (described in the previous section) were regressed out of the timeseries and the residuals were then aligned to the onset of each trial and a GLM was fit at each timestep using ordinary least squares (OLS). The timecourse GLMs were similar to the ones described in the previous section for the whole-brain analysis, but focussed on decision-related factors and did not model outcome events. This was not necessary because regressions were run at each time step separately, across trials. The GLM included 21 regressors: six constant, six value difference and six value sum regressors, one for each of the 6 conditions (cf. table 3.1), additionally, logarithm of the reaction time, time in the task and side of the choice were modelled. Beta coefficients for the regressors, or contrasts of betas, were therefore estimated at each timepoint and then averaged across the 48 sessions.

It should be noted that for the whole-brain analysis, because the regressors are HRF convolved and they include time derivatives, the exact timing of regressor onsets with respect to task events is relatively unimportant and so we chose to align the regressors to

600ms before the response in order to maximize the total number of trials included in the analysis. On the other hand, in the timecourse analysis the exact timing is essential and if regressor onsets and trials are misaligned then effects may be averaged out. As a consequence, we aligned the timecourse to the stimulus onset and discarded trials with RT>3s (10%); to compensate for this, the beta estimates for each session were weighted by the number of trials included in the analysis.

An additional analysis was run to discriminate the contribution of the value of the chosen stimulus and the value of the unchosen stimulus. In this analysis, the six value difference regressors and the six value sum regressors were removed and replaced with chosen value and unchosen value, for the same conditions. All the confound regressors were kept the same.

Because the ROIs for the timecourses were defined around the peak effects at the whole-brain level for the same GLM, the extent of the parametric modulation illustrated in the timecourse figures may have been inflated. To avoid double-dipping a leave-one-out procedure could be used in order to select a specific ROI for each session based on all other sessions bar that one. However, the timecourse figures were included not only to illustrate the temporal evolution of the effect, but also importantly to illustrate the signs of the value difference modulations in familiar and novel trials separately, as well as their components: the chosen value and unchosen value modulations.

3.3 Results

As reported in the previous chapter, the macaque monkeys readily adapted to the novel testing trials, making meaningful choices among new stimuli by taking into account both their magnitude and their probability with approximately the same weight, and reaching a

high level of performance from the first session onward. As already mentioned computational modelling identified biases in the decision process, in particular the tendency to add rather than multiply dimensions and prospect theory-like distortions in each dimension; these biases were consistent across all monkeys and stronger for familiar than novel trials. Because the subjective distortions of representations of value may or may not match the activity in different brain regions and because subjective distortions were less apparent in novel trials, objective choice values were used in the initial analysis of neural activity. The results were subsequently compared with the results obtained when the same analyses were run but with subjective rather than objective values.

Table 3.2 lists all the analyses that were run and are reported in this section; the three main regressors that we discuss are the constant regressor, representing the average effect of decision versus rest, the value difference regressor, representing the key decision variable of chosen minus unchosen value, and the value sum regressor, representing the combined value of both available options. The beta weights of these regressors were either averaged across all conditions, novel and familiar (left column of table 3.2), or contrasted, i.e. novel minus familiar (right column of table 3.2). Timecourses of some of these effects were measured. Further analyses contrasted specifically the value difference regressors in the conflicting conditions. We then tested how some of these effects change with experience across the twelve sessions of experiment 1. Finally, these analyses were repeated using a subjective definition of value.

The key results that are statistically significant at a standard criterion (cluster corrected with $Z>2.3$, $p<0.05$) are highlighted in green. Other results are either not significant, or significant at different thresholds, and are reported because they may help giving a more complete picture of the neural processes engaged by the present task. In some cases, given the constraints of our design involving monkey fMRI, the statistical

power may have been limited. However, the specific anatomical location of some subthreshold results reported here is suggestive of genuine activations consistent with previous knowledge of primate brain function.

Choice-related effects	Novelty-specific effects	Figures
Decision main effect – all trials	Decision main effect – novel vs familiar	3.5
Value difference – all trials	Value difference – novel vs familiar	3.6 3.8
Value difference effects, split by subject, in OFC and AMFC		3.10
Chosen value – all trials	Chosen value – novel vs familiar	3.11
Timecourse of value difference and chosen value – all trials, in OFC		3.7
Timecourse of value difference and sum, chosen and unchosen value – all trials, in AMFC	Timecourse of value difference – novel versus familiar, in AMFC	3.9
	Value sum – novel vs familiar	3.12
Conflicting vs other conditions	Novel conflicting vs familiar conflicting	3.13
	Trend across sessions of value difference novel vs familiar	3.14
Subjective value: confirmation of all previous analyses		
Trend across sessions of subjective value difference in familiar and all trials		3.15

Table 3.2 List of all experiment 1 neuroimaging analyses

The GLM revealed a positive main effect of decision on the BOLD signal throughout large swathes of sensorimotor cortex in frontal and parietal cortex, as well as ACC and dlPFC (figure 3.5, top). The contrast of novel versus familiar conditions remained below the threshold for cluster correction, but revealed a tendency towards a stronger activation

for novel trials at decision time in areas including bilateral dlPFC (figure 3.5, bottom), visual cortex, superior parietal lobule (area PE) and left superior temporal sulcus (STS).

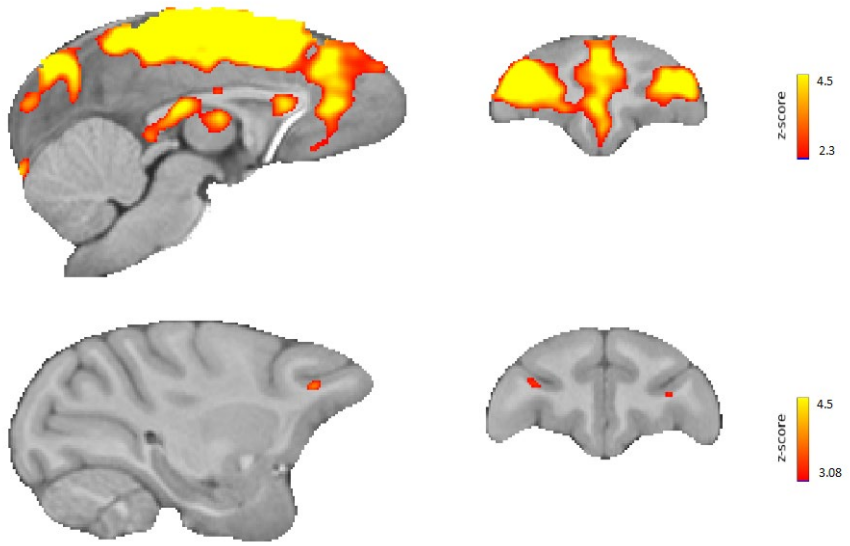


Figure 3.5 Main effect of decision. **Top:** average of all conditions, Z>2.3 cluster p<0.05. **Bottom:** Contrast of novel versus familiar decision, Z>3.08, p<0.001 uncorrected. This and all following brain images are displayed in radiological orientation (the left hemisphere is printed to the right and vice-versa).

The main effect of decision versus rest is a very coarse contrast that might capture activity related to the taking of the decision but which might also capture activity related to perception and to the initiation of action. We further hypothesised that regions specifically encoding the value of the options available, or comparing them, may show a trial-by-trial parametric signal at the time of the decision, proportional to the difference in value between the option ultimately chosen and the alternative one.

We observed that when both familiar and novel trials were included, the only region in prefrontal cortex showing such a parametric effect of value difference after whole-brain cluster correction was left OFC, area 13, extending from the medial to the lateral orbital sulci (figure 3.6, medial peak Z=-3.99, coordinates [-10 15 2], lateral peak Z=-3.89, coordinates [-15 16 4]). This effect was negative, meaning that the overall neural

activity increased whenever the two options were closer in value and the decision was more difficult, and activity decreased when the options were more different and the decision was easier. Given the elongated shape and the bimodal amplitude of the OFC cluster, a double-sphere ROI was created, centred on the two peaks, as illustrated in figure 3.6.

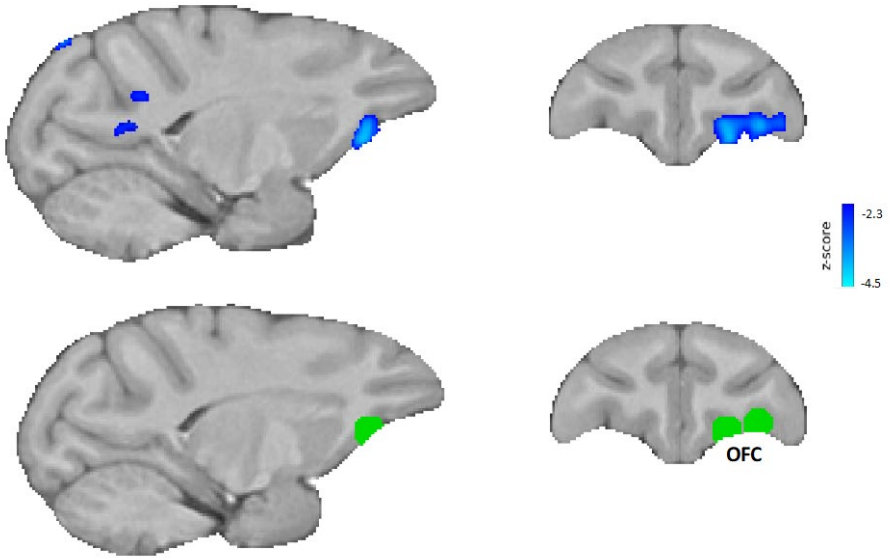


Figure 3.6 Top: effect of chosen value minus unchosen value across all conditions, whole-brain corrected with cluster threshold $Z>2.3$, $p<0.05$. **Bottom:** OFC bimodal ROI in green, defined by two spheres of 3mm radius, centred on the peaks of the value difference effect.

Focusing on such OFC ROI, a consistent negative effect of chosen minus unchosen value was revealed in each subject (see methods, section 3.2.6). The timecourse of the BOLD signal in OFC was investigated, averaging across all 6 choice conditions, but no statistical test was based on the timecourse analyses here and later in this chapter. The timecourse plot confirmed a parametric negative effect of value difference. Additionally, studying the separate contribution of the chosen option value and the unchosen option

value, it was observed that the signal becomes more negative when choosing high value stimuli and slightly more positive for high value unchosen stimuli, and while the unchosen effect is transient, the negative chosen effect appears at the beginning of the trial and continues to build up until after the outcome event (figure 3.7).



Figure 3.7 Analysis of the OFC ROI. **Left:** average effect in each subject of chosen minus unchosen value. **Centre:** timecourse of the parametric regressors linked to value difference. **Right:** timecourse of the chosen value and unchosen value effects separately. Dotted vertical lines represent stimulus onset and average outcome onset. The grey shading represents ±1 standard deviation in the outcome time. Coloured shadings represent the standard error in the timecourse estimate across sessions.

The value difference effect in OFC was present for both familiar and novel trials, and stronger for novel trials. However, the difference was not significant, $t(143)=1.46$, $p>0.05$.

Because the OFC effect was lateralised in the left hemisphere, a symmetrical ROI in the right hemisphere was created to investigate how it was modulated by value difference. While the average z-score within the left ROI mask was -2.51, in the right ROI the parametric effect was considerably weaker although similar in sign (average z-score of -0.87), with a significant difference between sides $t(47)=3.15$, $p<0.01$.

Next, we contrasted the beta coefficients for value difference in novel and familiar trials. A significant difference between these two conditions was found in ventral striatum extending to left OFC (peak Z=3.71, coordinates [-11, 8, 4]), as well as in dlPFC (peak Z=3.39, coordinates [-15, 20, 14]); the largest peak of activation, however, was found in the most anterior part of area 32, at the border of perigenual ACC, vmPFC, and the medial frontal pole (figure 3.8, peak Z=3.58, coordinates [0 24 7]). As discussed in the introduction (section 1.3.3), we will refer to this area as anterior medial frontal cortex (AMFC). This strong effect for the contrast between the BOLD responses to variations in the decision variable between novel and familiar trials, clearly indicates that different neurocognitive mechanisms are recruited for familiar and novel decisions. This holds true, even if the average value sum, value difference and dimensionality of the choices were matched between the familiar and novel conditions.

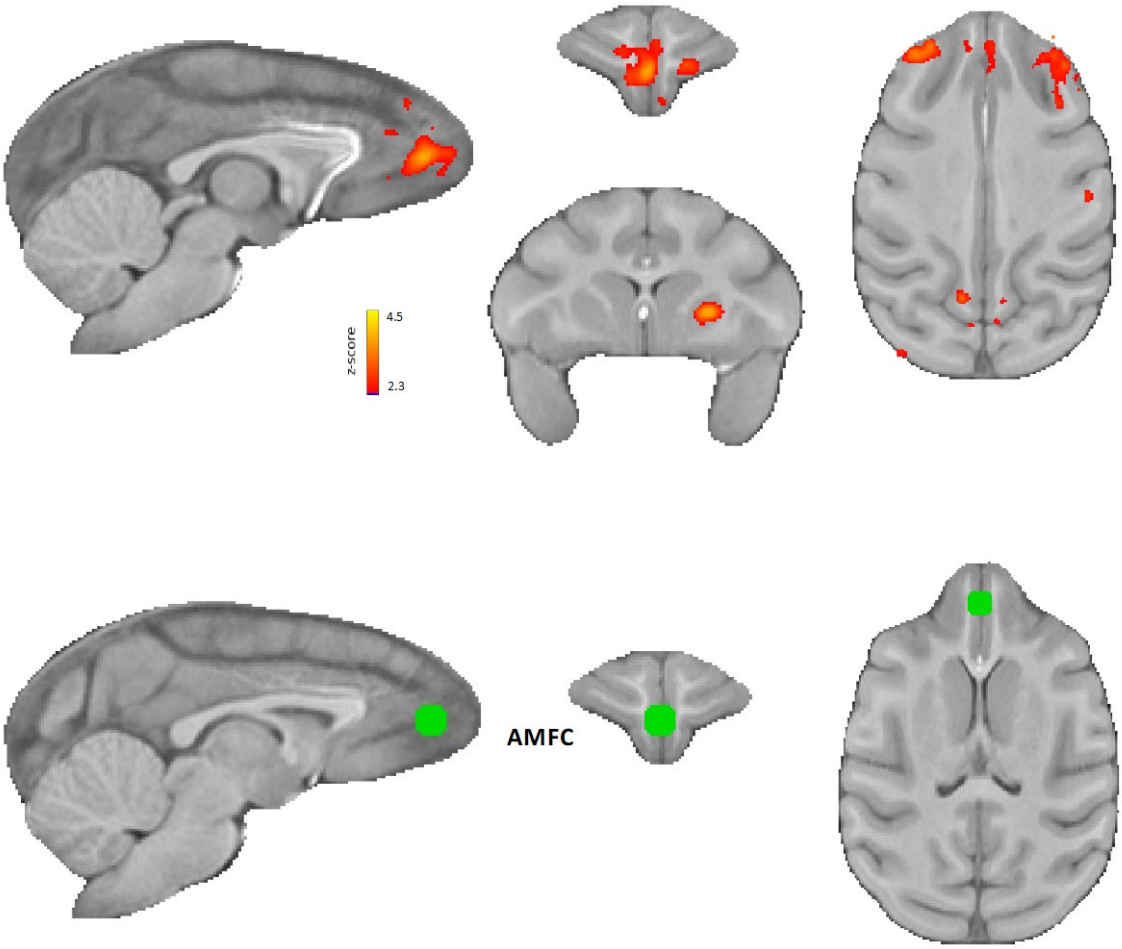


Figure 3.8 Contrast of value difference effects in novel versus familiar choice, whole-brain cluster-corrected ($Z>2.3$, $p<0.05$), except in the right panel. **Top left:** AMFC. **Top centre:** AMFC and ventral striatum. **Top right:** bilateral dlPFC. The dlPFC effect survives cluster correction only in the left hemisphere, but an uncorrected axial image is shown to illustrate the bilateral effect. **Bottom:** AMFC ROI with 3mm radius centred on coordinates [0 24 7].

To further investigate the role of AMFC in novel and familiar decision-making, we created an ROI centred on the AMFC effect peak (figure 3.8, bottom) and ran a timecourse analysis in such ROI, for novel and familiar trials separately. Unlike OFC, AMFC showed a clear difference in activity between familiar and novel trials. In familiar trials, value difference (green in figure 3.9) modulated positively the neural activity during and following decision, meaning that the area was more active for trials where the

chosen option had a high value compared to the unchosen option. On the other hand, in novel trials the modulation was negative, meaning that neural activity increased when the two options were more similar to each other and the choice was more difficult. As a consequence, the contrast between conditions was negative during the decision, but turned positive after the outcome.

It was also found that the effect of the two options' combined value (value sum, purple in figure 3.9) was approximately similar in the familiar and novel conditions, and modulated positively AMFC activity only at the earlier stages of the decision, before the peak of the value difference effect.

Disentangling the relative contributions of the value of the chosen option and the unchosen option, revealed that unchosen value (light blue in figure 3.9) was modulating activity initially positively and later on the effect turned negative, at a time potentially corresponding to the action-outcome delay. On the other hand, the value of the chosen option (red in figure 3.9) had different effects on the BOLD signal timecourse between conditions which were consistent with the profile of the value difference regressor. In familiar trials the chosen effect was transiently positive and then null, whereas in novel trials the effect was negative during the action-outcome delay and then positive again after the outcome.

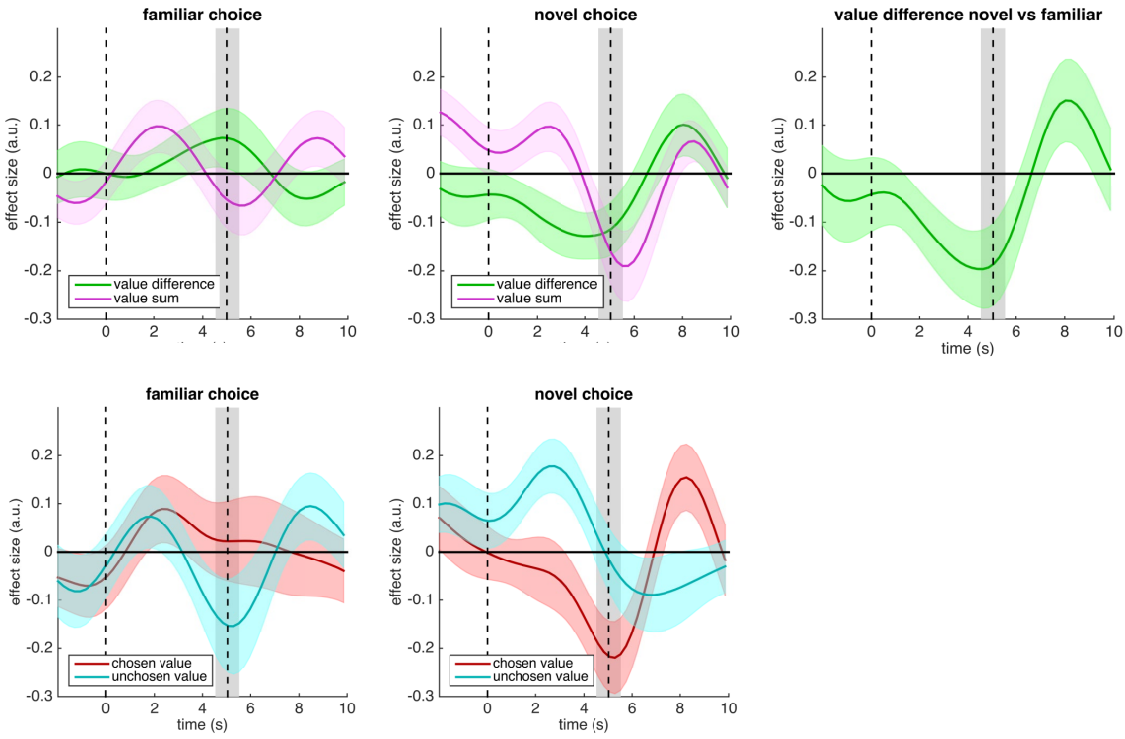


Figure 3.9 Timecourses of the value-related regressors in AMFC. **Top left:** value difference and value sum in familiar trials. **Top centre:** value difference and value sum in novel trials. **Top right:** contrast of value difference in familiar and novel trials. **Bottom left:** chosen option and unchosen option value in familiar trials. **Bottom centre:** chosen option and unchosen option value in novel trials. Dotted vertical lines represent stimulus onset and average outcome onset. The grey shading represents ±1 standard deviation in the outcome time. Coloured shadings represent the standard error in the timecourse estimate across sessions.

By comparing the neural effects in OFC and AMFC of the 4 subjects separately, we found consistency in the value difference effect for all trials in OFC (figure 3.10, top left panel) and value difference in novel minus familiar trials in AMFC (figure 3.10 bottom right panel). However, there was no consistency in the size of novel-versus-familiar effects in OFC or value difference effects regardless of novelty in AMFC (figure 3.10, top right and bottom left panels).

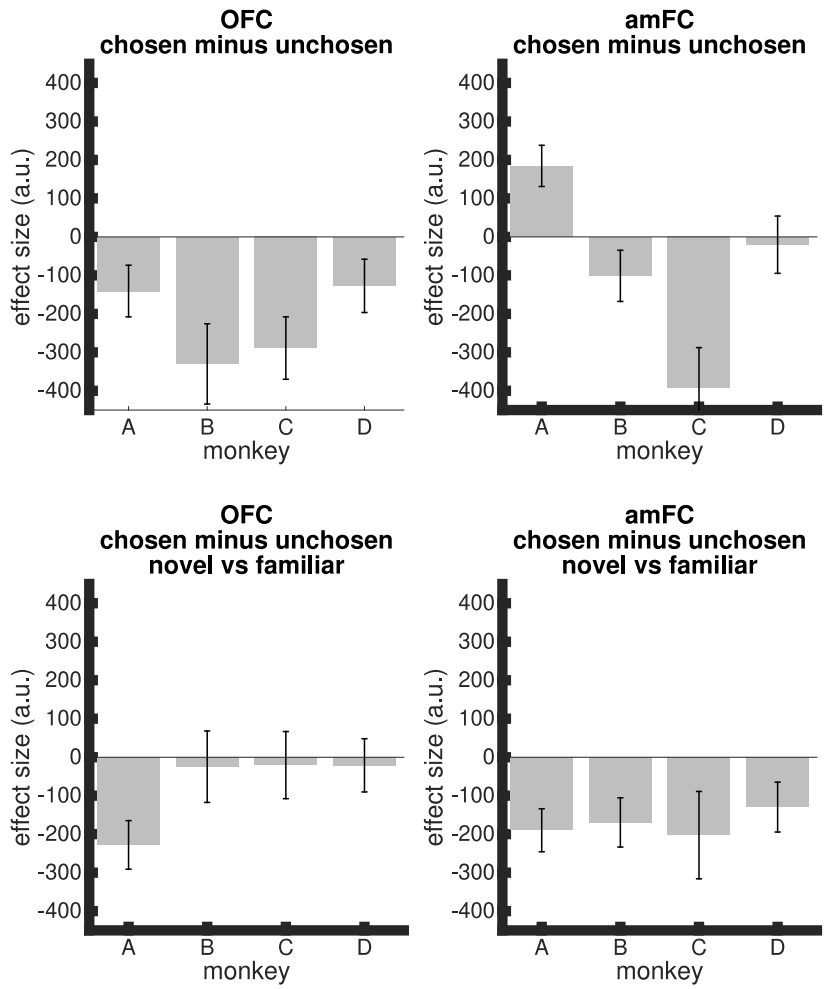


Figure 3.10 Individual subject contributions to value difference in all trials and in the novel versus familiar contrast, in OFC and AMFC.

A separate neuroimaging analysis was run, with a GLM including chosen value and unchosen value as parametric regressors of interest (but not including value difference and value sum). The chosen and unchosen value timecourses reported above were derived from this analysis. At the whole-brain level, significant clusters of activity were found, positively correlated with the value of the chosen option in dorsomedial prefrontal cortex (dmPFC) area 9, extending into area 10 near AMFC (figure 3.11, top, peak Z=3.65, coordinates [0 20 19]), as well as in right insula, basal ganglia and motor cortex. At the same time, no area displayed a significant negative effect of chosen value, nor any significant effect of unchosen value. Nevertheless, at a weaker threshold (Z>2, not

corrected), we observed a negative effect of chosen value in OFC, consistent with the timecourse results reported above (figure 3.7, centre) for the same GLM. Subthreshold effects of unchosen value were observed in the insula and the midbrain. A contrast of chosen value in familiar and novel trials revealed a difference in bilateral entorhinal cortex (figure 3.11, bottom, right peak Z=3.84, coordinates [10 -6 -15], left peak Z=3.63, coordinates [-7 -8 -16]), with a positive sign for chosen value in familiar trials and a negative sign for chosen value in novel trials; given the small size of the clusters, they did not survive multiple comparison correction.



Figure 3.11 Effect of chosen option value at the time of decision. **Top:** average of all conditions, whole-brain cluster-corrected, threshold Z>2.3, p<0.05. **Bottom:** contrast of familiar versus novel conditions, Z>3.08, p<0.001, uncorrected.

A further result of interest (figure 3.12) is that the combined value of both options (value sum) was found to be stronger for familiar compared to novel trials in the

midbrain, in a location anatomically consistent with the ventral tegmental area (VTA) and the substantia nigra.

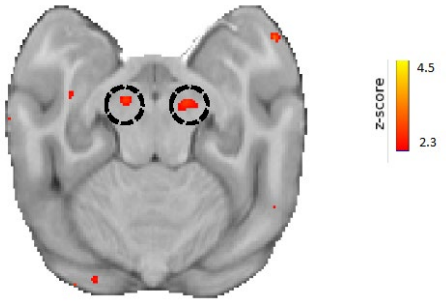


Figure 3.12 Areas more strongly activated by the total value of both stimuli in familiar compared to novel trials, Z>2.3, not corrected.

The results presented up to this point are based on either the average of all six conditions (familiar coherent, familiar one-dimensional, familiar conflicting, novel coherent, novel one-dimensional, novel conflicting) or on the contrast of the last three (novel) minus the first three (familiar). Conflicting conditions, in which each option is more valuable in one dimension but less valuable in the other, are particularly challenging and are the only ones where the integration of the two dimensions of value is crucial for making a correct choice. We hypothesised that specific neural mechanisms may be recruited for decision-making in the conflicting conditions, and these may differ between novel and familiar trials.

Therefore, a first contrast was created between conflicting and non-conflicting conditions, regardless of the novelty of the decision (conditions 3+6-1-2-4-5, as numbered in table 3.1, section 3.2.1). This contrast suggested that some modulation by task dimensionality may be present in the anterior part of vmPFC (peak Z=3.92), dorsal ACC, motor and visual cortex, but no effect survived cluster correction. Furthermore,

and more importantly, a contrast specifically between the novel conflicting and the familiar conflicting conditions (6 minus 3) was run; the result was a relatively shallow but large ventromedial cluster, which was significant only with a different voxel threshold (Z<1.96) but a standard cluster significance of p=0.05. The peak of such effect (Z=3.5, coordinates [-1 27 6]) was within the AMFC ROI, but the cluster extended ventrally and posteriorly through area 14m; the posterior end of the cluster corresponded to the PMFC ROI, defined below. A second cluster was identified in the visual cortex.

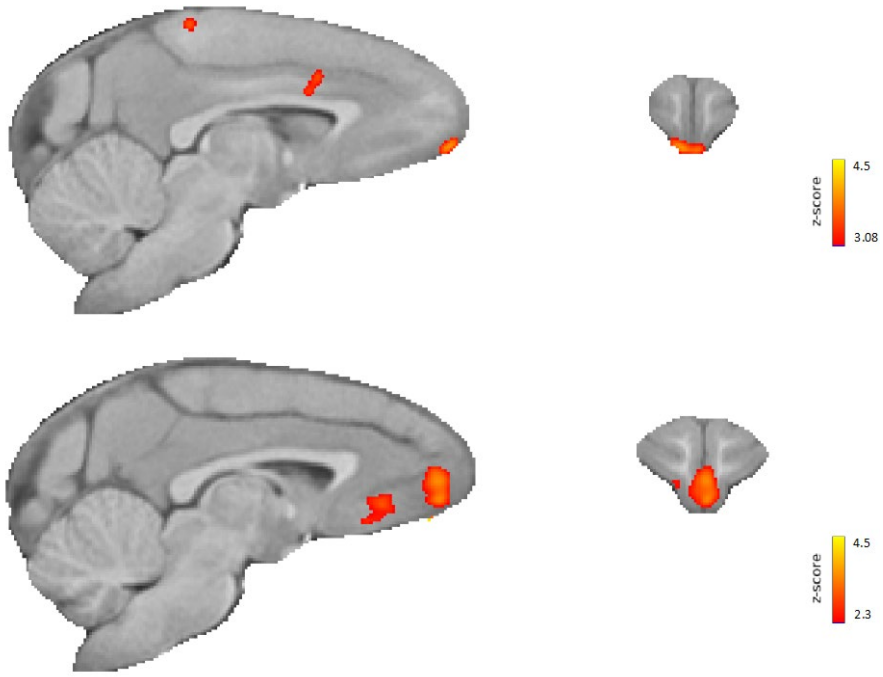


Figure 3.13 Analysis of conflicting conditions. **Top:** value difference in conflicting versus non-conflicting conditions, regardless of novelty (voxel threshold Z>3.08, p<0.001, uncorrected). **Bottom:** value difference in novel conflicting versus familiar conflicting conditions, cluster corrected (Z>1.96, p<0.05).

A further investigation addressed the evolution of the novel signal with time. It is possible that the effect of stimulus novelty may be short lived, therefore we tested whether the novel versus familiar value difference contrast changed linearly across the

twelve sessions of experiment 1. The higher-level linear trend contrast did not reveal any effect surviving correction for multiple comparisons, but activity in an area in posterior medial frontal cortex, later referred to as PMFC, showed a tendency towards an interaction between the effects of novelty (novel vs familiar contrast) and session number (peak Z=3.37, coordinates [-2 15 2]); additionally, the same tendency was present in bilateral globus pallidus, as depicted in figure 3.14. This effect was not significant at the whole-brain level, but was used to define an ROI in the location where the difference between novel and familiar trials changed most with experience across sessions, namely in PMFC. This ROI will be used as the stimulation target in the control condition in experiment 3 (see chapter 5). An ROI analysis confirmed that in amFC the difference between novel and familiar trials did not disappear in the timeframe of experiment 1.

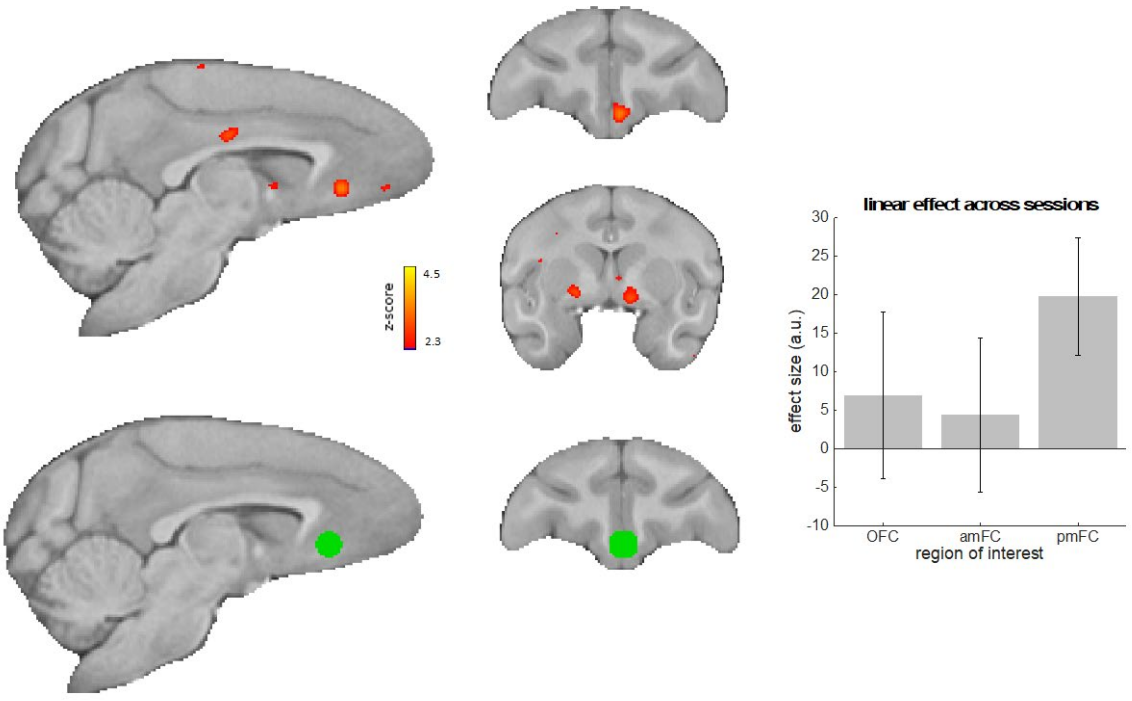


Figure 3.14 Linear effect of session number on the familiar vs novel value difference contrast. **Top left:** higher-level statistical map, effect thresholded at Z>2.3, uncorrected.

Bottom left: 3mm PMFC ROI. **Right:** comparison of the session number effect in three different ROIs. No effect was significant after cluster correction for multiple comparison.

All the results reported above were obtained using an objective definition of stimulus value, namely the amount of reward multiplied by the probability of receiving it (Pascalian expected value). However, as discussed in chapter 2, behavioural biases were observed in the decisions, which were consistent across subjects, but more pronounced in the familiar than the novel condition. A model for subjective value was used to construct subjective regressors (value sum and value difference) to be used in the fMRI analysis; the parameters of the subjective value model were fitted on behavioural choices in all six conditions at the same time.

The group-level whole-brain statistical results were qualitatively similar to the ones observed previously in the objective value analysis. In particular, as expected the main effect of decision, and the contrast novel versus familiar, were unchanged. The effect of value difference in OFC was qualitatively the same (i.e. negative in all conditions but more negative for novel than familiar trials). The value sum contrast between novel and familiar trials also revealed a qualitatively similar effect in the midbrain for subjective value.

On the other hand, the contrast of value difference in novel versus familiar trials was reduced in all three locations: AMFC, ventral striatum and principal sulcus, not surviving cluster correction. A further analysis showed that, while the (positive) effect of value difference in familiar trials remained very similar across subjective and objective value, the (negative) effect in novel trials was reduced for subjective value. Similarly, the contrast of value difference in conflicting novel versus conflicting familiar conditions was also reduced in vmPFC, and this was explained by an unchanged (positive) effect of

value difference in conflicting familiar choices but an absence of (negative) effect of value difference in conflicting novel choices.

Additionally, the subjective value analysis revealed a clear impact of exposure to the new multidimensional choice task in the temporal lobe. Firstly, the effect of value difference in hippocampus was significantly modulated by session number, across all conditions (peak Z=3.64, coordinates [-12 -3 -14]); secondly, the effect of value difference specific to familiar trials was significantly modulated in a close location and extending to the entorhinal cortex (peak Z=4.17, coordinates [-16 -4 -17]).

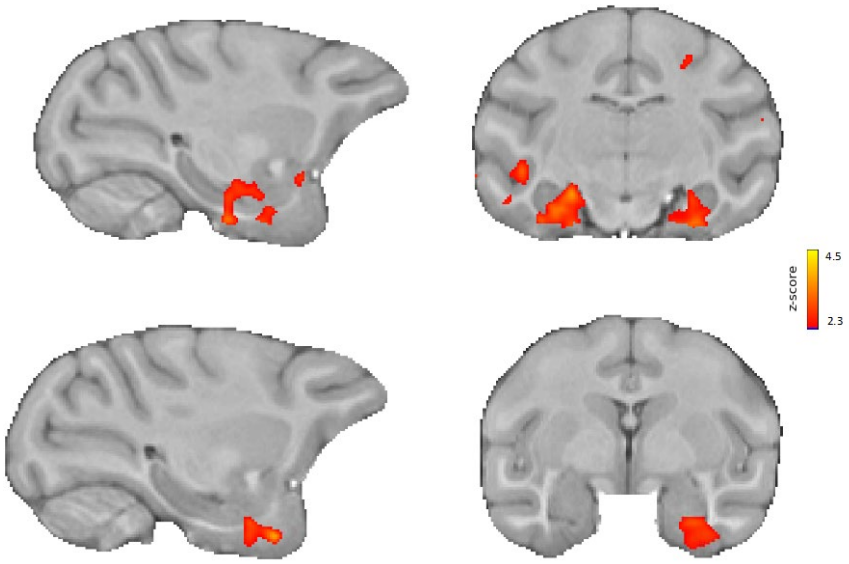


Figure 3.15. Changes in subjective value difference encoding with practice. **Top:** linear effect of session number on value difference in all conditions Z>2.3. The left hippocampus cluster survives correction for multiple-comparisons (p<0.05) but in the top right panel a non-corrected image is reported to illustrate the bilateral nature of the effect. **Bottom:** linear effect of session number specific to value difference in familiar trials (cluster-corrected Z>2.3, p<0.05).

These effects were also present in the objective value analysis, but below the multiple comparison correction threshold, whereas they were significant when including the

subjective bias in the value difference definition. These effects across time reveal that neural activity in the medial temporal lobe was modulated as a function of subjective value difference in the first few sessions of experiment 1, but stopped being so in later sessions.

3.4 Discussion

Experiment 1 took a neuroimaging approach to the study of novel and familiar decision-making in non-human primates. Whereas electrophysiological research has focused on recording single-unit activity in selected brain regions in primates, most often in repetitive and overlearned tasks, we were able to show for the first time whole-brain activity during novel decision-making, with an approach similar to the one commonly taken for humans; additionally, we were able to make a specific comparison between the brain activity associated with familiar and novel choices thanks to the specific design of the training and testing conditions.

It is interesting to notice (figure 3.5) that if on the one hand the overall brain activity during choice and response engaged large parts of the brain, the contrast of novel versus familiar trials highlighted an enhanced activation in the principal sulcus for novel decisions. This region is known to be part of a multiple-demand network that is consistently recruited in a large set of demanding tasks, such as the Stroop task, working memory tasks and challenging perceptual tasks (Duncan, Owen, 2000; Duncan, 2010). It is plausible to assume that choices between pairs of novel stimuli were more demanding than corresponding familiar choices, even if they were matched for average value, average value difference, and dimensionality of the comparison. Thus, this result can be seen as a confirmation that macaque dlPFC may have a similar role in coping with demanding tasks as in humans. We found less clear evidence for the involvement of the other nodes of this network, such as ACC and insula (Mitchell *et al.* 2016), in the novel choice task.

A parametric approach to value difference encoding in the brain revealed a strong effect in OFC (figure 3.6). Notice that by OFC here we refer to areas 11 and 13, which,

as discussed in the introduction (section 1.3.2), are part of the lateral orbitofrontal network. Such region allows the evaluation of expected reward, especially in animal experiments involving primary rewards, and is sensitive to the identity and the sensory features of the outcome. We observed an increase in BOLD signal when the value difference between stimuli was lower, which can be interpreted as a sign that the region is more active, and for a longer time, when the two options are more similar to each other and therefore the choice is more challenging. Given the temporal resolution of fMRI and the range of decision times displayed by monkeys (in the order of a few hundred milliseconds when excluding perception and motor response times), a longer decision process would be captured as a stronger, but not necessarily longer, BOLD effect.

The clear effect of value difference in OFC is consistent with the importance of this region in the electrophysiological literature (Padoa-Schioppa, Assad, 2006; 2008) and with recent fMRI results in non-human primates (Chau *et al.* 2015; Papageorgiou *et al.* 2017), but contrasts with its comparatively rare involvement in human fMRI value-guided tasks (but see Klein-Flügge *et al.* 2013). This may be either due to an inter-species difference, or to a difference in the type of reward used: whereas most human experiments used secondary rewards, typically money, those studies involving primary rewards in humans, either only allowed the reward consumption at the end of the fMRI session, or employed erotic rather than edible rewards. Thus, it is possible that the strong involvement of OFC in value-guided decision-making may be specific to situations where the reward is multisensory (not purely visual) and expected on a timescale of seconds.

The contrast of value difference regressors in novel and familiar trials (figure 3.8), reveals what is potentially the most interesting result of this experiment, namely a strong effect of novelty, consistent in all four subjects, in AMFC and in two additional areas,

the principal sulcus and the putamen. The effect cannot be explained by a difference in stimulus value, or by the difficulty of the comparison, but is instead specific to the nature of the options compared: in one case they are very familiar options, experienced thousands of times, in the other case they are novel options, whose component features are understood, but which have not been experienced together before. As a consequence, the compound value of familiar options may be readily available at the time of the choice for familiar trials, but may need to be constructed on-the-fly for novel trials.

This result is consistent with previous knowledge about these regions. The involvement of dlPFC in novel trials indicated by both the main effect and the parametric effect highlights the key role played by this region when coping with novel choices and sustained cognitive demands (Duncan, Owen, 2000). Ventral striatum (vSTR) is known to play a role in reward-guided learning and receives dopaminergic inputs from the midbrain (Haber, Knutson, 2010). Finally, it has been argued that AMFC (vmPFC) encodes a context-independent, economic representation of both expected value and outcome value and is necessary to guide value-based decision making (Levy, Glimcher, 2012). Even if this latter region has been seldom studied in non-human primates, the activation in the present experiment bears a striking resemblance with many human fMRI results in investigations of value-guided decision-making (Clithero, Rangel, 2010; Bartra *et al.* 2013). This region in humans is commonly referred to as vmPFC, although peak activations in different experiments range over various parts of medial frontal cortex, including proper vmPFC, area 14m, ACC, area 32, and frontal pole, area 10.

A further investigation of the AMFC effect made clear that the contrast is driven by a positive effect of value difference in familiar trials, and a negative effect in novel trials. The nature of this difference is not entirely known. We hypothesize that the familiar positive effect may be driven by a process similar to the one observed in human fMRI.

As discussed in the Introduction (section 1.3.3), AMFC may encode the expected value of the outcome, relative to a baseline given by the unchosen option; this can be seen mechanistically as the outcome of a decision process implemented in a mutual-inhibition attractor model, if the network maintains its activity after the decision has been reached. Consistent with this possibility, value is encoded positively in human AMFC. On the other hand, we hypothesize that the negative effect of novel trials may be driven by a different phenomenon, similar to the one described above in relation to the OFC. It is possible that the challenge of computing the value of both novel options online and comparing them overrides, by orders of magnitude, the positive value signal in AMFC. Because in familiar trials the value of each option is available due to long practice, the value retrieval process may take a similar amount of time, with little relation to value difference. Conversely, in novel trials an accurate estimation of the value of each option depends on a distinct process; if we assume that the process of comparison starts while the process of precise estimation is still ongoing, it is plausible that in trials with a large value difference this process of value estimation of each option may be stopped short as soon as a preference is reached, but in trials with a small value difference, the value estimation process may be protracted for longer, or performed more accurately, because it is more challenging to reach a clear preference for one option rather than the other. Consequently, this hypothesis would predict a larger BOLD effect for novel (but not familiar) trials with small value difference. This is what we observed in AMFC and dlPFC. The different contributions of these two regions may be explainable by a role of the medial PFC in the motivational aspect of decision-making, and a role of the lateral PFC in its cognitive control (Kounehier *et al.* 2009).

The novelty effect in the striatum (figure 3.8) can be related to a previous study showing that contextual novelty changes reward representations in the striatum (Guitart-

Masip *et al.* 2010). The precise location of the activation we observe is in ventral putamen, more lateral than the most frequently mentioned nucleus accumbens, but it is equally part of the reward circuit and has been associated with a reward prediction error signal (O'Doherty *et al.* 2004, McClure *et al.* 2003).

An inspection of the timecourse of activity in familiar and novel choices in AMFC supports this view (figure 3.9). In familiar choices, an initial increase in value sum, plausibly at the time of the decision, is followed by a positive effect of value difference, plausibly during the inter-trial interval. This is precisely the pattern of activity predicted by an attractor model and observed in human vmPFC with MEG (Hunt *et al.* 2012); although we did not expect *a priori* to see such an effect due to the low temporal resolution of the BOLD signal, it is possible that the signal in macaques during the action-outcome delay is more sustained than in humans. This may be partly related to the immediate and primary nature of the reward. In novel choices, instead, the effect of value difference was negative throughout the decision and delay periods, confirming the different nature of the cognitive process underlying novel choices.

The chosen value effect on the signal timecourse is markedly different after the outcome between novel and familiar trials. We observe no modulation in the familiar condition but a rapid switch from negative to positive modulation in the novel condition. It is possible that the effect observed after the outcome is a form of reward prediction error in AMFC: whereas the outcome in familiar trials is fully predicted, and therefore does not modulate activity, the outcome expectation in novel trials is more noisy and unprecise and therefore the actual outcome causes surprise and a reward prediction error signal that is captured by the chosen value regressor.

It is remarkable that the location identified by our familiar versus novel contrast, where a sign reversal of the value difference effect takes place, corresponds to the area

most commonly associated to value encoding in human fMRI. Just dorsal to this location, in area 9, we observed a different pattern, with chosen value being encoded in a consistently positive way in all types of trials (figure 3.11). This result is compatible with other human fMRI studies, described in section 3.1, finding encoding of chosen value in dmPFC for novel or "modelled" decisions.

The bilateral effect in the midbrain for value sum in familiar vs novel trials indicated, differently from other contrasts discussed above, that the region was more active for familiar trials and less active for novel ones (figure 3.12). This effect, although it was not robust to cluster correction given its small size, was in a location consistent with the ventral tegmental area (VTA) and the substantia nigra, which are known to encode a reward prediction error signal (Schultz *et al.* 1997) and to send dopaminergic projections to the nucleus accumbens and the prefrontal cortex (Haber, Knutson, 2010). We observed an effect of value sum at decision time, which may be read as a prediction error at the time of the presentation of the two stimuli, before a choice was made; such an effect is expected if the options on a given trial are better or worse than the average of all trials. The fact that a dopaminergic response to the presentation of valuable stimuli was observed for familiar but not novel trials is an indication of the fact that the value of familiar options was encoded by the habitual reward system by the end of the long training, but this was not the case for novel options during experiment 1.

AMFC has been consistently associated with value encoding and value-guided decision-making in the human fMRI literature. However, a consensus on the precise contribution of the region at the algorithmic level has not been reached, and both the type of tasks employed and the specific locations activated vary across studies. A ventral location, corresponding to cytoarchitectonic area 14, is a candidate for value integration in multidimensional choice (Fellows, 2006). It is interesting that if on the one hand we

found a global effect of value difference in all novel versus all familiar trials in AMFC, when we used the more specific contrast between novel conflicting versus familiar conflicting trials, we found a specific effect in area 14, with a peak in its most anterior part (figure 3.13). Conflicting trials, by definition, require the integration of each attribute's contribution into a combined value in order to make a correct choice, because focusing on a single dimension does not allow identification of the best option.

The use of the subjective value model based on behaviour allowed us to create value difference regressors that took into account the biases described in chapter 2. As suggested at the end of that chapter, behaviour was more biased for familiar than novel trials, even if accuracy was higher in the familiar trials. This may be interpreted as a sign that habitual value representations, based on lengthy training, are encoded in idiosyncratic ways; such representations are immediately available at the time of the familiar decision. On the other hand, no such habitual value representation is available for a novel choice, and therefore compound value needs to be constructed on-the-fly. It is possible to assume that this mechanism for online value integration may be slower and noisier, but not affected by the same type of systematic biases observed in familiar value representations. If this is the case, then a neuroimaging analysis using subjective rather than objective value may find strengthened effects of value difference in familiar trials but weakened effects in novel trials. This is the pattern that we observed. The main effect of decision, the value difference effect for all conditions including the familiar condition in OFC, and the familiar value effect in the midbrain were maintained in the subjective value analysis. The trend for effects to change across sessions in the familiar condition was strengthened. On the other hand, the novel versus familiar effects in AMFC and dlPFC were weakened perhaps because, as discussed above, the values in the novel condition did not match the biased subjective value regressors.

According to our hypothesis that the negative effect of novel value difference in these regions may be due to the duration over which the values of novel, multidimensional options have to be integrated, it is possible to suggest that the duration or the intensity of such process depends on the objective similarity of the two options to be compared, rather than their subjective similarity. This is not implausible, because before the computation itself their subjective similarity is not even available.

In conclusion, the overall pattern of results seems to suggest that different processes are involved in novel and familiar value-guided decision-making. All decisions involve OFC, plausibly because it represents the expected value of the primary reward used (fruit juice), and more so for difficult decisions where the values of the two alternatives are similar. Familiar trials activate regions that have also been linked to a habitual reward system, including VTA/substantia nigra, ventral striatum, hippocampus and surrounding cortex, and OFC. It is important to stress however, that we did not study habitual stimulus-response associations, because even if the nineteen familiar options were relatively few in number and well known, they could be involved in a total of 361 different possible binary decisions. Also in many case the pairs of alternative familiar options presented in experiment 1 had not been compared before, because training pairs were always differing by a single dimension at a time, but testing pairs could differ in both dimensions.

On the other hand, novel choices did not involve the midbrain to the same degree as familiar ones. They were also associated with an enhanced activity in a core part of the multiple-demand network: dlPFC. Most importantly, both types of decisions recruited the medial frontal cortex, but by different mechanisms. We suggested that this pattern of results may be consistent with the idea that, for familiar decisions, value difference is encoded positively after the decision has been taken, whereas for novel decisions a

negative effect of value difference is manifested because when the options are similar in value, considerable integration time is required to estimate their value precisely.

Our data (when comparing the contrast of all conditions and the contrast of conflicting conditions only) may also support the hypothesis that a more ventral region identified with vmPFC implements attribute integration, and therefore estimates the value of each option separately, whereas slightly more dorsally, AMFC performs a comparison of the two alternative options and selects the preferred one.

Chapter 4
Neural representations of reward-predictive stimuli in value space within prefrontal cortex

4.1 Introduction

Decision-making is essentially composed of two processes: the evaluation of each available option and the selection of the preferred one. As described in the Introduction (section 1.3.1), these logically sequential steps may be implemented as dynamical processes in the brain and be partially simultaneous. Our research investigates the neural bases of novel decision-making and the novelty is defined as a property of the stimuli, rather than a property of the comparison. Experiment 1 reported in chapter 3 focussed on the comparison process during decision making, although care was taken to equate the difficulty and complexity of the comparison process across familiar and novel conditions.

Further insights into the construction of such novel representations may be gained by the study of how individual options are perceived, independently from a comparison process. For this reason, experiment 2 uses the same novel stimuli as experiment 1 but they are presented one at a time to the subjects, followed by the probabilistic outcomes associated with them. This experiment enables us to observe neural signatures of the construction of novel stimulus representations, in two essential respects: first, these stimuli represent outcome in a complex way, varying in both amount and probability. A useful representation of the meaning of a stimulus should therefore combine these features into a multidimensional construct. Second, because the combinations are novel, and no integration was required during training (because training sessions were

unidimensional), we expect to observe such a representational system building up across the duration of the experiment, as it becomes needed for the first time by the subjects. For this reason, sessions of experiment 2 were interleaved with those of experiment 1, to cover the period of exposure to the novel choices and test changes in the neural representation of multidimensional value space.

We suggest that such a representational system, if it is tuned to what matters in the end to the subject, should learn to compute the expected value of incoming stimuli; this means to categorise stimuli not merely as a function of their component features, but as a function of their overall value; we use fMRI adaptation to test whether candidate areas for novel value representation, such as AMFC, are indeed sensitive to such abstract properties of the stimuli. Moreover, if they are, the next question to be addressed is how the integration is performed in practice. We suggest that a candidate mechanism for such process is a grid-like encoding of an abstract value space, similar to the encoding of physical space in entorhinal cortex; we test this hypothesis by measuring the modulation of the neural signal as a function of the trajectory between one option and the next in value space.

Repetition suppression, or adaptation, is a well-known phenomenon by which repeated or prolonged presentation of a stimulus causes a reduced neural response (Blakemore, Campbell, 1969); this effect has been widely used in fMRI studies (Grill-Spector, Malach, 2001). Multiple potential mechanisms have been suggested to cause adaptation: one is habituation, another one is visual priming (Kourtzi, Grill-Spector, 2006). An advantage of repetition suppression in fMRI is that it may identify neuronal subpopulations within the same voxel that could not otherwise be discriminated.

Here, we use fMRI adaptation to probe whether a neural response is sensitive to the composite value of an option (magnitude times probability), in a way that is not affected

by the common correlates of value expectation, such as arousal and a stronger motor response. This is possible because we compare trials that are on average equally valuable, but depending on the value of the preceding trial (similar or different), they may be expected to display different degrees of adaptation.

The second analytical approach we take is inspired by studies of spatial mapping in the temporal lobe. Neurons in the hippocampus called place cells have a spatial receptive field (O'Keefe, Dostrovsky, 1971); they receive their input from the medial entorhinal cortex (Brun *et al.* 2002). For their part, entorhinal neurons also have a spatial receptive field, but are responsive to multiple locations (Fyhn *et al.* 2004) and these cover the whole environmental space with a specific triangular pattern (Hafting *et al.* 2005). The grid of each neuron in the same region has the same alignment, but is randomly shifted in phase. Such structure enables the representation of position and movement in space, relative to distal landmarks (Hafting *et al.* 2005).

Grid cells have been extensively studied in rodents, but an innovative design allowed Doeller and colleagues (2010) to reveal a signature of the grid structure through fMRI in human volunteers. Because the grids of all cells within a given region are parallel, a movement in space aligned with the grids will increase the firing rate of more cells compared to a misaligned movement. Using a virtual reality setup, it was found that fMRI signal in bilateral entorhinal cortex was modulated as a function of the direction of movement, and more so when moving faster. Additionally, multiple brain areas including entorhinal cortex and AMFC displayed speed-dependent fMRI adaptation to runs at 60º from the current direction (Doeller *et al.* 2010).

Intriguingly, Constantinescu and colleagues (2016) suggested that such mechanism of neural encoding of space may not be limited to physical space, but be adaptable to represent novel arbitrary two-dimensional spaces. Using human fMRI, they asked their

participants to learn to categorise bird drawings into families, each one predicting a different outcome; importantly birds were taken from a continuous "bird space" (never explicitly revealed to the subjects) within which birds varied by two dimensions: legs length and neck length. Although bird space was purposefully arbitrary, subjects after training revealed a hexagonal symmetry in the BOLD response to "morphing trajectories" when bird images were transformed into one another, precisely as would be predicted by a grid-like encoding of bird space. Such an effect was observed in a network of regions similar to the one reported by Doeller and colleagues, and including AMFC and entorhinal cortex (Constantinescu *et al.* 2016).

Our experiment aims to assess whether a two-dimensional value space, which may be new to the subjects in this specific form, but which is also ecologically more relevant to them, can be represented with a grid-like code, hinting at a general and flexible mechanism for learning novel two-dimensional abstract maps.

4.2 Methods

4.2.1 Experimental design and task

The experimental setup was the same as described in chapter 2 section 2.2.2 for behavioural training. Briefly, head-fixed animals in a sphynx position in a chair were trained in a custom-built wooden mock-scanner and tested in an MRI scanner. They were positioned facing a screen and two custom-built infrared sensors. Reward was delivered into the mouth through a plastic spout.

Four Rhesus macaques, after extended training with familiar trials, performed 17 sessions with novel trials in consecutive days (excluding weekends) belonging to either experiment 1 or experiment 2, as described in table 4.1.

2	1	1	1	2	1	1	1	2	1	1	1	2	1	1	1	2

Table 4.1 Alternating session order of experiments 1 and 2.

As described previously, experiment 1 consisted of twelve sessions of binary choice trials and experiment 2 consisted of five sessions of rapid single-option trials. Because monkeys were exposed for the first time to novel choices across the twelve sessions of experiment 1, it was expected that their mental representation of the value space would form gradually. Experiment 2 spanned the full period of exposure to the novel choices, and aimed to capture the presence and development of such multidimensional value representations.

The available options were the same as in experiment 1. There were ten levels of reward magnitude, varying from 1 drop to 10 drops of juice and ten levels of reward

probability varying between 10% and 100%. Magnitude was cued by colour and probability by number of dots (figure 4.1). The monkeys did not experience the full space of possibilities until the testing phase in the scanner. Training was performed in two sub-regions of this space: on even days they were only presented with stimuli with a probability of 60% and a magnitude drawn randomly from one to ten drops (figure 4.1, top left and same figure, centre, vertical grey line); by contrast on odd days, they were presented only with stimuli with a magnitude of 5 but with a probability that was drawn randomly between 10% and 100% and in steps of 10% (figure 4.1, left bottom and same figure, centre, horizontal grey line). Jointly, these stimuli are defined as the familiar stimuli. Training trials were either forced choices (a single option was presented on the screen) or binary choices of familiar stimuli (animals chose between two options). Only in the final testing phase were subjects presented with novel combinations of magnitude and probability not experienced before (white squares in figure 4.1, centre).

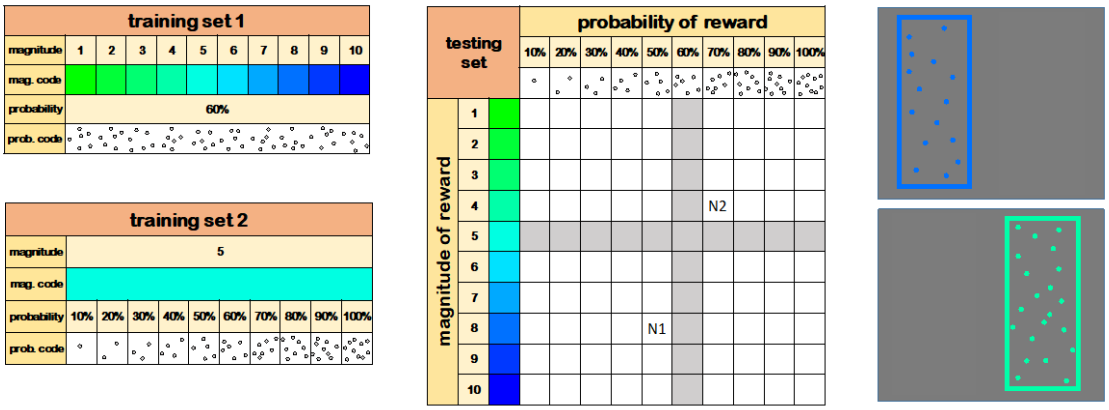


Figure 4.1 Value space and experiment 2 task. **Left:** Separate training sets for magnitude (top) and probability (bottom) and illustration of the outcome coding. **Centre:** Two-dimensional value space and illustration of the outcome coding. Familiar options in light grey, novel options in white, unused options in dark grey. **Right:** visual appearance of the task. In the top example, a novel option (N1) gives 8 drops with a 50% chance, in the bottom example, another novel option (N2) gives 4 drops with a 70% chance.

In contrast to experiment 1, testing sessions of experiment 2 included exclusively novel single-option trials and the timing of the sequence was designed to minimise the interval between a stimulus and the next one. The temporal structure of the task was the following: a single option was presented randomly to the left or the right of the screen for a fixed duration of approximately 1 second (adjusted to each subject's training performance but fixed for each subject for the duration of the experiment). If a response was recorded within that deadline, the outcome was delivered at the end of the fixed stimulus time; otherwise the trial counted as a miss. The outcome period lasted for 3 seconds, regardless of whether and how much reward was delivered. An inter-trial interval followed with a duration drawn from a uniform distribution between 0 and 0.5 seconds. The exact stimulus duration was 1.1, 1, 1.5 and 1.4 seconds for monkeys A, B, C and D respectively. This temporal structure assured that the duration of the stimulus presentation was constant across trials, the interval between a stimulus and the next was on average 4.5s (±0.25s) and it was not affected by performance or by whether a reward was delivered.

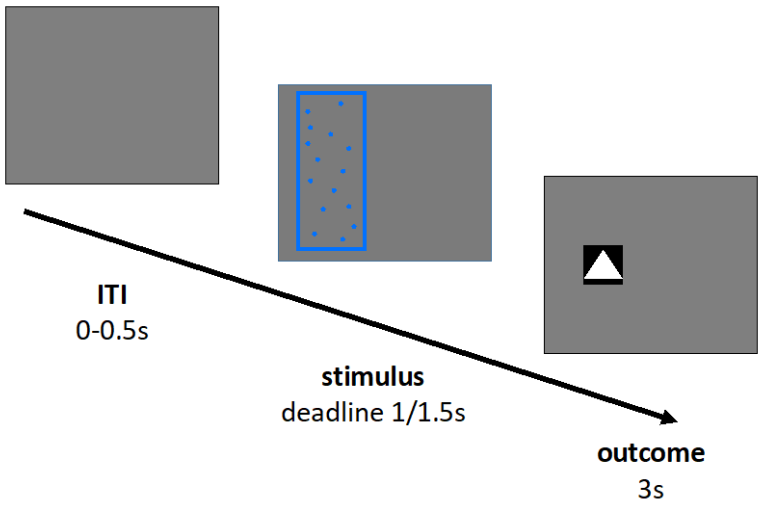


Figure 4.2 Experiment 2 task timeline

Subjects performed a short sequence (156 planned trials) in session 1, before the beginning of the novel choice experiment, and long sequences (234 planned trials) in sessions 2 to 5, spaced during the novel choice experiment. The first session was shorter in order to minimise the exposure to novel options before the start of experiment 1 the following day. Because of the very fast pace of the task (~13 trials per minute) and in order to avoid too short sessions and the delivery of too large amounts of reward in a short time, sequences were slowed down with the addition of a 30 second break every 26 trials. The only exception to this was monkey B, who performed 5 short sequences without breaks.

Given the short deadlines, subjects struggled to perform this task, and even if missed trials were repeated at the end of planned sequences, on some occasions the originally planned sequences were not completed. In particular session 2 of monkey C could not be analysed, and session 1 could only be partially analysed, because of the low number of completed trials. This was compensated by acquiring one extra sessions at the end of the experiment.

All procedures were conducted under a Home Office licence according to the Animal (Scientific Procedures) Act 1986 and the European Union guidelines (EU Directive 2010/63/EU).

4.2.2 Repetition suppression design and analysis

The first use of the sequential single-option data was to test the neural representation of value as an abstract variable distinct from its component dimensions, through fMRI adaptation. Because all stimuli were novel (not familiar) and only one stimulus appeared per trial, the key comparison was between stimulus at trial n and stimulus at trial n-1. All

trials were categorised as belonging to one of five conditions, depending on the relation between the present and the previous stimulus:

- ID – identical, if the present stimulus is identical to the previous stimulus (the one presented in the previous trial);
- SM – same magnitude, if it has the same magnitude, but not the same probability, as the previous stimulus;
- SP – same probability, if it has the same probability, but not the same magnitude, as the previous stimulus;
- SV – similar value, if both the magnitude and the probability differ from the previous stimulus but the expected value (their product) is close to the previous one, as defined by an expected value difference smaller than 0.33 drops;
- DV – different value, if both the magnitude and the probability differ from the previous stimulus and the expected value is not close to the previous one.

The stimuli sequences were designed in such a way that 30 or 45 predefined pairs of successive stimuli would appear for each condition during each session, in short and long sessions, respectively. Planned pairs were concatenated, in such a way that each stimulus would count both as the second element of a pair and as the first element of the next pair. In this way, each short sequence contained 156 pairs (30 x 5 + 6) and each long sequences 234 pairs (45 x 5 + 9), where 30 and 45 are the number of trials per condition, 5 denotes the number of conditions and 6 and 9 denote the number of trials following the long breaks that could not be concatenated. If a subject missed the deadline or did not respond at all, the relevant pair(s) would be repeated at the end of the session; only pairs in which responses were recorded for both trials within the deadline have been analysed.

Sequences never contained more than two identity pairs in a row (implying three identical stimuli in a row), and at most nine of these triplets.

Animals C and D failed to respond to some of the stimuli on the first occasion they were presented. As a result, some of the original planned pairs (a given stimulus on trial n-1 and a given stimulus on trial n) were not acquired. But because the missed pairs were repeated at the end of the session, many of them could be acquired then, and new pairs of successive stimuli were also acquired because of the unplanned concatenations during the repetition of pairs originally far apart in the sequence. By this mechanism the number of correct trials was generally superior than the number originally planned, but the number of correct pairs was approximately equal for at least 3 animals. If we include in the count the unplanned pairs, but discard the ones in excess of the planned amount within each session and condition, animals completed respectively 2%, 0%, 23% and 4% fewer pairs than planned on average.

The full set of successful trials for each session was therefore resampled *post hoc*, discarding some trials, in order to make use of most of the new unplanned pairs but at the same time maintaining the balance of the design. Both original sequences and resampled sequences were designed in such a way that the average value of stimuli in each of the five conditions, and particularly in the SV, ID and DV conditions, would be as uniform as possible (the average of all conditions within a session lay within a range of 0.1 drops, where the average expected value of any given trial across the whole experiment was ~2.5 drops); moreover, the average value difference between the preceding stimulus and the following stimulus was also kept as uniform as possible across conditions (while the absolute value change was by definition 0 for ID, small for SV and large for DV, the average relative value change, indicating whether the present option is better or worse than the previous one, was kept as similar as possible between the SV and the DV

conditions, with an error smaller than 0.1 drops). The absolute value difference of similar (SV) pairs was always below 0.33 drops and on average below 0.2; the absolute value difference for different value (DV) pairs was always above 0.33 and on average above 1.5 drops. The sequences were also designed to make as uniform as possible the mean magnitude, the mean probability, the mean magnitude change and the mean probability change across conditions. Given these constraints and the animals' performance, a set of 37 trials (±12) per condition was analysed, with the same number of trials across conditions within each session.

A GLM was created, as had also been the case for the choice data of experiment 1, to perform neuroimaging analysis using the FEAT tool of FSL. The five regressors of interest (ID, SM, SP, SV, DV) represented one category of trials each, with a constant value of 1 for those trials, starting at the stimulus onset and with a long duration of 3 seconds. Other task-related confound regressors were added to the GLM to control for irrelevant factors: more constant regressors of same duration were created to capture trials when the response deadline was missed, trials following trials with a missed deadline (these trials could not be used because of the comparative nature of the adaptation design), trials following the planned break and other trials discarded in order to balance the design. Simultaneous with these constant regressors, four parametric regressors were added to index each stimulus's expected value, reaction time, trial number and expected value difference with the previous stimulus. Finally, a binary regressor denoted response side. Additionally, three separate constant regressors were used to model positive outcome, negative outcome on trials on which animals had responded in time and negative outcome on which animals had not responded in time, respectively. Simultaneous with these, a parametric regressor modelled the reward amount. All the parametric regressors were normalised and all regressors, both

parametric and non-parametric were HRF-convolved. Three further non-convolved regressors were added to capture immediate motion-related artefacts due to field distortions: left side response, right side response, and juice delivery. The first two had a duration of one repetition time (TR), and the last had a duration corresponding to the juice delivery time for the precise amount of each trial, but a fixed amplitude of 1.

Task-unrelated confounds were those used in experiment 1 (motion regressors, low quality-volume confounds), plus an extra set: non-brain ICA components. Low quality volume regressors where null at all points except one each, flagging up TRs where the signal distortions were 2.5 standard deviations larger than the average. Thirty additional regressors were added to the GLM, obtained by the first thirty components of an independent component analysis (ICA) run with the FSL tool MELODIC (Multivariate Exploratory Linear Decomposition into Independent Components) (Beckmann, Smith 2004) version 3.14, on the 4D images where all brain voxels were zeroed. The ICA components were therefore able to model the field distortions as observed on non-brain tissue and subtract them from the signal within the brain, increasing the relative proportion of the remaining signal that was due to the BOLD response. Preliminary testing confirmed that this approach improved the final statistical results, as assessed by effects associated to task-related regressors other than the key regressors of interest.

The regressors of interest, namely the five constant regressors representing the five types of valid pairs, had a correlation $r<0.14$ with any other regressor on average, an and $r<0.28$ in any given session.

A subjective value design was also considered; however, given the categorical nature of the stimuli (a given trial is assigned to either the SV or the DV condition), the two designs were very highly correlated. The results based on objective values are therefore reported.

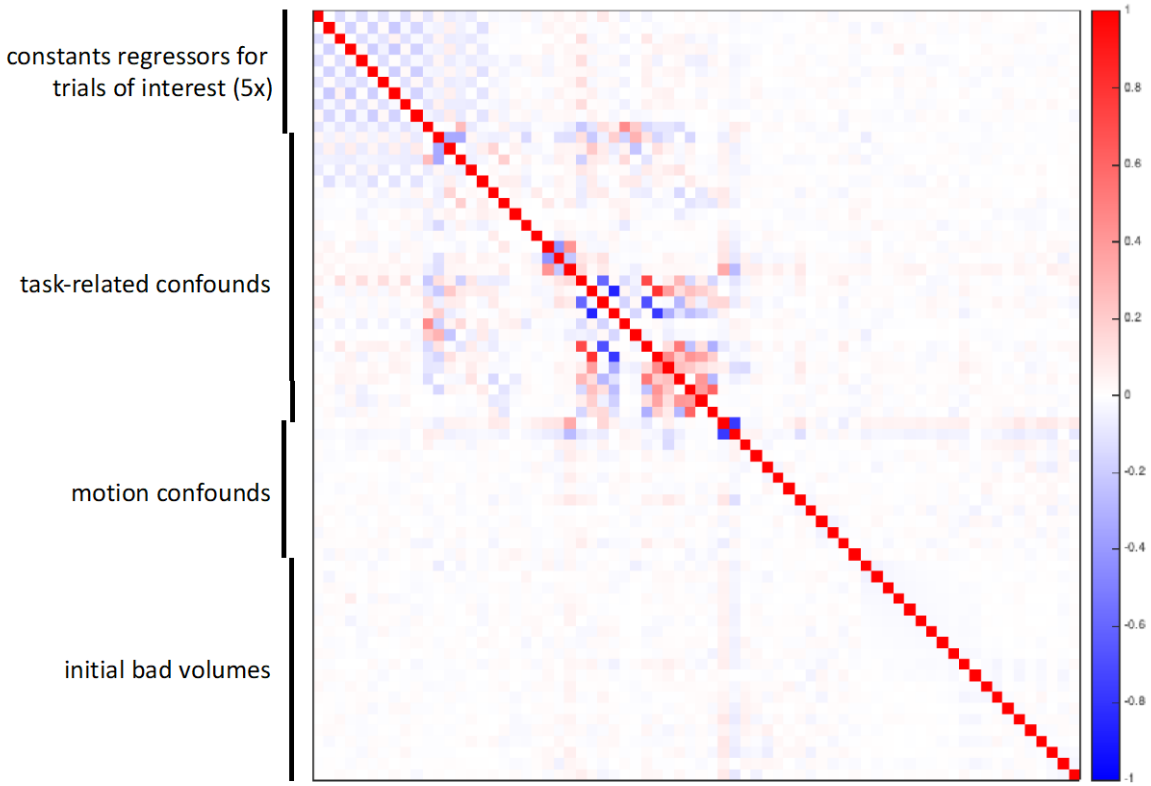


Figure 4.3 Correlation matrix for the repetition-suppression design. Each task-related regressor is followed by its temporal derivative. The first 20 low quality volume regressors are illustrated; their number varies across sessions. For the regressors of interests all $r<0.14$.

For each session two contrasts of interest were calculated: identity minus different value (ID-DV) and similar value minus different value (SV-DV). As a control, also same magnitude minus different value (SM-DV) and same probability minus different value (SP-DV) were tested. The first contrast was designed to identify areas that show adaptation when the very same option is presented twice in a row, compared to a different option. It is important to notice that although identical options are visually similar, they are not identical because the side of the screen changes half of the time and

the exact location of the points is also reshuffled. Therefore, identical options are not necessarily visually identical. The second contrast is instead designed to identify those regions that do not code for the identity of the stimulus, but for the value that it represents. This is a more abstract representation because stimuli with different magnitude and different probability may have the same value. Regions activated by this contrast are sensitive to a combination of multiple components of value. It is also important to note that, as described above, although options in the DV condition have by definition a value different from the previous stimulus, this change can be either for better or worse, so that on average within each session a DV transition is equally valuable as a SV or ID transition.

At the higher level, sessions were combined according to two strategies; the underlying assumption being that neural representations of abstract value may or may not be present when a subject has no experience with the multidimensional value space. In the second case, fMRI adaptation may appear only after such representations have been constructed with practice across the duration of the experiment. Therefore, to keep the assumptions to a minimum, we ran one higher-level analysis averaging all 5 sessions, and one analysis spanning the second half of experiment 1, averaging sessions 3, 4 and 5 of experiment 2 for each subject.

4.2.3 Grid-like code analysis

The second way sequential data from experiment 2 were used was to search for a grid-like encoding of value space. Under the assumption that abstract two-dimensional value space may be encoded by cells in medial frontal cortex with a grid code, we looked for an fMRI signature of this representation.

The fast succession of stimuli in experiment 2 may be conceived as a series of movements within the abstract value space, with a "jump" per trial. Therefore, every valid trial (validity will be defined below) was associated with the angle of the trajectory between the location of the preceding stimulus and the location of the present stimulus. Initially, this angle was defined in relation to a standard orientation parallel to the probability axis. The actual mean orientation of the grid was then estimated. If a triangular grid exists, movements along trajectories aligned with the grid will cause stronger neural activity than unaligned movements, because a larger number of cells will be activated in the former case. If all possible trajectories are tested a pattern of signal increases and decreases every 60 degrees is predicted, corresponding to the cyclical alignments and misalignments of the trajectory to the axes of the triangular grid. As a consequence, the signal is expected to be a sinusoidal function of the trajectory angle with a periodicity of six. The number six derives from the fact that within a full 360º cycle there are six ways of aligning to the grid.

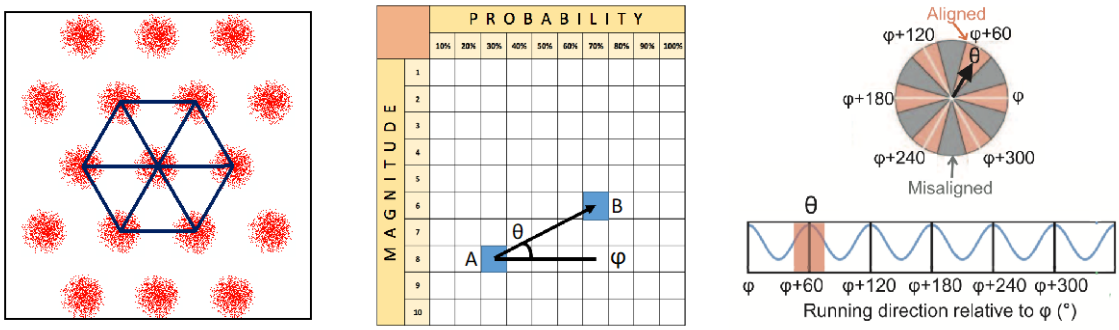


Figure 4.4 Grid-like code analysis design. **Left:** schematic example of the receptive field of a grid cell, following a triangular pattern (in red); this implies a hexagonal symmetry (black lines). **Centre:** stimulus pairs define directional movements in the abstract two-dimensional value space. In the example, stimulus A is followed by stimulus B. The angle between them (θ) is defined with reference to a given orientation (the probability axis in the example, labelled φ). **Right:** movements at an angle aligned with the grid will

induce an enhanced neural response, compared to misaligned movements. Right panel figure from Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science*, *352*(6292), 1464-1468. Reprinted with permission from AAAS).

Because the grid orientation is unknown, following Constantinescu and colleagues (2016), we assumed an arbitrary standard orientation, in our case parallel to the probability axis, and created two regressors for the fMRI analysis: one proportional to the sine of six times the trajectory angle and one proportional to the cosine of the same quantity. If one or both of these regressors correctly predict BOLD signal variations, this is an indication that neural activity follows the six-fold pattern derived from the grid code hypothesis. We therefore ran an F-test over both the sine and the cosine regressor beta estimates to individuate brain regions sensitive to this effect. The same test was repeated for all periodicities between four-fold and eight-fold to establish whether the hypothesis of a six-fold symmetry was a better explanation of the neural data.

For the purpose of the grid code fMRI analysis, a separate GLM was created. Trials were defined as valid if the subject responded within the deadline to both the trial itself and the previous one; additionally, only non-identical trials were considered valid, because in ID trials no movement occurred. In this way 158 trials (±51) were analysed per session.

The GLM included five main constant regressors, for valid trials, ID trials, trials with no response, trials preceded by trials with no response, and trials preceded by a break; these five regressors were aligned with stimulus onset, had a duration of 3s and an amplitude of 1. Four parametric regressors representing log reaction time, time in the session, expected value, and expected value difference with respect to the previous trial, were aligned to the corresponding constant regressors for all trials where a response was

correctly recorded. Moreover, the two key parametric regressors of interest for the present analysis were sine(6xθ) and cosine(6xθ), where θ is the angle of the trajectory relative to an arbitrary orientation (i.e. the probability axis). These regressors were only present for valid trials, with same onset and duration as the valid trial constant regressor. Additionally, in the GLM two response side regressors (left and right) with onset at stimulus time and duration of 1s were included, and four additional outcome-related regressors, as described in the GLM for repetition suppression. Parametric regressors were normalised, but not the sine and cosine regressors. All these regressors were HRF-convolved, but, as for the repetition suppression GLM, three further regressors for each response side and juice duration were not convolved. The task-independent confound regressors were the same as above, namely motion confounds, low quality-volume confounds and components of activity outside the brain identified with an ICA. The correlation of the sine and cosine regressors was $r<0.14$ with any other regressor on average, and $r<0.33$ in any given session.

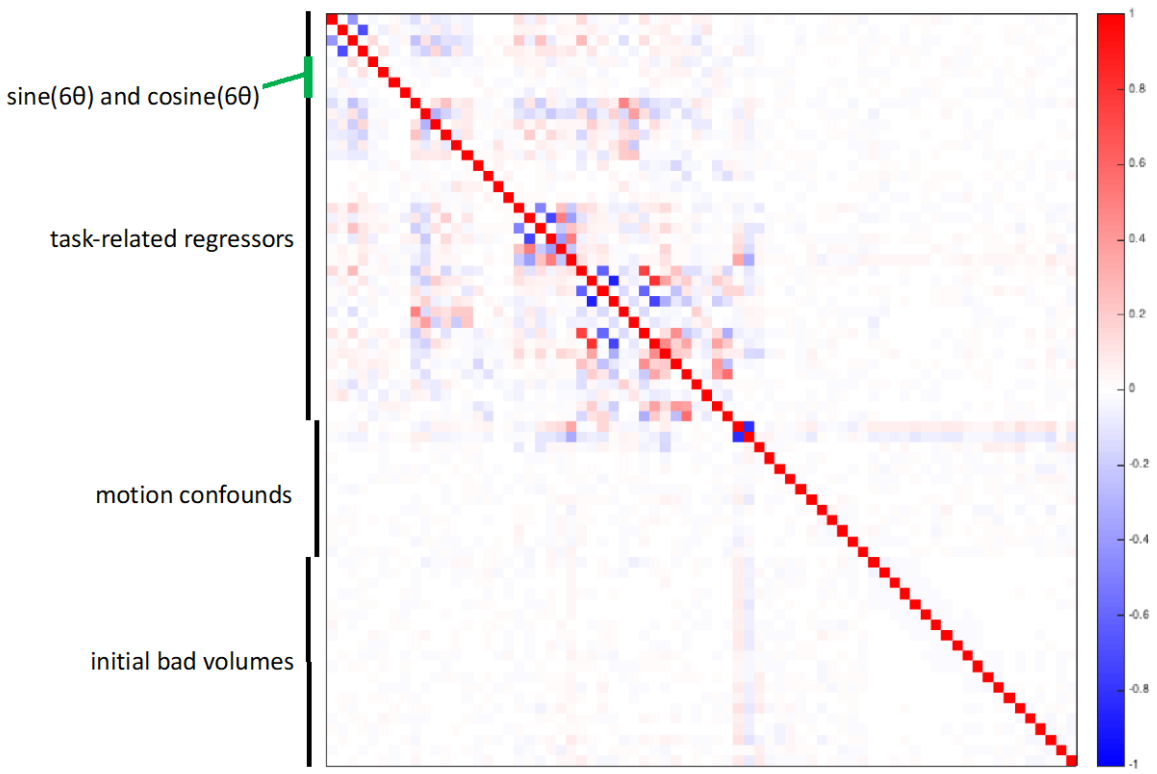


Figure 4.5 Correlation matrix for the grid-like code design. Each task-related regressor is followed by its temporal derivative. The first 20 low quality volume regressors are illustrated; their number varies across sessions. For the regressors of interest, all $r<0.14$.

In each session, locations revealing a potential grid-like code are identified by an F-test of the two regressors of interest: sine and cosine; this result however cannot be correctly used at the second level to test across subjects and sessions because the skewness of the F-distribution may increase the false-positive rate (Constantinescu *et al.* 2016). Therefore, the F-test was used to select ROIs that were subsequently analysed with an approach orthogonal to the initial ROI selection. The F-values were averaged across the last 3 sessions of experiment 2 for each subject, corresponding to the second half of experiment 1. These sessions were selected *a priori* under the assumption that, if it exists, any value space grid-like representation may take time to develop through exposure to the novel stimuli. ROIs were defined on the two peaks of the effects

determined in this manner within prefrontal cortex. For illustration purposes (see figure 4.9) average F values (df>1, 207)[5] were thresholded at F>8, p<0.005. The threshold was chosen arbitrarily, given that there is no established convention about F-tests in fMRI analysis.

To provide an independent test of the grid code hypothesis in the selected ROIs, trials were split in two arbitrary subsets; after excluding all non-valid trials, even-numbered trials and odd-numbered ones were analysed separately. A new GLM was created with twice as many regressors relative to the decision period as the one reported above (two constant regressors for odd and even trials, two sine and two cosine regressors, two logRT and two time-in-task regressors, two expected value and two expected value difference regressors), for even and odd trials respectively, together with one copy of all other regressors relating to the response and outcome periods as in the previous GLM, and the same confounds. A neuroimaging analysis was performed with FEAT and beta estimates for the sine and cosine regressors were used to infer the orientation of the triangular grid in each subset independently (Doeller *et al.* 2010). Under the assumption that BOLD signal would fluctuate as a function of trajectory angle with a six-fold symmetry (cf. figure 4.4), the relative strength of the sine and cosine estimates relative to an arbitrary angle can be used to determine the grid orientation. Note that the sine and cosine regressors were created as functions of six times the angle of the movement in value space and therefore the beta estimates were relative to such six-fold periodicity. The formula employed to estimate the grid orientation φ (modulo 60º) was:

$$\varphi = \text{arctangent}\left(\frac{\beta\ sine}{\beta\ cosine}\right) / 6 \qquad \text{eq. 4.1}$$

[5] The degrees of freedom at the first level depend on the number of volumes acquired; therefore, they vary across sessions. We indicate that df2 was always larger than 207.

with the quadrant of the arctangent adjusted depending on the signs of β sine and β cosine.

The same procedure was repeated twice for odd trials and even trials, giving two independent estimates of the grid orientation, as an angle varying between 0º and 60º. The consistency of these two angles was therefore used as evidence for the hypothesis that location and movement in value space are indeed represented with a grid-like code in the brain. If there was no real pattern, the two estimates would diverge on average by 30º. As for the repetition suppression analysis, we took two approaches to combine results across sessions: first, we averaged the last three sessions of each subject, corresponding to the second half of the novel choice experiment; second, we averaged the consistency results across subjects, separately for each of the five sessions, spanning the full length of the novel choice experiment. This allowed us to test the appearance of a grid-like code through experience in an unbiased way.

4.3 Results

4.3.1 Repetition suppression

Aiming to study the neural representation of individual options, as opposed to the decision process, experiment 2 used a fast sequence of single options presented to the subjects; each one was followed by its probabilistic outcome as long as the animal responded within a short deadline.

We used an fMRI adaptation design to reveal whether the key areas individuated in experiment 1 were encoding the abstract property of expected value of an option, as opposed to its components, namely the magnitude and the probability of reward.

Using the GLM for the repetition suppression analysis described in section 4.2.2, options preceded by identical options were contrasted with options preceded by different options. A whole-brain analysis in all sessions, including the first one before the start of experiment 1, revealed a single cluster, corrected for multiple comparisons, in lateral intraparietal cortex (LIP, coordinates [9 -24 13]) (figure 4.6, top). When restricting the analysis to the last three sequence sessions, corresponding to the second half of experiment 1 (see methods, section 4.2.2), after some exposure to the novel choice task, a new significant cluster was revealed in prefrontal cortex, spanning perigenual ACC, dlPFC and the anterior part of OFC (figure 4.6, bottom). A subthreshold fMRI adaptation was also observed bilaterally near entorhinal cortex and amygdala ($Z>2.32$, $p<0.01$, uncorrected). In all these prefrontal and temporal areas, a linear modulation by session number was identified ($Z>2.32$, $p<0.01$, uncorrected) suggesting a gradual appearance of this neural signature across the novel choice experiment.

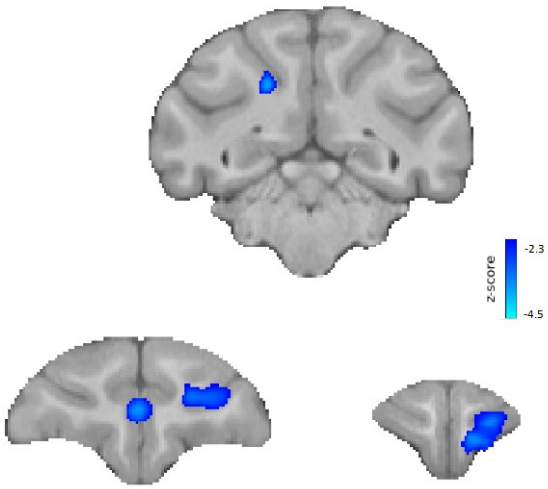


Figure 4.6 Repetition suppression after identical stimuli. **Top:** average effect across the whole experiment. **Bottom:** effect in the second half of the experiment, after experience with the novel task. All images cluster corrected (threshold Z>2.3, p<0.05).

We subsequently contrasted single options preceded by options of similar value (i.e. similar product of magnitude and probability) but different identity, with options preceded by options of different value. This contrast is designed to identify areas concerned with the abstract property of value of an option, independently from its components or its identity. No cluster survived a whole-brain correction for multiple comparisons, but interestingly the strongest subthreshold result was observed in entorhinal cortex (figure 4.7, peak Z=-3.66, coordinates [7 -2 -10]). This did not show variations with experience during the experiment.

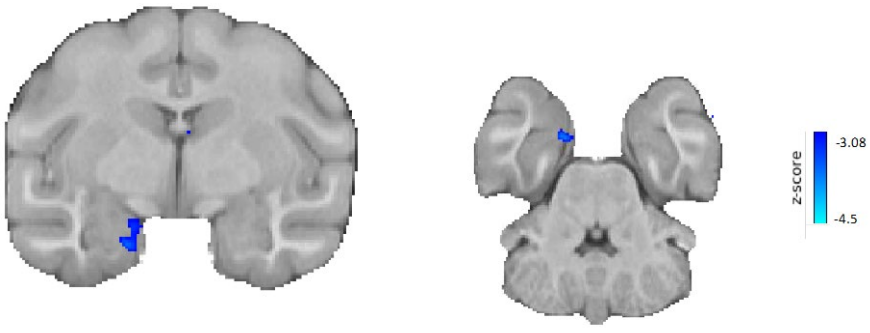


Figure 4.7 Repetition suppression for options of similar value across all sessions of experiment two, Z>3.08, p<0.001, uncorrected.

Because some of the results illustrated so far were not strong enough to survive cluster correction, given the constraints of this experiment discussed below in section 4.4, we subsequently focused the analysis to the regions of interest in prefrontal cortex derived from experiment 1, namely AMFC, dlPFC, striatum, OFC, PMFC and dmPFC. As reported in chapter 3, the first three ROIs had been identified by the contrast of value difference novel versus familiar; OFC was identified by value difference in all conditions; PMFC showed a small linear trend of value difference across sessions and dmPFC showed an effect of chosen value in all conditions. For this repetition suppression analysis, we focused on the last three sequence sessions, ranging through the second half of the novel choice experiment. Among the selected ROIs, we found a significant adaptation effect in two of them: AMFC and dlPFC (figure 4.8, paired t-test p=0.05).

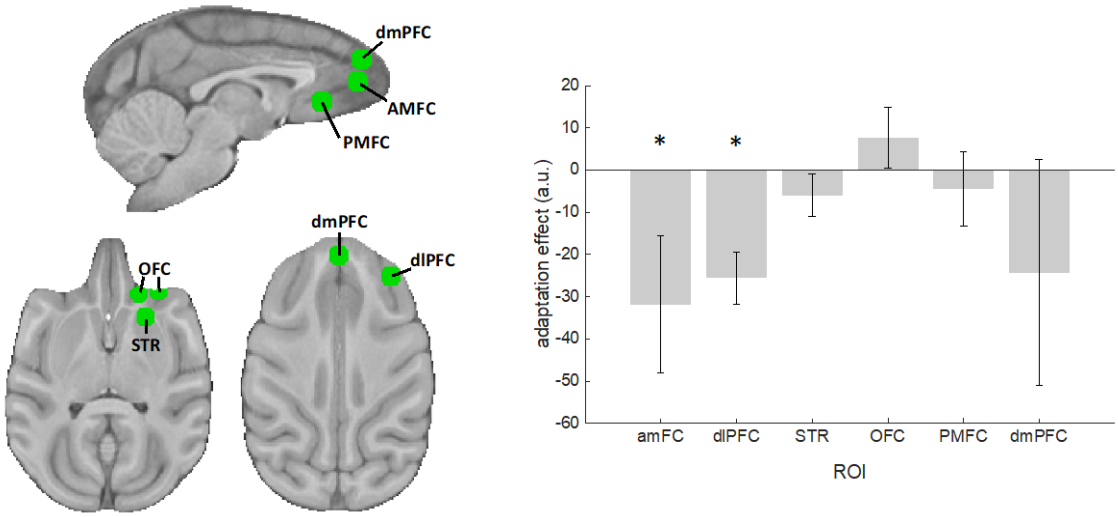


Figure 4.8 Repetition suppression after stimuli of similar value but different component dimensions. **Left:** tested ROIs. **Right:** paired t-tests between betas in similar value trials and in different value trials during the second half of the experiment (df=11, p=0.05).

AMFC and dlPFC, as well as dmPFC, also showed a significant suppression for pairs of identical stimuli and the contrast for same magnitude adaptation was only significant in dlPFC and the contrast for same probability adaptation in AMFC (all paired t-tests df=11, p=0.05).

4.3.2 Grid-like code

A whole brain fMRI analysis searched for a signature of grid-like encoding of the two-dimensional value space used in this task, in the form of a hexagonal symmetry in the neural effects of movement from one option to the next. Again, we focused on the second half of the novel decision-making experiment (sessions 3-5). An F-test identified only two areas in prefrontal cortex with an average F-statistics (df>1, 207) larger than 8 ($p<0.005$, see methods, section 4.2.3): ventral AMFC (overlapping with the AMFC ROI identified in experiment 1, but centred on the novel vs familiar conflicting effect) and right dlPFC (more anterior than the experiment 1 dlPFC ROI). No effect of similar strength was identified in the temporal cortex.

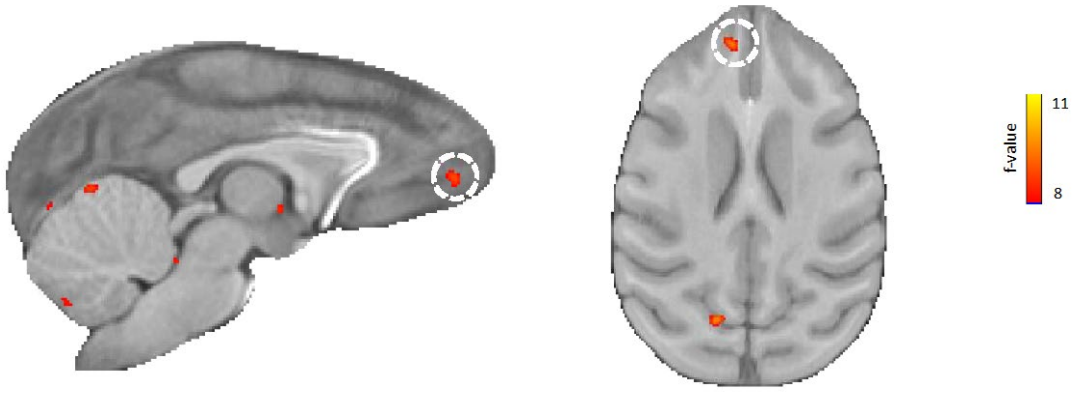


Figure 4.9 Search for a hexagonal symmetry effect of trajectories in value space. F>8, df>(1,207), averaged across sessions, uncorrected.

We further showed that out of these two areas, only AMFC was selectively responding to a six-fold symmetry compared to other periodicities.

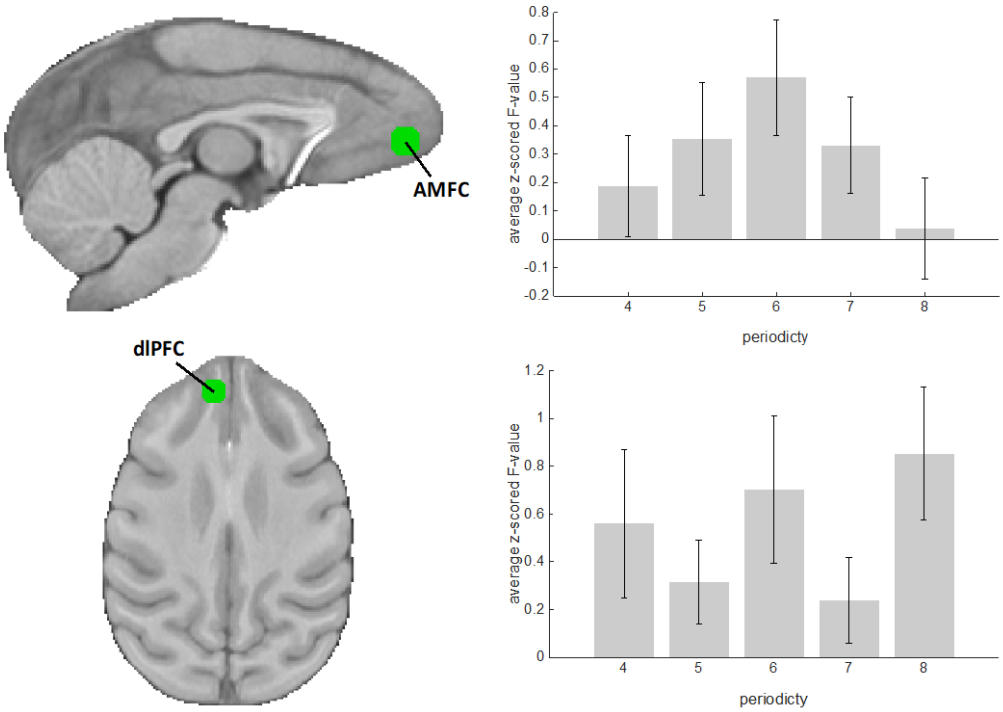


Figure 4.10 Specificity of the hexagonal pattern in AMFC, coordinates [0 26 5] **(top)** but not dlPFC, coordinates [4 25 11] **(bottom)**.

To further validate this result, trials were split in two random subsets (odd and even valid trials) and the orientation of the hexagonal grid was estimated separately in each subset and each session. In AMFC, but not dlPFC, the orientation estimated in odd trials was significantly more similar than chance to the orientation estimated in even trials in the last three sessions (one-sided t-tests, AMFC $t(11)=2.74$, $p<0.01$, dlPFC $t(11)=0.01$, n.s.). Additionally, the grid orientation became more consistent between the two subsets of trials in AMFC, but not dlPFC, with further exposure to the two-dimensional value space during the novel choice experiment. A linear trend was fit for each subject and

tested across subjects, t(3)=-4.2, p<0.05 for AMFC, t(3)=0.73 n.s. for dlPFC. This confirms the presence and the gradual construction of a hexagonal grid-like code in the AMFC ROI with a test that is orthogonal to the way the ROI was selected.

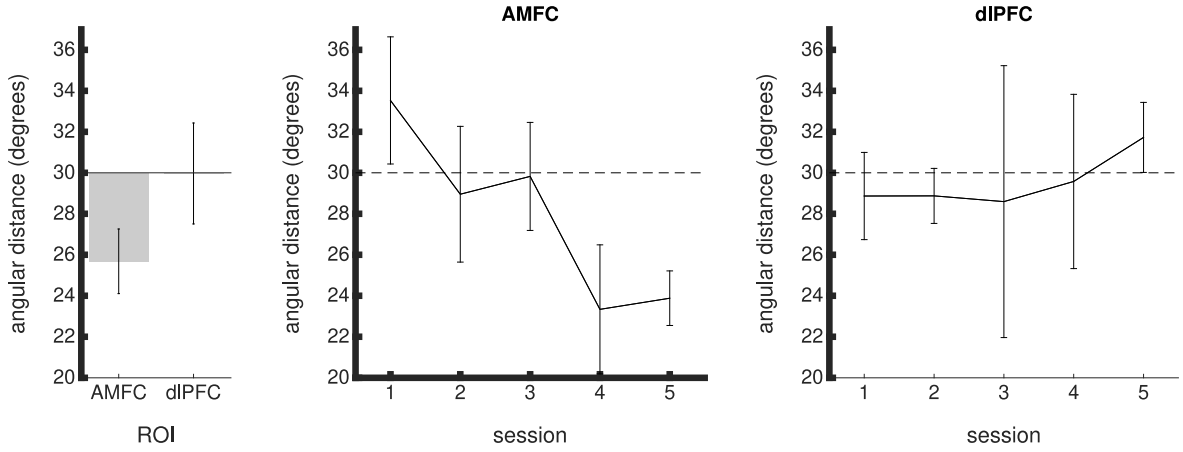


Figure 4.11 Angular difference between grid orientations estimated in odd and even trials separately. Chance level = 30º. **Left:** average of last three sessions in two ROIs. **Centre and right:** trend across the novel experiment in AMFC and dlPFC respectively.

4.4 Discussion

Complementing our investigation of novel decision-making in the brain, experiment 2 brings new insights into how novel complex representations of value are encoded. Employing single rewarding options, in the absence of choice, and focusing on their temporal sequence, we adopted two advanced approaches to fMRI analysis in order to uncover the formation of value representations: repetition suppression and a test for a hexagonal symmetry in response to trajectory angle.

It must be noted that the design of experiment 2 was very challenging because of requirements such as the interval between stimuli and the number of trials, and therefore the signal to noise ratio was relatively low. Nevertheless, effects observed in both the repetition suppression and the grid code analyses were focal and meaningful, consistent with each other and with previous studies, strengthening our confidence in the results of this challenging project. Firstly, in order to maintain the distinction between novel and familiar options in experiment 1, the number of planned trials in the no-choice task of experiment 2 conducted in parallel, was purposely capped to a reasonable minimum. Secondly, adaptation studies require a fast sequence of stimuli; in humans they are generally performed with no intervening event between stimuli; for macaques, however, we had to deliver the reward (probabilistically) between stimuli. Within these limits, we aimed to minimise the interval between stimuli, reaching an average of approximately 4.5 seconds. The downside of short trials was, on the other hand, a larger amount of bodily movement per volume compared to experiment 1, impacting on the homogeneity of the magnetic field. This was nevertheless, we argue, the best possible compromise to implement an innovative design tracking the presence and the development of value representations in non-human primates facing a novel task. For these reasons, most

analyses were not conducted at the whole-brain scale in experiment 2, but in ROIs defined independently either in experiment 1, or by a separate analysis within experiment 2.

Firstly, the only significant trial-to-trial adaptation effect in all five sessions was localised in LIP, which thus appears to be independent from practice with the novel task. This area is known to contribute to choice in both perceptual and value-based decision-making (Platt, Glimcher, 1999; Louie, Glimcher, 2010); however, in the repetition suppression design no choice is involved. It is possible that the effect is linked to a representation of stimulus identity that is automatically triggered even in single option trials in animals accustomed to perform choice tasks on a daily basis. It is also possible that the effect was due to the role of LIP in representing numerosity (Roitman *et al.* 2004), given that stimulus identity is in part determined by the number of dots it contains.

Interestingly, other prefrontal locations, such as OFC, dlPFC and ACC started to display adaptation to identical stimuli only later during the experiment. As discussed above (sections 1.3.3 and 3.4), OFC is known to represent object identity, especially with regard to their sensory properties (Grabenhorst, Rolls, 2011). However, the location identified in experiment 2 is more anterior than the OFC effect for value difference observed in experiment 1. The suppression in dlPFC and ACC following identical stimuli may be a sign that in all other conditions these areas work more to compute a novel representation, and expectation, for a given stimulus, but less so when the stimulus is the same as in the previous trial. The fact that this form of adaptation is not visible in the first few sessions may be an indication that even simple features of a novel stimulus, such as its identity independent from its expected value, may require experience to be processed in full. In particular, it is possible that because the identity of the stimuli in this

experiment was determined by two attributes, some way of integrating these attributes may have still been required in order to react to their identity, and show adaptation.

The focus of the repetition suppression design was nevertheless the same value condition (SV). Stimuli in this category differ in both magnitude and probability from the preceding stimulus, but they have a similar expected value (product of the two). If a subject was not sensitive to the average expected reward, he would not display adaptation for otherwise perceptually different stimuli; however, three areas showed evidence of repetition suppression when stimuli with the same value were presented on successive trials, and these were the entorhinal cortex, AMFC and dlPFC, which, remarkably, are the same locations showing a strong contrast between novel and familiar choice trials in experiment 1.

We have suggested in the previous chapter that AMFC may be the key area computing the motivational value of options, and comparing them. The fact that the same area displays adaptation for similar value stimuli, confirms that AMFC possesses a value-oriented representation of the available option(s), but for novel options this effect is only visible after an exposure of a few days to the novel task. This is a confirmation that novel option representations are formed on the fly and helps explain why in experiment 1 novel, but not familiar, choices induce an enhanced response for comparisons with low value difference, where the similarity of the options plausibly requires a more precise representation of each. dlPFC equally shows adaptation after same value stimuli, confirming the role of the multiple-demand network in the computation of complex and novel representations whether in the context of decision-making or not.

Entorhinal cortex shows repetition suppression for similar value stimuli in all sessions, even the first ones. This confirms its important role in novel decision-making and may be

seen as a sign that it contributes to the encoding of the value of novel options, to be stored in long-term representations, and that it does not need practice with the novel value space to play this role, but this happens from the initial period of exposure to the new stimuli.

It is interesting to note that OFC reveals some effect of identity repetition suppression, but not same value repetition suppression. This is consistent with our hypothesis that OFC represents the association between a stimulus identity and its outcome, but may not have a specific role in the novel choice circuit, that relies on an online estimation of the value through attribute integration.

The second part of this chapter referred to the idea that abstract two-dimensional spaces may be encoded with a grid code, in a similar way as physical space is encoded by grid cells in entorhinal cortex. The grid refers to the multiple locations of the receptive field of the neurons, structured as a triangular grid, and parallel across neurons. Although grid cells have been recorded directly in entorhinal cortex of non-human animals, fMRI studies suggest that they also exist in humans (Doeller *et al.* 2010). The same study, relying on the hexagonal symmetry of the grid, found evidence of a grid-like code in a network of other brain regions, including large parts of the medial frontal cortex. Based on this fMRI approach, and the suggestion by Constantinescu and colleagues (2016) that a grid-like code may be used for non-physical space as well, we tested the hypothesis that value space, constituted by the two dimensions of magnitude and probability of reward, is represented with a grid-like code in medial prefrontal cortex. Remarkably, only one area responded to the F-test for a six-fold symmetry but not for other periodicities, namely AMFC. In a separate analysis, when trials were divided in two subsets, the orientation of the grid estimated in each set separately was more consistent than chance.

These results fit with the assumption that brain organisation is based on functional maps and moreover makes a link with a very specific, computationally well-defined structure, namely the grid code observed in entorhinal cortex. The location of the activation we observed was consistent at the same time with studies of neural representations of value (Clithero, Rangel, 2010) and with studies of grid-like spatial representations (Doeller *et al.* 2010). Not only does this strengthen the credibility of our results, but hints towards a more general theoretical understanding of how complex and novel information about rewarding options is organised in prefrontal cortex. The pace at which value-related adaptation was displayed in AMFC was approximately the same pace at which a grid structure seemed to appear in the region. It seems plausible to assume that a grid-like like code in medial frontal cortex is therefore a candidate mechanism for the ability to integrate multidimensional information and although a population of neurons may have the potential to learn and organise any arbitrary space (such as the "bird space" studied by Constantinescu *et al*. 2016), doing so requires time and training (in total monkeys experienced 17 sessions in experiments 1 and 2 together).

Recent theoretical contributions have suggested that the nature of the computations performed by place and grid cells in medial temporal lobe may not be a unique mechanism dedicated to physical space, but rather an instantiation of a more general algorithm that predicts equally well spatial and conceptual maps, with two or multiple dimensions and varying degrees of density (Mok, Love, 2018). In fact, in a two-dimensional space, a triangular organisation allows the most efficient homogeneous sampling across space, and this predicts specifically a 6-fold symmetry. Any other periodicity would necessarily leave more uncovered space between locations in the receptive fields.

It is intriguing that dlPFC has been implicated by both the repetition suppression and the hexagonal symmetry analyses (albeit in different hemispheres) given its central position in the multiple-demand system (Duncan, 2010). However, it is revealing that four- and eight-fold symmetries matched equally well its pattern of activity. These patterns are consistent with the possibility that it constructs representations of the cardinal directions of the abstract decision space.

Even within AMFC, we do not suggest that a grid-like encoding of the full value space is the only way by which novel decisions are taken, given that the representation may take days to form, whereas choices have a high accuracy level from the first session. It is possible that an understanding of each of the two attributes separately may be sufficient to guide decision-making in most cases, but a full model-based understanding of the task can come only with further practice.

Finally, it should be pointed out that the medial frontal regions activated by the value difference contrast in experiment 1and identified by the F-test for hexagonal symmetry in experiment 2 are very close to each other but may not be identical. The peak activation of value difference in novel versus familiar trials corresponds quite precisely to the average activation of most human fMRI experiments involving value-guided decision-making and is located at the border between perigenual ACC (area 32) and ventromedial prefrontal cortex (area 14). On the other hand, one of the more specific contrasts performed in experiment 1, namely value difference in conflicting novel versus conflicting familiar trials, was centred in a slightly more ventral location, lying entirely in area 14. The importance of conflicting trials is that each option is better according to one but not another attribute, and therefore an integration of both attributes is required in order to identify the better choice. Very remarkably, the same location in area 14 was the peak of the F-test for hexagonal symmetry and was used as an ROI for the further test of

orientation consistency between odd and even trials. The grid code is a candidate mechanism for the integration of multiple dimensions and it has been shown that people with vmPFC lesions are impaired in performing multi-attribute choice (Fellows, 2006).

Although the distinction between these regions should be interpreted with caution, given the degree of spatial resolution that is possible with fMRI, the results obtained from both experiment 1 and experiment 2 are consistent with the hypothesis that a ventral network potentially located in area 14 is instrumental in the integration of multiple attributes, and one mechanism for this would be a grid-like encoding of any abstract space that becomes relevant through experience; this network would therefore determine the value of each available option. On the other hand, a slightly more dorsal network, nearer perigenual ACC, could perform the option comparison in a choice context, or otherwise represent value expectation in a no-choice situation.

Chapter 5
Effects of medial prefrontal ultrasound stimulation on novel multidimensional value integration

5.1 Introduction

In experiments 1 and 2 we have shown results consistent with the hypothesis that anterior medial frontal cortex (AMFC) plays a role in the integration of orthogonal value dimensions into a single abstract value representation in a relatively unbiased fashion and is instrumental in producing adaptive choice behaviour.

While consistent results from different experiments and different analyses strengthen each other; fMRI is essentially a measuring technique, showing correlations between neural activity and modelled cognitive processes. In order to validate the hypothesis that AMFC plays a causal role in the cognitive process required for adaptive multidimensional value-guided choice, we designed a third experiment to disrupt the activity of this area and assess the behavioural change.

In humans, a vmPFC lesion strongly alters the process of multi-attribute integration. Fellows (2006) asked patients with vmPFC or dlPFC lesions and controls to hypothetically choose an apartment to buy. Multiple apartments, with multiple attributes (price, location…) were described in a table. Importantly, attributes were initially hidden but participants could reveal them one by one, by clicking on them, in order to acquire the necessary information to guide the decision. vmPFC patients showed a pattern of exploration markedly different from controls and dlPFC patients: instead of exploring information by apartment (considering each of the attributes of an apartment one after another before moving on to the next apartment), they explored by attribute (they

reviewed a given attribute, such as price, across several apartments before moving on to the next attribute, such as location). While control participants appeared to integrate the different component attributes of a given option into a single evaluation, this was no longer the case after a vmPFC lesion.

In our task, where only two options are presented, varying in two respects, magnitude and probability, we considered the hypothesis that adaptive behaviour requires the integration of the values of the two component dimensions of each option into a single representation of total value. However, if the integration mechanism is disrupted, the subjects may fall back on using two separate representations of the magnitude difference between options and the probability difference between options. Whereas naturally occurring brain lesions in humans may affect large parts of the cortex and are unfortunately not reversible, we induced specific and reversible disruption in macaque brains using focused transcranial ultrasound stimulation (TUS).

Ultrasound has been used to stimulate animal tissue for almost a century (Harvey, 1929). Ultrasound can be used at high intensities to ablate neural tissue (Martin *et al.* 2009) but it can also be used at lower ones to reversibly modulate neural activity (Tufail *et al.* 2011). The first proof of the possibility of reversibly exciting the nervous system with ultrasound was provided by Fry and colleagues (1958); as a mean of neurostimulation, it provides a good compromise between precision and invasiveness compared to other available techniques (Kubanek, 2018). On the one hand, deep brain stimulation can target individual neuronal circuits, but it requires surgery and the implantation of an electrode; on the other hand, transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (tDCS) are not invasive but cannot reach deep areas in species with larger brains, and their spatial resolution is low (above one

centimetre). In contrast, ultrasound can reach deep brain areas with a specificity of possibly few millimetres without requiring a craniotomy (Kubanek, 2018).

Short pulses of transcranial focused ultrasound stimulation (TUS) have been shown to elicit specific neuromodulatory effects in the brain, as assessed by electrophysiological measurements in different species (Velling, Shklyaruk, 1988; Gavrilov *et al.* 1996) including humans (Legon *et al.* 2014), and it can also induce limb movements through the motor cortex (Tufail *et al.* 2010).

The exact means by which TUS modulates neural activity are still the object of ongoing research. At high power (>100 W/cm^2), the impact on the target is mainly mediated by thermal effects; however, neuromodulation can be achieved by low-frequency pulses, typically on the order of 0.2 - 1.5 MHz, for a power <1 W/cm^2 (King *et al.* 2013, Constans *et al.* 2017). At these intensities, thermal effects are virtually absent and mechanical bioeffects or cavitation are believed to be candidate mediators of TUS impact on the neural tissue. It is speculated that ultrasound waves may affect membrane conductance and therefore neuron excitability (Fini, Tyler, 2017; Kubanek, 2018); further research is, however, required.

A recent controversy has been sparked by the report that some ultrasound effects may be mediated by the auditory system (Sato *et al.* 2018; Guo *et al.* 2018); nevertheless, this cannot be the only source of TUS effect: neurostimulation has also been reported in *in vitro* preparations (Tyler *et al.* 2008) and it has been shown to be mediated by mechanoreceptors in the worm *Caenorhabditis elegans* (Kubanek *et al.* 2018). Additionally, as described below, offline effects of TUS have also been reported, that cannot be due to an auditory response (Verhagen *et al.* 2018; Folloni *et al.* 2018).

While the exact mechanism underlying TUS effects is investigated, a number of recent studies targeting the prefrontal cortex in non-human primates have been able to

both induce a connectivity change detected with fMRI (Folloni *et al.* 2018; Verhagen *et al.* 2018) and disrupt higher cognitive functions (Deffieux *et al.* 2013; Fouragnan *et al.* 2018) with spatially specific TUS.

In particular, offline TUS over pre-SMA, frontal pole (Verhagen *et al.* 2018), amygdala and ACC (Folloni *et al.* 2018) in separate sessions, produced lasting changes in neural activity specific to the targeted areas, as assessed by both an fMRI seed-based connectivity analysis and a comparison of the connectional fingerprints (Neubert *et al.* 2015). The connectional fingerprints assessed the functional connectivity of the target area with respect to a number of other areas; a reduction in the coupling implies that the target area has become less responsive to inputs from connected areas. At the same time, histological examination revealed an absence of any damage to the tissue (Verhagen *et al.* 2018).

Deffieux and colleagues (2013) were the first to examine the effect of TUS on complex behaviour in non-human primates. While macaques were performing an antisaccade task (they were rewarded for fixating the location opposite to the one cued), a short TUS pulse was occasionally delivered briefly after the cue. Saccade latency was selectively increased by ipsilateral TUS.

More recently, Fouragnan and colleagues (2018) performed TUS on macaque ACC and then measured performance in a probabilistic binary choice task. They observed an impairment in the ability to translate counterfactual choice values into actual behavioural change.

Importantly, three of the last studies discussed (Folloni *et al.* 2018; Verhagen *et al.* 2018; Fouragnan *et al.* 2018), as well as the present experiment, used an "offline protocol". TUS trains each with a duration of 40s were applied and behavioural or neural data were then collected in the one to two hour period that followed. Moreover, in each

case, the effects were selectively induced by stimulation of a target area but not of a control area. This conclusively excludes the possibility that the effects were mediated by any acoustic artefact associated with the stimulation or via activation of the auditory cortex.

In this chapter, experiment 3 of our research on novel decision-making is presented. We used a similar TUS paradigm to the one employed by Fouragnan and colleagues but we applied it to the area that the earlier fMRI experiments described in chapters 3 and 4 had suggested was critical – anterior medial frontal cortex (AMFC) – and a control area, posterior medial frontal cortex (PMFC), to determine whether AMFC exerted any causal role in the online integration of two dimensions of value during novel choices in the macaque.

5.2 Methods

5.2.1 Experimental design and task

Three of the same Rhesus macaques, monkeys A, B and C, participated in experiment 3. After a break of 140 days (± 30), in which they participated in a different experiment involving an independent cognitive task, the monkeys again underwent extensive training in the familiar version of the task. They performed 11000 familiar trials (± 750) in 52 sessions (± 6). In total, there was an interval of 230 days (± 35) between experiment 2 and experiment 3.

Monkeys were gradually exposed to the equipment and the procedures required for transcranial ultrasound neurostimulation (TUS), for a transition phase of at least ten sessions. They then experienced the full procedure but in the absence of an actual TUS train for at least seven more sessions before the beginning of the experiment. Thus, by the beginning of the experiment, the monkeys were completely familiar with the experimental equipment and procedures.

The experimental setup was the same as described in chapter 2 section 2.2.2 for behavioural training. In experiment 3, both training and testing were performed in a custom-built wooden mock-scanner. Head-fixed animals in a sphynx position in a chair were moved into the mock scanner facing a screen and two custom-built infrared sensors in a dark room. Reward was delivered into the mouth through a plastic spout. Pre-recorded MRI scanner noise was played at a moderate intensity (82dB).

The task was analogous to the task of experiment 1. There were ten levels of reward magnitude, varying from 1 drop to 10 drops of juice and ten levels of reward probability varying between 10% and 100%. Magnitude was cued by colour and probability by number of dots (figure 5.1). However, the monkeys did not experience the full space of

possibilities until the testing phase with TUS. Monkeys were trained in two sub-regions of this space: on even days they were only presented with stimuli with a probability of 60% and a magnitude drawn randomly from one to ten drops (figure 5.1, top left and same figure, centre, vertical grey line); by contrast, on odd days, they were presented only with stimuli with a magnitude of 5 but with a probability that was drawn randomly between 10% and 100% and in steps of 10% (figure 5.1, left bottom and same figure, centre, horizontal grey line). Jointly, these stimuli are defined as the familiar stimuli. Training trials were either forced choices (a single option was presented on the screen) or binary choices of familiar stimuli (animals chose between two options).

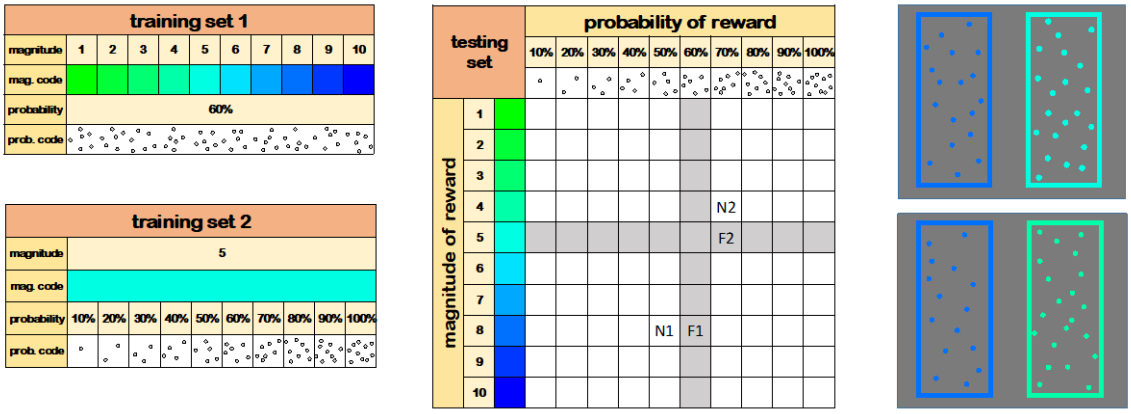


Figure 5.1 Left: Separate training sets for magnitude (top) and probability (bottom) and illustration of the outcome coding. **Centre:** Two-dimensional value space and illustration of the outcome coding. Familiar options in light grey, novel options in white, unused options in dark grey. **Right:** visual appearance of the task. In the top example, a familiar conflicting trial where the left option (F1) gives 8 drops with a 60% chance and the right option (F2) gives 5 drops with a 70% chance. In the bottom example, a novel conflicting trial, where the left option (N1) gives 8 drops with a 50% chance and the right option (N2) gives 4 drops with a 70% chance.

Experiment 3 was based on a choice task with the following timing (figure 5.2): inter-trial interval (ITI) between 3s and 5s; decision between two options with free response

time, and a deadline of 60s; action-outcome delay between 1.5s and 2.5s; reward delivery time of 3s. If monkeys did not respond within the deadline the trial terminated and was repeated at the end of the session. After response, the outcome time was constant for all trials regardless of whether the reward was delivered and how large it was.

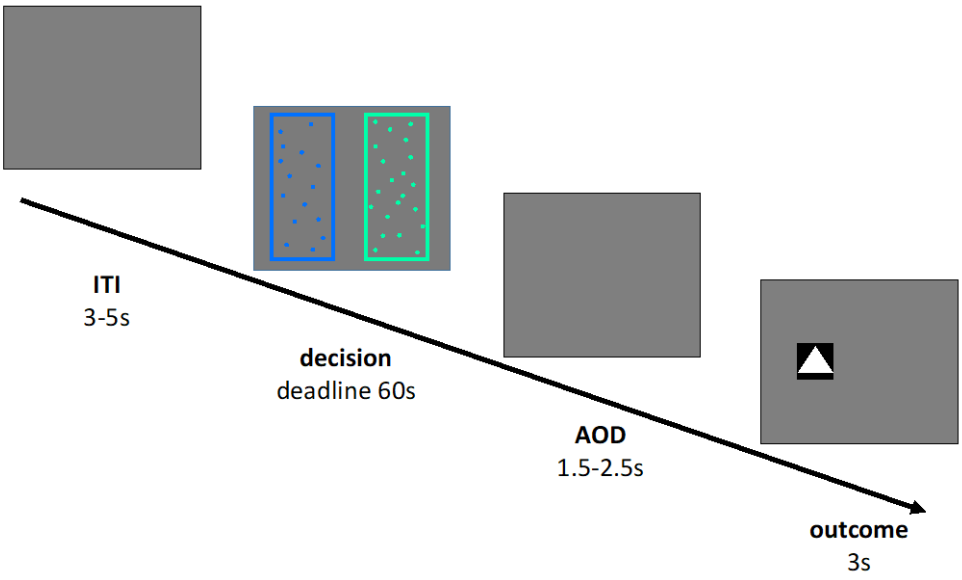


Figure 5.2 TUS task timeline

Sequences included 220 binary choice trials, half of them familiar and half novel and all of them conflicting, meaning that in each trial, the option with higher magnitude had lower probability, and vice-versa. Notice that, on the other hand, training trials were one-dimensional (i.e. they varied in either only probability or only magnitude on a given training day), as in the previous experiments.

The experiment consisted of 12 sessions per subject, in three conditions: TUS of AMFC target, TUS of the control PMFC target, and a further control – a mock TUS procedure (see section 5.2.2, figure 5.4). Therefore, four sessions per condition per subject were acquired, for a total of 36 sessions.

In order to remove all possible confounds related to the precise nature of the stimuli used in the between-session design, only four different stimulus schedules were created, and each schedule was experienced three times by each subject, once in each condition (AMFC target, PMFC target, mock procedure). Therefore, each of the three subjects in each of the three conditions responded to the same four stimulus schedules. Trials in which a response was not given within 60 seconds were interrupted and repeated at the end of the session. All subjects completed all the pre-planned schedules.

The three stimulation conditions were interleaved in pseudo-random order, making sure that no correlation was present between session number and stimulation condition; the four stimulus schedules were also balanced in the same way. Additionally, no two consecutive sessions were assigned to the same condition or schedule.

To avoid after-effects of one session condition on the next, we ran the twelve sessions per subject on alternate days, leaving a minimum of 48 hours before consecutive sessions. Because of external circumstances, monkey C experienced a nine-day gap between the first six sessions and the following six. Overall, the subjects completed the twelve sessions of experiment 3 in 28 days (± 4).

5.2.2 Focused transcranial ultrasound neurostimulation procedure

Before the testing, offline transcranial focused ultrasound stimulation (TUS) was performed. Monkeys sat in a chair with their heads fixed. The normal opening in the chair through which animals made hand responses was closed for the duration of the stimulation procedure. Head registration was based on a previously acquired structural image with fiducial markers; frameless stereotaxic neuronavigation was performed with Brainsight software and a Polaris camera (Rogue Research, Montreal, Canada).

A single-element ultrasound transducer (H115-MR, diameter 64 mm, Sonic Concept, Bothell, WA, USA) with a 51.74 mm focal depth was used with region-specific coupling cones filled with degassed water and sealed with a latex membrane (figure 5.3). The ultrasound wave frequency was set to a 250 kHz resonance frequency and 30 ms bursts of ultrasound were generated every 100 ms with a digital function generator (Handyscope HS5, TiePie engineering, Sneek, the Netherlands). Overall, the stimulation lasted for 40 seconds. A 75-Watt amplifier (75A250A, Amplifier Research, Souderton, PA, USE) was used to deliver the required power to the transducer. A TiePie probe (Handyscope HS5, TiePie engineering, Sneek, The Netherlands) connected to an oscilloscope was used to monitor the voltage delivered. The recorded peak-to-peak voltage was constantly maintained in the range 128-134 V and corresponded to a peak negative pressure of 1.15-1.27 MPa measured in water with an in-house heterodyne interferometer (Constans *et al.* 2017).

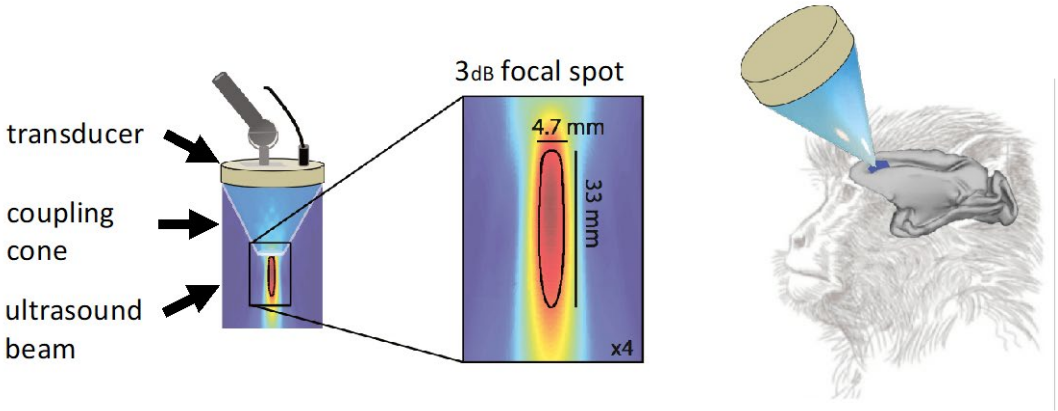


Figure 5.3 Transcranial focused ultrasound stimulation setup and pressure generated. Image (left) adapted from Deffieux *et al.* (2013) with permission from Elsevier; the size of the focal spot may slightly differ in our setup.

Target sites were defined based on the neuroimaging results of experiment 1 (see chapter 3, section 3.3) in two medial prefrontal regions: anterior medial frontal cortex (AMFC) and posterior medial frontal cortex (PMFC) as illustrated in figure 5.4. Each site was targeted four times per subject. In addition, on interleaved days, animals underwent four mock TUS procedures (sham condition); each of these additional sessions were in all respects identical to AMFC sessions but the TUS transducer was turned off.

The neural tissue reached by the ultrasound beam extended necessarily beyond the target sites, but we can safely assume that it was not overlapping between the two conditions. First, the entry point of the beam was different (more dorsal for AMFC than PMFC); second, the size of the coupling cone used meant that the converging focus was more shallow for AMFC than PMFC.

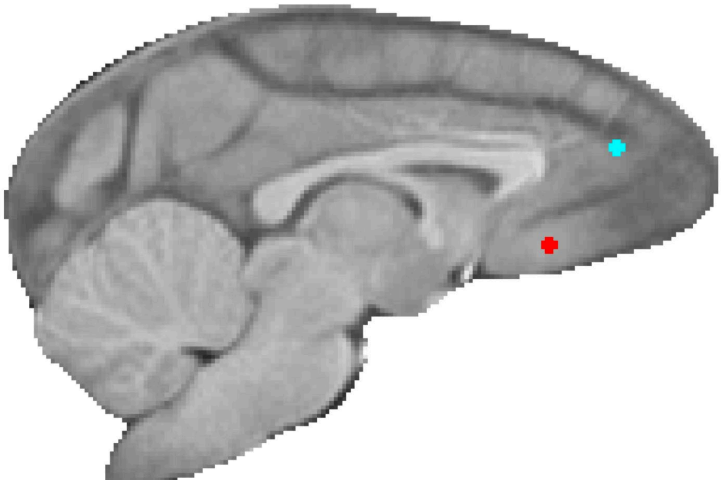


Figure 5.4 TUS target sites: AMFC in light blue, PMFC in red.

Following stimulation, the monkeys were moved to a different room for behavioural testing. The TUS procedure lasted on average 20 minutes; subsequent behavioural testing lasted on average 47 minutes (σ=9). All procedures were conducted under a UK Home

Office licence according to the Animal (Scientific Procedures) Act 1986 and the European Union guidelines (EU Directive 2010/63/EU).

5.2.3 Data analysis and regression simulation

It was hypothesised that, following AMFC, but not PMFC disruption, the subjects' ability to adaptively integrate the two dimensions of value to guide choice would be impaired. As shown in chapter 2, monkeys' decision-making was driven by a combination of two ways of estimating value: (optimal) multiplicative value defined as reward magnitude times probability of an option, and additive value (a heuristic) defined as magnitude plus probability weighted by a coefficient. Decisions driven by additive value can be expressed as decisions driven by a weighted sum of magnitude difference and probability difference, but this heuristic leads to suboptimal choices, especially if some stimuli are very high in one dimension but very low on the other (see figure 2.9, section 2.2.6).

A GLM was therefore created where three regressors predicted choice side:

- the reward magnitude difference between left and right (MAG)
- the reward probability difference between left and right (PROB)
- the optimally integrated value (magnitude times probability) difference between left and right (EXP_VAL)

The relative weight of the third regressor on the choice can be considered a measure of the ability to optimally integrate the two components of value in a single multidimensional construct. The GLM, with parameter estimates constrained to be positive, was fit once for each session separately (12 estimates per condition, each estimate based on 110 novel trials). As a validation, sessions for the same subject and

stimulation condition were merged together (3 estimates per condition, each based on 440 novel trials).

The integration coefficient was defined as the relative size of the expected value coefficient relative to the other beta coefficients:

$$\text{Integration coefficient} = \frac{\beta_{EXP_VAL}}{\beta_{MAG}+\beta_{PROB}+\beta_{EXP_VAL}} \quad \text{(eq. 5.1)}$$

A repeated-measures analysis of variance (ANOVA) with three within-subject conditions: AMFC target, PMFC target and sham stimulation, was used to test for differences in the integration coefficient across conditions.

Additionally, a simulation was run to assess the ability of the GLM described above to correctly recover the integration coefficient. Because trials were all conflicting, a negative correlation between magnitude difference and probability difference was inevitable. In the novel condition, across the 4 stimulus sequences used, the correlations are reported in table 5.1.

Correlation (Pearson's *r*)	Magnitude difference	Probability difference	Expected value difference
Magnitude difference	1		
Probability difference	-0.85	1	
Expected value difference	0.22	0.18	1

Table 5.1 Correlations among regressors in reward dimension integration GLM

Nevertheless, because only the sum of β_{MAG} and β_{PROB} was part of the definition of the integration coefficient (eq. 5.1), their anti-correlation was not expected to prevent the

correct estimation of the overall integration coefficient, as long as the sum of the estimations of the two parameters was reliable.

8000 artificial agents were created, with β_{MAG}, β_{PROB} and β_{EXP_VAL} varying between 0.5 and 10 in 20 steps of 0.5; each one performed the task four times, once for each stimulus schedule, for a total of 32000 datasets. The GLM was then run and the generative parameters were compared to the beta coefficients estimated separately from each artificial dataset. In particular the correlation between the real and the recovered coefficient of integration was assessed.

5.3 Results

As in experiment 1 (chapter 2), performance was at ceiling during most of the training on familiar trials: accuracy was on average 0.86 (σ=0.26), with no extra linear improvement related to practice (t-test of linear trend estimates, df=3, $p>0.05$). Testing performance was also good, considering the challenge posed by the conflicting trials, with an accuracy of 0.77 in familiar and 0.69 in novel trials (σ=0.07), averaging across stimulation conditions. Overall performance was not significantly different across the three stimulation conditions (paired t-tests over GLM coefficient estimates, df=11, $p>0.05$).

In order to test the ability to integrate magnitude and probability dimensions in a single representation, we ran a GLM to estimate how much the combined value difference was driving choice, compared to the two independent value dimensions. We compared stimulation in AMFC, PMFC and no stimulation (see methods, section 5.2.2). In novel trials, a repeated-measures ANOVA showed that the coefficient of integration (eq. 5.1) was significantly different across stimulation conditions $F(2,4)=31.5$, $p<0.01$.

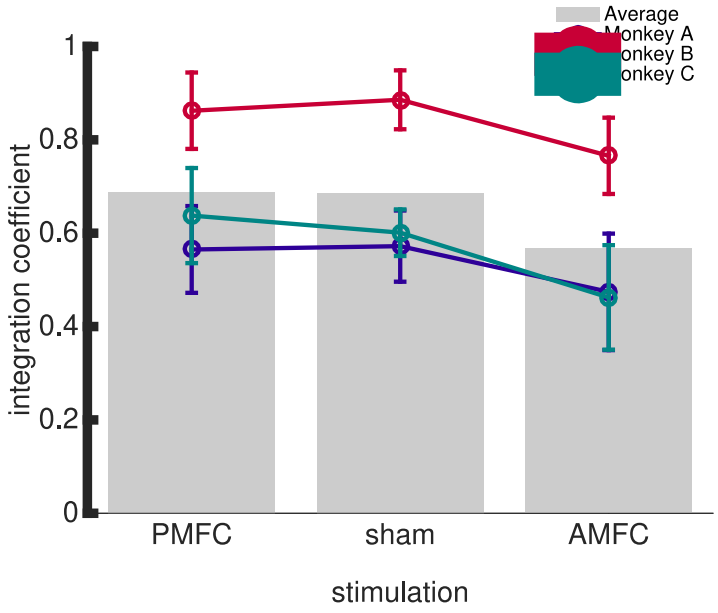


Figure 5.5 Effect of network disruption on optimal integration of magnitude and probability in novel trials. The integration coefficient represents the ratio of the beta coefficient for the expected value regressor (magnitude times probability) to the sum of all three beta coefficients (magnitude, probability and their product). See methods, section 5.2.3.

In the simulation study, the GLM was found to be reliable. When fitting each session (110 trials) separately, the correlation between the real and recovered values of the integration coefficient was equal to 0.64 (Pearson's *r*) for novel trials.

To confirm this result, we merged all trials within each condition and subject (440 trials). The simulation study revealed that the correlation between generative and recovered parameters with these larger sets of trials was 0.87 (Pearson's *r*). A single integration coefficient per subject per condition was estimated from the animal behavioural data and a paired t-test between AMFC and sham conditions confirmed a significant effect of ultrasound $t(2)=5.17$, $p<0.05$.

5.4 Discussion

Experiments 1 and 2 brought converging evidence to the idea that anterior medial frontal cortex (AMFC) is a key region for value-guided decision-making, also in non-human primates, and that it plays a central role in novel decision-making in particular, for both evaluating available options and comparing them. Evidence from different analysis approaches consistently pointed to AMFC; fMRI, however, can only reveal a correlation between a cognitive process and a neural event, but cannot prove the causal link (it cannot exclude the presence of a third common cause). In order to complete the investigations presented in the previous three chapters and determine the causal role of AMFC in novel value-guided decision-making, we ran a neurostimulation experiment to disrupt neural activity in AMFC and another prefrontal region, PMFC, and compare the behavioural performance with control sessions. Because previous studies (Fellows, 2006) and our experiments have implicated medial frontal cortex with multi-attribute integration in particular, we focused on conflicting choices, i.e. choices where focusing on a single attribute cannot allow determination of which is the overall best option.

Although stimulation in both target regions did not strongly disrupt performance and caused inconsistent behavioural effects, a key prediction we made was that the role of AMFC in multi-attribute integration would be affected by TUS. By using a specific definition of the amount of integration (equation 5.1) displayed by the subjects, we observed a clear and consistent effect of AMFC, but not PMFC, disruption compared to sham sessions. The definition was based on the insight that the animals obviously prefer stimuli with higher magnitude, and also stimuli with higher probability, however if the integration mechanism is impaired, they may hold these two representations independently and switch randomly between focusing on the one or on the other. As we

saw in chapter 2 (section 2.2.6), the optimal way of integrating magnitude and probability of reward is to multiply the two; however, some human and animal behaviour may be better described as if they summed the two values instead. Summing the two values is analogous to averaging the magnitude difference and the probability difference, which is a suboptimal strategy compared to first multiplying the attributes and then taking the difference.

It was not possible to compare the additive and the multiplicative models directly, because their predictions are very highly correlated. On the other hand, the three-regressor GLM described in the methods (section 5.2.3) allowed to separate the contribution of the multiplicative factor and the contribution of the two independent dimensions (anti-correlated with each other by design, but not correlated with the product).

Humans and other primates, including our four subjects, display a behaviour that can be described as a combination of the two approaches, both relying on the individual dimensions in isolation but also combining them in an integrative manner. The weight of the suboptimal strategy (without integration) varied across individuals. Experiment 3 shows that, regardless of their normal level of optimal (multiplicative) integration, all subjects show an increased reliance on the suboptimal strategy of considering the two attribute differences separately, after disruption of the AMFC network.

This result clearly proves the causal role of AMFC in novel value-guided decision-making, and also strengthens the more specific hypothesis that AMFC may be instrumental in the attribute integration process and may potentially do this by forming a spatial representation with a grid-like code, in accord with experiments 1 and 2.

Experiment 3 also contributes towards the establishment of transcranial focused ultrasound stimulation (TUS) at low intensities as a reliable scientific tool to induce focal and reversible offline neuromodulation in primates in a safe way.

Chapter 6
General discussion

This series of experiments has revealed the crucial role of medial prefrontal cortex, in particular the area we named AMFC, in novel value-guided decision-making in non-human primates. Through experiment 1 we discovered that the strongest difference in activation by value difference in novel versus familiar choices is centred in AMFC. This seems to imply that while the area plays a role in both types of decision, its role may be different in each case, which in turn may be taken as a confirmation of our initial hypothesis that a specific neural mechanism exists devoted to novel choices (see introduction, section 1.5). At the same time, this reveals a more complex picture; if two separate mechanisms exist, they may be anatomically overlapping, or alternatively the same mechanism may act in different ways. In familiar decisions the representation of the value of each option may be readily available and the role of AMFC may be one of comparing them, in a way that can, for example, be described by a mutual inhibition model (Wang, 2002). However, the role of AMFC in novel decisions is even more crucial: because the value of each option is not available, this may need to be computed on the fly. A potentially distinct neural population within the anterior medial frontal cortex, potentially more ventral than the comparator, is the most likely candidate for the construction of value representations of multi-attribute objects.

Although all decisions require some form of comparison, habitual decisions are more rapid and stereotyped, they are also potentially more prone to biases as suggested in chapter 2. Novel decisions, instead, differ by the additional requirement of comparing the value representations of alternatives that have not been actually experienced before, and

thus such value representations cannot be the outcome of a reinforcement learning process. It would be more appropriate to conceive of them as inferential representations.

Experiment 1 suggests that a network including midbrain, basal ganglia, hippocampus is recruited by familiar value-guided decisions but in a lesser degree by novel decisions; on the other hand, OFC is recruited by both, which means that it encodes the value of the expected outcome but it does not compute it itself, it is therefore unlikely to be a necessary step in the process of making decisions for the first time. Finally, the anterior medial frontal cortex (AMFC) seems to be implicated in both familiar and novel choice but with a clearly different pattern. This may suggest that the area performs the comparison, and the duration of the comparison is influenced by novelty and value difference. The same area, or a neighbouring one, may also compute the value of novel options on the fly. This interplay between constructing a value representation and using it to make a choice, we suggest, is the reason why the effect of value is dissimilar between novel and familiar choices. dlPFC is also activated by the same contrast, and this can be understood as an increased recruitment of the multiple-demand network in novel compared to familiar choices, even with a matched level of difficulty in all other respects.

Experiment 2 confirmed this picture by showing that AMFC and dlPFC after some exposure to the novel task come to encode the expected value of an option, even outside a choice context (adaptation effect). The same experiment also provided evidence for the hypothesis that neurons in AMFC may represent value as a two-dimensional space with a grid-like code. In sum, AMFC may contribute to decision-making for options that have not been learned by reinforcement through a rapid online integration of attributes, which, with practice, may lead to the construction of a consistent grid-like code accompanied by a value representation associated to each new option.

Finally, experiment three validated the causal role of AMFC in novel value-guided decision-making. Moreover, it particularly emphasized the integration of component attributes' values, on the fly, in order to derive an integrated value estimate to drive choice.

This result obtained by the combination of functional neuroimaging and transcranial ultrasound stimulation on awake behaving macaque monkeys achieves a number of goals. Firstly, while novel decisions are routinely employed in human fMRI studies, they have not been studied by themselves, in contrast with familiar decisions. Therefore, assuming a similarity across species, this study brings some light to the nature of human decision-making as well, which is particularly adaptable to novel contexts. Additionally, this study illustrates how Rhesus macaques are able – if the right conditions are provided – to perform flexible decision-making in novel contexts as well. This ability has seldom been exploited in neuroscientific experiments that tended to focus on habitual patterns of decisions.

Based on our results, we can draw some conclusions on the interactions between measuring techniques and species. Electrophysiologists working with macaques have tended to stress the role of OFC in value-guided decision-making, whereas researchers using fMRI with humans have highlighted AMFC, despite these being parts of different anatomical networks. Although it is possible that the two species truly differ in neurocognitive mechanisms underlying choice, the present results suggest a more nuanced interpretation. OFC is indeed much more clearly engaged in the task compared to what is usually found in humans. This may indicate that there is an inter-species difference, or, as we suggested in section 3.4, that it depends on the type of reward used. But this result excludes the possibility that the difference depends on the measuring technique. On the other hand, OFC cannot be part of a mechanism specific to novel

decision-making because it is also activated, and with the same sign, by value difference in familiar decisions.

On the other hand, AMFC was clearly implicated by all three of our experiments, but little evidence of value encoding has been found there in electrophysiological studies (cf. the Introduction, section 1.3.3). This is compatible with the hypothesis that there is a difference in measuring technique, for example fMRI averaging of the metabolic activity of potentially multiple populations of neurons may detect modulations in activity difficult to measure in single cell studies. Alternatively, this may again be an effect of the task design, which means that animal tasks based on very familiar patterns of decisions may miss out the capacity of medial prefrontal neurons to flexibly respond to novel choices. Recording in AMFC with a task based on relatively untrained value-guided decisions may be required to determine whether single neurons in AMFC perform value computations only in novel contexts, which would explain why so far this area has not been highlighted in electrophysiological studies of decision making.

Overall, these experiments strengthen the hypothesis that cognitive and neural mechanisms for value-guided decision-making are highly similar in humans and non-human primates, whereas the differences observed across studies can be more easily explained by different task designs.

This study was limited by a number of factors connected to the specificity of the innovative design we used. Non-human primate studies are always constrained by practical and ethical concerns. Notably, the number of subjects must be kept to a minimum and cannot compare with statistical standards in human psychology.

Whole-brain imaging is particularly challenging in non-human primates because of the large body movements, the smaller brains, and the necessity to provide rewards

between trials. We needed to compromise between the physical requirements of fMRI, the necessities of the experimental design, and the requirements of the monkeys.

The very nature of the contrast between novel and familiar choices required a fine balance between acquiring a sufficient amount of data and preventing the novelty of the options to diminish. Because no piloting was possible and the number of animals was limited, some guesswork was required to plan the number of sessions and the relations between experiment 1 and 2. Additionally, ideally experiment 3 should have been performed on animals never previously exposed to novel choices, but this was not possible and the same animals as experiments 1 and 2 were used, albeit after several months of interval and retraining on just the familiar decision trials in the absence of any novel trials.

Due to the complexity of the transcranial ultrasound technique, and the necessity to keep the number of novel sessions to a minimum, only two areas were targeted in experiment 3, AMFC and a control area. It would be interesting to know what would be the effect of disrupting other areas, such as dlPFC and OFC.

Further research would be required to fully investigate the new and intriguing patterns of results obtained in this work. fMRI is a flexible tool, well-suited to perform whole-brain investigations but it lacks the spatial and temporal resolution of electrophysiology. A single-cell study of novel decision-making in primates would be highly desirable, especially if it could target multiple prefrontal areas such as AMFC, dlPFC and OFC. This would facilitate a clearer understanding of the specific details of the computational process underlying novel inferential choices and their relation to these brain regions.

The evidence we found for hexagonal symmetry in activity in AMFC activity deserves a special mention; future research may describe more precisely the nature of abstract multi-dimensional representations and their link with physical space

representations in the temporal lobe. For example, a three-dimensional value space may be employed to test the flexibility of the neural population underlying value integration to represent spaces with more dimensions.

The observation that performance was high for novel trials even at early stages, before clear evidence of integrated value representations in AMFC, invites further examination of the precise nature of activity patterns in the very first sessions that include novel stimuli. Such a design may only have sufficient power if a larger cohort of subjects is recruited, but it would reveal how early and in which way complete multidimensional representations drive behaviour (in a model-based fashion) and what other heuristics are used as a back-up in their absence.

Notwithstanding its limitations, the series of experiments presented in this thesis has contributed to uncovering a specific neural network for novel value-guided choice, a typical human ability seldom previously investigated and contrasted it with familiar decision-making. It suggests at the same time a mechanistic explanation of how this ability to choose among new items can be achieved, through the use of a grid-like code. In doing so, it may have opened the door to further study of the relation between novel and habitual behaviour in the domain of value-guided choice and contributed to bridging neuroscientific research in animal models and in humans both methodologically and theoretically.

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.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*(4), 268.

Ang, Y. S., Manohar, S., Plant, O., Kienast, A., Le Heron, C., Muhammed, K., ... & Husain, M. (2018). Dopamine Modulates Option Generation for Behavior. *Current Biology*, *28*(10), 1561-1569.

Antzoulatos, E. G., & Miller, E. K. (2011). Differences between neural activity in prefrontal cortex and striatum during learning of novel abstract categories. *Neuron*, *71*(2), 243-249.

Ariely, D. (2008). *Predictably irrational* (p. 20). New York: HarperCollins.

Badre, D. (2008). Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes. *Trends in cognitive sciences*, *12*(5), 193-200.

Badre, D., & D'Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of cognitive neuroscience*, *19*(12), 2082-2099.

Balleine, B. W., Daw, N. D., & O'Doherty, J. P. (2009). Multiple forms of value learning and the function of dopamine. In *Neuroeconomics* (pp. 367-387).

Balleine, B. W., & O'Doherty, J. P. (2010). Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, *35*(1), 48.

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

Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. Journal of Behavioral Decision Making, 16(3), 215-233.

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage*, *76*, 412-427.

Basten, U., Biele, G., Heekeren, H. R., & Fiebach, C. J. (2010). How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Sciences*, *107*(50), 21767-21772.

Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2003). General multilevel linear modeling for group analysis in FMRI. *Neuroimage*, *20*(2), 1052-1063.

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE transactions on medical imaging*, *23*(2), 137-152.

Beckmann, M., Johansen-Berg, H., & Rushworth, M. F. (2009). Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *Journal of Neuroscience*, *29*(4), 1175-1190.

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

Belliveau, J. W., Kennedy, D. N., McKinstry, R. C., Buchbinder, B. R., Weisskoff, R., Cohen, M. S., ... & Rosen, B. R. (1991). Functional mapping of the human visual cortex by magnetic resonance imaging. *Science, 254(5032),* 716-719.

Benjamin Jr, L. T. (2007). *A brief history of modern psychology*. Blackwell publishing.

Bennett, C. M., Miller, M. B., & Wolford, G. L. (2009). Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: an argument for multiple comparisons correction. *Neuroimage*, *47*(Suppl 1), S125.

Bernoulli, D. (1738). Specimen Theoriae Novae de Mensura Sortis. Commentarii academiae scientiarum imperialis Petropolitanae, 5, 175-192. Reprint in *Econometrica 22(1), 23-36, translated by Louise Sommer.*

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *Journal of Neuroscience*, *17*(1), 353-362.

Blakemore, C., & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *The Journal of physiology*, *203*(1), 237-260.

Bloch, F. (1946). Nuclear induction. *Physical review*, *70*(7-8), 460.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review*, *113*(4), 700.

Boorman, E. D., Behrens, T. E., Woolrich, M. W., & Rushworth, M. F. (2009). How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron*, *62*(5), 733-743.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological review, 108(3),* 624.

Bouret, S., & Richmond, B. J. (2010). Ventromedial and orbital prefrontal neurons differentially encode internally and externally driven motivational values in monkeys. *Journal of Neuroscience, 30(25),* 8591-8601.

Broca, P. (1861). Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). *Bulletin et Memoires de la Societe anatomique de Paris*, *6*, 330-357.

Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science, 307(5712),* 1118-1121.

Brun, V. H., Otnæss, M. K., Molden, S., Steffenach, H. A., Witter, M. P., Moser, M. B., & Moser, E. I. (2002). Place cells and place recognition maintained by direct entorhinal-hippocampal circuitry. *Science*, *296*(5576), 2243-2246.

Cai, X., & Padoa-Schioppa, C. (2012). Neuronal encoding of subjective value in dorsal and ventral anterior cingulate cortex. *Journal of Neuroscience*, *32*(11), 3791-3808.

Camilleri, A. R., & Newell, B. R. (2013). Mind the gap? Description, experience, and the continuum of uncertainty in risky choice. In *Progress in brain research* (Vol. 202, pp. 55-71). Elsevier.

Carmichael, S. T., & Price, J. L. (1995). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *Journal of Comparative Neurology*, *363*(4), 615-641.

Carmichael, S. T., & Price, J. L. (1995). Sensory and premotor connections of the orbital and medial prefrontal cortex of macaque monkeys. *Journal of Comparative Neurology*, *363*(4), 642-664.

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

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem.

Chau, B. K., Kolling, N., Hunt, L. T., Walton, M. E., & Rushworth, M. F. (2014). A neural mechanism underlying failure of optimal choice with multiple alternatives. *Nature neuroscience*, *17*(3), 463.

Chau, B. K., Sallet, J., Papageorgiou, G. K., Noonan, M. P., Bell, A. H., Walton, M. E., & Rushworth, M. F. (2015). Contrasting roles for orbitofrontal cortex and amygdala in credit assignment and learning in macaques. *Neuron*, *87*(5), 1106-1118.

Chen, M. K., Lakshminarayanan, V., & Santos, L. R. (2006). How basic are behavioral biases? Evidence from capuchin monkey trading behavior. *Journal of Political Economy*, *114*(3), 517-537.

Chomsky, N. (1959). Review of Verbal Behavior, by B.F. Skinner. *Language*. 35: 26–57.

Cieslik, E. C., Zilles, K., Caspers, S., Roski, C., Kellermann, T. S., Jakobs, O., ... & Eickhoff, S. B. (2012). Is there "one" DLPFC in cognitive action control? Evidence for heterogeneity from co-activation-based parcellation. *Cerebral cortex*, *23*(11), 2677-2689.

Clithero, J. A., & Rangel, A. (2013). Informatic parcellation of the network involved in the computation of subjective value. *Social cognitive and affective neuroscience*, *9*(9), 1289-1302.

Cohen, J. D., McClure, S. M., & Angela, J. Y. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362*(1481), 933-942.

Comte, A. (1853) *The positive philosophy of Auguste Comte*, London: J. Chapman, freely translated and condensed by Harriet Martineau from (1830–1842) *Cours de philosophie positive*, Paris: Rouen first, then Bachelier (6 vols).

Constans, C., Deffieux, T., Pouget, P., Tanter, M., & Aubry, J. F. (2017). A 200–1380-kHz Quadrifrequency Focused Ultrasound Transducer for Neurostimulation in Rodents and Primates: Transcranial In Vitro Calibration and Numerical Study of the Influence of Skull Cavity. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, *64*(4), 717-724.

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

Croxson, P. L., Walton, M. E., O'Reilly, J. X., Behrens, T. E., & Rushworth, M. F. (2009). Effort-based cost–benefit valuation and the human brain. *Journal of Neuroscience*, *29*(14), 4531-4541.

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

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

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

Dayan, P., & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(4), 429-453.

Deffieux, T., Younan, Y., Wattiez, N., Tanter, M., Pouget, P., & Aubry, J. F. (2013). Low-intensity focused ultrasound modulates monkey visuomotor behavior. *Current Biology*, *23*(23), 2430-2433.

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases, and rational decision-making in the human brain. *Science*, *313*(5787), 684-687.

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

Descartes R. (1641) Meditations on First Philosophy. Reprint and English translation by E.S. Haldane and G.R.T. Ross (1993), London: Routledge.

de Visser, L., Homberg, J., Mitsogiannis, M., Zeeb, F., Rivalan, M., Fitoussi, A., ... & Dellu-Hagedorn, F. (2011). Rodent versions of the iowa gambling task: opportunities and challenges for the understanding of decision-making. *Frontiers in neuroscience*, *5*, 109.

Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Phil. Trans. R. Soc. Lond. B*, *308*(1135), 67-78.

Dickinson, A. (1994). Instrumental conditioning. In: N.J. Mackintosh (ed.), *Animal Cognition and Learning*. London: Academic Press, pp. 4–79.

Dickinson, A., & Balleine, B. (1994). Motivational control of goal-directed action. *Animal Learning & Behavior*, *22*(1), 1-18.

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

Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, *80*(2), 312-325.

Domenech, P., Redouté, J., Koechlin, E., & Dreher, J. C. (2017). The neuro-computational architecture of value-based selection in the human brain. *Cerebral Cortex*, *28*(2), 585-601.

Doya, K., Samejima, K., Katagiri, K. I., & Kawato, M. (2002). Multiple model-based reinforcement learning. *Neural computation*, *14*(6), 1347-1369.

Dukas, R. (Ed.). (1998). *Cognitive ecology: the evolutionary ecology of information processing and decision making.* University of Chicago Press.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in cognitive sciences*, *14*(4), 172-179.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in neurosciences*, *23*(10), 475-483.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The quarterly journal of economics*, 643-669.

Farashahi, S., Azab, H., Hayden, B., & Soltani, A. (2018). On the flexibility of basic risk attitudes in monkeys. *Journal of Neuroscience*, 2260-17.

Fellows, L. K. (2006). Deciding how to decide: ventromedial frontal lobe damage affects information acquisition in multi-attribute decision making. *Brain*, *129*(4), 944-952.

Fellows, L. K., & Farah, M. J. (2007). The role of ventromedial prefrontal cortex in decision making: judgment under uncertainty or judgment per se?. *Cerebral Cortex*, *17*(11), 2669-2674.

Fiegenbaum, A., & Thomas, H. (1988). Attitudes toward risk and the risk–return paradox: prospect theory explanations. *Academy of Management journal*, *31*(1), 85-106.

Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012). Prefrontal contributions to metacognition in perceptual decision making. *Journal of Neuroscience*, *32*(18), 6117-6125.

Folloni, D., Verhagen, L., Mars, R. B., Fouragnan, E., Aubry, J. F., Rushworth, M. F., & Sallet, J. (2018). Manipulation of deep brain activity in primates using transcranial focused ultrasound stimulation. *bioRxiv*, 342303.

Fouragnan, E. F., Chau, B. K., Folloni, D., Kolling, N., Verhagen, L., Klein-Flugge, M., ... & Rushworth, M. F. (2018). The macaque anterior cingulate cortex translates counterfactual choice value into actual behavioral change. *bioRxiv*, 336917.

Fry, F. J., Ades, H. W., & Fry, W. J. (1958). Production of reversible changes in the central nervous system by ultrasound. *Science*, *127*(3289), 83-84.

Fyhn, M., Molden, S., Witter, M. P., Moser, E. I., & Moser, M. B. (2004). Spatial representation in the entorhinal cortex. *Science*, *305*(5688), 1258-1264.

Gavrilov, L. R., Tsirulnikov, E. M., & Davies, I. A. I. (1996). Application of focused ultrasound for the stimulation of neural structures. *Ultrasound in Medicine and Biology*, *22*(2), 179-192.

Gazzaniga, M. (1984). Preface. *Handbook of Cognitive Neuroscience*. Springer

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology, 62*, 451-482.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. Evolution and Cognition, OUP.

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

Glimcher, P. W. (2004). *Decisions, uncertainty, and the brain: The science of neuroeconomics*. MIT press.

Glimcher, P. W. (2011). *Foundations of neuroeconomic analysis*. OUP USA.

Glimcher, P. W. (2014). Value-based decision making. In *Neuroeconomics (Second Edition)* (pp. 373-391).

Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in cognitive sciences*, *5*(1), 10-16.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual review of neuroscience*, *30*.

Grabenhorst, F., & Rolls, E. T. (2009). Different representations of relative and absolute subjective value in the human brain. *Neuroimage*, *48*(1), 258-268.

Grabenhorst, F., & Rolls, E. T. (2011). Value, pleasure and choice in the ventral prefrontal cortex. *Trends in cognitive sciences*, *15*(2), 56-67.

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.*, *31*, 359-387.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389-400.

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta psychologica*, *107*(1-3), 293-321.

Guitart-Masip, M., Bunzeck, N., Stephan, K. E., Dolan, R. J., & Düzel, E. (2010). Contextual novelty changes reward representations in the striatum. *Journal of Neuroscience*, *30*(5), 1721-1726.

Guo, H., Hamilton, M., Offutt, S. J., Gloeckner, C. D., Li, T., Kim, Y., ... & Lim, H. H. (2018). Ultrasound produces extensive brain activation via a cochlear pathway. *Neuron*.

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

Hald, A. (2003). *A history of probability and statistics and their applications before 1750* (Vol. 501). John Wiley & Sons.

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

Hare, T. A., Hakimi, S., & Rangel, A. (2014). Activity in dlPFC and its effective connectivity to vmPFC are associated with temporal discounting. *Frontiers in neuroscience*, *8*, 50.

Hare, T. A., Malmaud, J., & Rangel, A. (2011). Focusing attention on the health aspects of foods changes value signals in vmPFC and improves dietary choice. *Journal of Neuroscience*, *31*(30), 11077-11087.

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

Harvey, E. N. (1929). The effect of high frequency sound waves on heart muscle and other irritable tissues. *American Journal of Physiology-Legacy Content*, *91*(1), 284-290.

Hatfield, T., Han, J. S., Conley, M., Gallagher, M., & Holland, P. (1996). Neurotoxic lesions of basolateral, but not central, amygdala interfere with Pavlovian second-order conditioning and reinforcer devaluation effects. *Journal of Neuroscience*, *16*(16), 5256-5265.

Hayden, B. Y., Heilbronner, S. R., Nair, A. C., & Platt, M. L. (2008). Cognitive influences on risk-seeking by rhesus macaques. *Judgment and decision making*, *3*(5), 389.

Hayden, B. Y., Pearson, J. M., & Platt, M. L. (2011). Neuronal basis of sequential foraging decisions in a patchy environment. Nature neuroscience, 14(7), 933.

Hayden, B. Y., & Platt, M. L. (2007). Temporal discounting predicts risk sensitivity in rhesus macaques. *Current Biology*, *17*(1), 49-53.

Heilbronner, S. R. (2017). Modeling risky decision-making in nonhuman animals: shared core features. *Current opinion in behavioral sciences*, *16*, 23-29.

Heilbronner, S., & Hayden, B. (2013). Contextual factors explain risk-seeking preferences in rhesus monkeys. *Frontiers in Neuroscience*, *7*, 7.

Heilbronner, S. R., & Hayden, B. Y. (2016). The description-experience gap in risky choice in nonhuman primates. *Psychonomic bulletin & review*, *23*(2), 593-600.

Heilbronner, S. R., Rosati, A. G., Stevens, J. R., Hare, B., & Hauser, M. D. (2008). A fruit in the hand or two in the bush? Divergent risk preferences in chimpanzees and bonobos. *Biology Letters*, *4*(3), 246-249.

Hergenhahn, B. R., & Henley, T. (2013). *An introduction to the history of psychology*. Cengage Learning.

Herrnstein, R. J. (2000). *The matching law: Papers in psychology and economics*. Harvard University Press.

Holland, P. C. (2004). Relations between Pavlovian-instrumental transfer and reinforcer devaluation. *Journal of Experimental Psychology: Animal Behavior Processes*, *30*(2), 104.

Huettel, S. A., Song, A. W., & McCarthy, G. (2008). *Functional magnetic resonance imaging*. Sunderland, MA: Sinauer Associates.

Hunt, L. T., Behrens, T. E., Hosokawa, T., Wallis, J. D., & Kennerley, S. W. (2015). Capturing the temporal evolution of choice across prefrontal cortex. *Elife*, *4*, e11945.

Hunt, L. T., & Hayden, B. Y. (2017). A distributed, hierarchical and recurrent framework for reward-based choice. *Nature Reviews Neuroscience*, *18*(3), 172.

Hunt, L. T., Kolling, N., Soltani, A., Woolrich, M. W., Rushworth, M. F., & Behrens, T. E. (2012). Mechanisms underlying cortical activity during value-guided choice. *Nature neuroscience*, *15*(3), 470.

Hutcherson, C. A., Plassmann, H., Gross, J. J., & Rangel, A. (2012). Cognitive regulation during decision making shifts behavioral control between ventromedial and dorsolateral prefrontal value systems. *Journal of Neuroscience*, *32*(39), 13543-13554.

James, W. (1890) *The principles of psychology*. New York: Holt.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*(2), 825-841.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). Fsl. *Neuroimage*, *62*(2), 782-790.

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical image analysis*, *5*(2), 143-156.

Jueptner, M., Frith, C. D., Brooks, D. J., Frackowiak, R. S. J., & Passingham, R. E. (1997). Anatomy of motor learning. II. Subcortical structures and learning by trial and error. *Journal of neurophysiology*, *77*(3), 1325-1337.

Kable, J. W., & Glimcher, P. W. (2009). The neurobiology of decision: consensus and controversy. *Neuron*, *63*(6), 733-745.

Kacelnik, A. (2006). Meanings of rationality. *Rational animals*, 87-106.

Kacelnik, A., & Bateson, M. (1996). Risky theories—the effects of variance on foraging decisions. *American Zoologist*, *36*(4), 402-434.

Kaelbling, L. P. (1993). *Learning in embedded systems*. MIT press.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263-292.

Kahnt, T., Chang, L. J., Park, S. Q., Heinzle, J., & Haynes, J. D. (2012). Connectivity-based parcellation of the human orbitofrontal cortex. *Journal of Neuroscience, 32(18),* 6240-6250.

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

Kaiser, S., Simon, J. J., Kalis, A., Schweizer, S., Tobler, P. N., & Mojzisch, A. (2013). The cognitive and neural basis of option generation and subsequent choice. Cognitive, Affective, & Behavioral Neuroscience, 13(4), 814-829.

Kastrup, A., Baudewig, J., Schnaudigel, S., Huonker, R., Becker, L., Sohns, J. M., ... & Witte, O. W. (2008). Behavioral correlates of negative BOLD signal changes in the primary somatosensory cortex. Neuroimage, 41(4), 1364-1371.

Kennerley, S. W., Behrens, T. E., & Wallis, J. D. (2011). Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. Nature neuroscience, 14(12), 1581.

Kikuchi, Y., Sedley, W., Griffiths, T. D., & Petkov, C. I. (2018). Evolutionarily conserved neural signatures involved in sequencing predictions and their relevance for language. *Current Opinion in Behavioral Sciences*, *21*, 145-153.

Kirk, G. S. (Ed.). (1954). *Heraclitus: The cosmic fragments*. Cambridge University Press.

Kolling, N., Behrens, T. E., Mars, R. B., & Rushworth, M. F. (2012). Neural mechanisms of foraging. Science, 336(6077), 95-98.

Kolling, N., Wittmann, M. K., Behrens, T. E., Boorman, E. D., Mars, R. B., & Rushworth, M. F. (2016). Value, search, persistence and model updating in anterior cingulate cortex. Nature neuroscience, 19(10), 1280.

Kolling, N., Wittmann, M., & Rushworth, M. F. (2014). Multiple neural mechanisms of decision making and their competition under changing risk pressure. Neuron, 81(5), 1190-1202.

Kolster, H., Mandeville, J. B., Arsenault, J. T., Ekstrom, L. B., Wald, L. L., & Vanduffel, W. (2009). Visual field map clusters in macaque extrastriate visual cortex. *Journal of Neuroscience*, *29*(21), 7031-7039.

Kouneiher, F., Charron, S., & Koechlin, E. (2009). Motivation and cognitive control in the human prefrontal cortex. *Nature neuroscience*, *12*(7), 939.

Kourtzi, Z., & Grill-Spector, K. (2005). fMRI adaptation: a tool for studying visual representations in the primate brain. *Fitting the mind into the world: Adaptation and after-effects in high level vision*, 173-88.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature neuroscience*, *13*(10), 1292.

Krebs, J. R., & Davies, N. B. (Eds.). (2009). *Behavioural ecology: an evolutionary approach*. John Wiley & Sons.

Kreps, D. M. (1990). *A course in microeconomic theory* (Vol. 41). Princeton, NJ: Princeton university press.

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

Kubanek, J. (2018). Neuromodulation with transcranial focused ultrasound. *Neurosurgical focus*, *44*(2), E14.

Kubanek, J., Shukla, P., Das, A., Baccus, S. A., & Goodman, M. B. (2018). Ultrasound elicits behavioral responses through mechanical effects on neurons and ion channels in a simple nervous system. *Journal of Neuroscience*, 1458-17.

Lauterbur, P. C. (1973). Image formation by induced local interactions: examples employing nuclear magnetic resonance.

Leathers, M. L., & Olson, C. R. (2012). In monkeys making value-based decisions, LIP neurons encode cue salience and not action value. *Science*, *338*(6103), 132-135.

Lebreton, M., Jorge, S., Michel, V., Thirion, B., & Pessiglione, M. (2009). An automatic valuation system in the human brain: evidence from functional neuroimaging. *Neuron*, *64*(3), 431-439.

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

Legon, W., Sato, T. F., Opitz, A., Mueller, J., Barbour, A., Williams, A., & Tyler, W. J. (2014). Transcranial focused ultrasound modulates the activity of primary somatosensory cortex in humans. *Nature neuroscience*, *17*(2), 322.

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: a neural common currency for choice. *Current opinion in neurobiology*, *22*(6), 1027-1038.

Louie, K., & Glimcher, P. W. (2010). Separating value from choice: delay discounting activity in the lateral intraparietal area. *Journal of Neuroscience*, *30*(16), 5498-5507.

Ludvig, E. A., & Spetch, M. L. (2011). Of black swans and tossed coins: is the description-experience gap in risky choice limited to rare events?. *PloS one*, *6*(6), e20262.

MacLean, E. L., Mandalaywala, T. M., & Brannon, E. M. (2012). Variance-sensitive choice in lemurs: constancy trumps quantity. *Animal cognition*, *15*(1), 15-25.

Macmillan, M. (2002). *An odd kind of fame: Stories of Phineas Gage*. MIT Press.

Mackey, S., & Petrides, M. (2010). Quantitative demonstration of comparable architectonic areas within the ventromedial and lateral orbital frontal cortex in the human and the macaque monkey brains. *European Journal of Neuroscience*, *32*(11), 1940-1950.

Malonek, D., & Grinvald, A. (1996). Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy: implications for functional brain mapping. *Science*, *272*(5261), 551-554.

Mandeville, J. B., Marota, J. J., Ayata, C., Moskowitz, M. A., Weisskoff, R. M., & Rosen, B. R. (1999). MRI measurement of the temporal evolution of relative CMRO2 during rat forepaw stimulation. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *42*(5), 944-951.

Mandler, G. (2011). *A history of modern experimental psychology: From James and Wundt to cognitive science*. MIT press.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. MIT Press. *Cambridge, Massachusetts*.

Martin, E., Jeanmonod, D., Morel, A., Zadicario, E., & Werner, B. (2009). High-intensity focused ultrasound for noninvasive functional neurosurgery. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *66*(6), 858-861.

Marx, K. (1867) *Das Kapital: Kritik der politischen Oekonomie*. Hamburg: Verlag von Otto Meissner. English translation by Samuel Moore and Edward Aveling. Reprint: (1983-84) *Capital: A Critique of Political Economy* (3 vols.), Lawrence & Wishart, London.

Matsumoto, K., Suzuki, W., & Tanaka, K. (2003). Neuronal correlates of goal-based motor selection in the prefrontal cortex. *Science*, *301*(5630), 229-232.

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

McCoy, A. N., & Platt, M. L. (2005). Risk-sensitive neurons in macaque posterior cingulate cortex. *Nature neuroscience*, *8*(9), 1220.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 115-133.

McDannald, M. A., Lucantonio, F., Burke, K. A., Niv, Y., & Schoenbaum, G. (2011). Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. *Journal of Neuroscience*, *31*(7), 2700-2705.

Meulders, Michel, 2010, *Helmholtz: From Enlightenment to Neuroscience*, Cambridge, MA: The MIT Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.

Miller, N. E. (1985). The value of behavioral research on animals. *American Psychologist*, *40*(4), 423.

Mishkin, M., Malamut, B., & Bachevalier, J. (1984). Memories and habits: Two neural systems. *Neurobiology of learning and memory*, 65-77.

Miyachi, S., Lu, X., Inoue, S., Iwasaki, T., Koike, S., Nambu, A., & Takada, M. (2005). Organization of multisynaptic inputs from prefrontal cortex to primary motor cortex as revealed by retrograde transneuronal transport of rabies virus. *Journal of Neuroscience*, *25*(10), 2547-2556.

Mitchell, D. J., Bell, A. H., Buckley, M. J., Mitchell, A. S., Sallet, J., & Duncan, J. (2016). A putative multiple-demand system in the macaque brain. *Journal of Neuroscience*, *36*(33), 8574-8585.

Miyamoto, K., Osada, T., Setsuie, R., Takeda, M., Tamura, K., Adachi, Y., & Miyashita, Y. (2017). Causal neural network of metamemory for retrospection in primates. *Science*, *355*(6321), 188-193.

Mok, R. M., & Love, B. C. (2018). A non-spatial account of place and grid cells based on clustering models of concept learning. *bioRxiv*, 421842.

Morgenstern, O., & Von Neumann, J. (1953). *Theory of games and economic behavior*. Princeton university press.

Morrison, S. E., & Salzman, C. D. (2011). Representations of appetitive and aversive information in the primate orbitofrontal cortex. *Annals of the New York Academy of Sciences*, *1239*(1), 59-70.

Nestler, E. J., & Hyman, S. E. (2010). Animal models of neuropsychiatric disorders. *Nature neuroscience*, *13*(10), 1161.

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

Niehans, J. (1990). *A history of economic theory: Classic contributions, 1720-1980.* Baltimore: Johns Hopkins University Press.

Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology*, *191*(3), 507-520.

Noonan, M. P., Chau, B., Rushworth, M. F., & Fellows, L. K. (2017). Contrasting effects of medial and lateral orbitofrontal cortex lesions on credit assignment and decision making in humans. *Journal of Neuroscience*, 0692-17.

Noonan, M. P., Walton, M. E., Behrens, T. E. J., Sallet, J., Buckley, M. J., & Rushworth, M. F. S. (2010). Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proceedings of the National Academy of Sciences*, 201012246.

O'Doherty, J. P. (2014). The problem with value. *Neuroscience & Biobehavioral Reviews*, *43*, 259-268.

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

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, *87*(24), 9868-9872.

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

Orban, G. A., Van Essen, D., & Vanduffel, W. (2004). Comparative mapping of higher visual areas in monkeys and humans. *Trends in cognitive sciences, 8(7),* 315-324.

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural computation, 18(2),* 283-328.

Ostlund, S. B., & Balleine, B. W. (2007). Orbitofrontal cortex mediates outcome encoding in Pavlovian but not instrumental conditioning. *Journal of Neuroscience*, *27*(18), 4819-4825.

Padoa-Schioppa, C. (2011). Neurobiology of economic choice: a good-based model. *Annual review of neuroscience*, *34*, 333-359.

Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, *441*(7090), 223.

Padoa-Schioppa, C., & Assad, J. A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nature neuroscience*, *11*(1), 95.

Paglieri, F., Addessi, E., De Petrillo, F., Laviola, G., Mirolli, M., Parisi, D., ... & Adriani, W. (2014). Nonhuman gamblers: lessons from rodents, primates, and robots. *Frontiers in behavioral neuroscience*, *8*, 33.

Papageorgiou, G. K., Sallet, J., Wittmann, M. K., Chau, B. K., Schüffelgen, U., Buckley, M. J., & Rushworth, M. F. (2017). Inverted activity patterns in ventromedial prefrontal cortex during value-guided decision-making in a less-is-more task. *Nature communications*, *8*(1), 1886.

Parkinson, J. A., Dalley, J. W., Cardinal, R. N., Bamford, A., Fehnert, B., Lachenal, G., ... & Everitt, B. J. (2002). Nucleus accumbens dopamine depletion impairs both acquisition and performance of appetitive Pavlovian approach behaviour: implications for mesoaccumbens dopamine function. *Behavioural brain research*, *137*(1-2), 149-163.

Pascal B. (1670) Pensées. Reprint and English translation by A.J. Krailsheimer (1995), Penguin Editions.

Passingham, R. (2009). How good is the macaque monkey model of the human brain?. *Current opinion in neurobiology*, *19*(1), 6-11.

Passingham, R. E., & Wise, S. P. (2012). *The neurobiology of the prefrontal cortex: anatomy, evolution, and the origin of insight* (No. 50). Oxford University Press.

Pauling, L., & Coryell, C. D. (1936). The magnetic properties and structure of hemoglobin, oxyhemoglobin and carbonmonoxyhemoglobin. *Proceedings of the National Academy of Sciences*, *22*(4), 210-216.

Pavlov, I. P. (1927). Conditional reflexes: an investigation of the physiological activity of the cerebral cortex.

Pearson, J. M., Watson, K. K., & Platt, M. L. (2014). Decision making: the neuroethological turn. *Neuron*, *82*(5), 950-965.

Petrides, M., & Pandya, D. N. (1999). Dorsolateral prefrontal cortex: comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. *European Journal of Neuroscience*, *11*(3), 1011-1036.

Plassmann, H., O'Doherty, J., & Rangel, A. (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *Journal of neuroscience*, *27*(37), 9984-9988.

Plassmann, H., O'Doherty, J. P., & Rangel, A. (2010). Appetitive and aversive goal values are encoded in the medial orbitofrontal cortex at the time of decision making. *Journal of neuroscience*, *30*(32), 10799-10808.

Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, *400*(6741), 233.

Platt, M. L., & Huettel, S. A. (2008). Risky business: the neuroeconomics of decision making under uncertainty. *Nature neuroscience*, *11*(4), 398.

Plous, S. (1993). *The psychology of judgment and decision making*. Mcgraw-Hill Book Company.

Poldrack, R. A., & Packard, M. G. (2003). Competition among multiple memory systems: converging evidence from animal and human brain studies. *Neuropsychologia*, *41*(3), 245-251.

Posner, M. I., & DiGirolamo, G. J. (2000). Cognitive neuroscience: origins and promise. *Psychological bulletin*, *126*(6), 873.

Prelec, D. (1998). The probability weighting function. *Econometrica*, 497-527.

Purcell, E. M., Torrey, H. C., & Pound, R. V. (1946). Resonance absorption by nuclear magnetic moments in a solid. *Physical review*, *69*(1-2), 37.

Pusey, E., Lufkin, R. B., Brown, R. K., Solomon, M. A., Stark, D. D., Tarr, R. W., & Hanafee, W. N. (1986). Magnetic resonance imaging artifacts: mechanism and clinical significance. *Radiographics*, *6*(5), 891-911.

Rabin, M., & Thaler, R. H. (2001). Anomalies: risk aversion. *Journal of Economic perspectives*, *15*(1), 219-232.

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature reviews neuroscience*, *9*(7), 545.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, *20*(4), 873-922.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, *2*, 64-99.

Rhodes, S. E., & Murray, E. A. (2013). Differential effects of amygdala, orbital prefrontal cortex, and prelimbic cortex lesions on goal-directed behavior in rhesus macaques. *Journal of Neuroscience*, *33*(8), 3380-3389.

Ridderinkhof, K. R., Van Den Wildenberg, W. P., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and cognition*, *56*(2), 129-140.

Roitman, J. D., Brannon, E. M., & Platt, M. L. (2007). Monotonic coding of numerosity in macaque lateral intraparietal area. *PLoS biology*, *5*(8), e208.

Rolls, E. T. (2007). Memory, attention, and decision-making: a unifying computational neuroscience approach, Oxford Univeristy Press.

Rolls, E. (2008). Functions of the orbitofrontal and pregenual cingulate cortex in taste, olfaction, appetite and emotion. *Acta Physiologica Hungarica*, *95*(2), 131-164.

Rudebeck, P. H., & Murray, E. A. (2011). Dissociable effects of subtotal lesions within the macaque orbital prefrontal cortex on reward-guided behavior. *Journal of Neuroscience*, *31*(29), 10569-10578.

Rudebeck, P. H., & Murray, E. A. (2014). The orbitofrontal oracle: cortical mechanisms for the prediction and evaluation of specific behavioral outcomes. *Neuron*, *84*(6), 1143-1156.

Rudebeck, P. H., Walton, M. E., Smyth, A. N., Bannerman, D. M., & Rushworth, M. F. (2006). Separate neural pathways process different decision costs. *Nature neuroscience*, *9*(9), 1161.

Ruff, C. C., Driver, J., & Bestmann, S. (2009). Combining TMS and fMRI: from 'virtual lesions' to functional-network accounts of cognition. *Cortex, 45(9)*, 1043-1049.

Ruff, C. C., Ugazio, G., & Fehr, E. (2013). Changing social norm compliance with noninvasive brain stimulation. *Science*, *342*(6157), 482-484.

Rushworth, M. F. S., Behrens, T. E. J., Rudebeck, P. H., & Walton, M. E. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. *Trends in cognitive sciences*, *11*(4), 168-176.

Rushworth, M. F., Kolling, N., Sallet, J., & Mars, R. B. (2012). Valuation and decision-making in frontal cortex: one or many serial or parallel systems?. Current opinion in neurobiology, 22(6), 946-955.

Rushworth, M. F., Noonan, M. P., Boorman, E. D., Walton, M. E., & Behrens, T. E. (2011). Frontal cortex and reward-guided learning and decision-making. *Neuron*, *70*(6), 1054-1069.

Rushworth, M. F. S., Walton, M. E., Kennerley, S. W., & Bannerman, D. M. (2004). Action sets and decisions in the medial frontal cortex. *Trends in cognitive sciences*, *8*(9), 410-417.

Saleem, K. S., Kondo, H., & Price, J. L. (2008). Complementary circuits connecting the orbital and medial prefrontal networks with the temporal, insular, and opercular cortex in the macaque monkey. *Journal of Comparative Neurology*, *506*(4), 659-693.

Saleem, K. S., Miller, B., & Price, J. L. (2014). Subdivisions and connectional networks of the lateral prefrontal cortex in the macaque monkey. *Journal of Comparative Neurology*, *522*(7), 1641-1690.

Sallet, J., Mars, R. B., Noonan, M. P., Neubert, F. X., Jbabdi, S., O'Reilly, J. X., ... & Rushworth, M. F. (2013). The organization of dorsal frontal cortex in humans and macaques. *Journal of Neuroscience*, *33*(30), 12255-12274.

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, *5*(17), 61-71.

San-Galli, A., Varazzani, C., Abitbol, R., Pessiglione, M., & Bouret, S. (2016). Primate ventromedial prefrontal cortex neurons continuously encode the willingness to engage in reward-directed behavior. *Cerebral Cortex*, *28*(1), 73-89.

Santos, L. R., & Platt, M. L. (2014). Evolutionary anthropological insights into neuroeconomics: What non-human primates can tell us about human decision-making strategies. In *Neuroeconomics (Second Edition)* (pp. 109-122).

Sato, T., Shapiro, M. G., & Tsao, D. Y. (2018). Ultrasonic neuromodulation causes widespread cortical activation via an indirect auditory mechanism. *Neuron*.

Scholl, J., Günthner, J., Kolling, N., Favaron, E., Rushworth, M. F., Harmer, C. J., & Reinecke, A. (2014). A role beyond learning for NMDA receptors in reward-based decision-making—a pharmacological study using d-cycloserine. *Neuropsychopharmacology*, *39*(12), 2900.

Scholl, J., Kolling, N., Nelissen, N., Wittmann, M. K., Harmer, C. J., & Rushworth, M. F. (2015). The good, the bad, and the irrelevant: neural mechanisms of learning real and hypothetical rewards and effort. *Journal of Neuroscience*, *35*(32), 11233-11251.

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

Semendeferi, K., Armstrong, E., Schleicher, A., Zilles, K., & Van Hoesen, G. W. (2001). Prefrontal cortex in humans and apes: a comparative study of area 10. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *114*(3), 224-241.

Sescousse, G., Redouté, J., & Dreher, J. C. (2010). The architecture of reward value coding in the human orbitofrontal cortex. *Journal of neuroscience*, *30*(39), 13095-13104.

Shadmehr, R., de Xivry, J. J. O., Xu-Wilson, M., & Shih, T. Y. (2010). Temporal discounting of reward and the cost of time in motor control. *Journal of Neuroscience*, *30*(31), 10507-10516.

Shafer, G. (2004). St. Petersburg Paradox. *Encyclopedia of statistical sciences*, *13*. Edited by S. Kotz. John Wiley & Sons.

Simon, H. A. (1972). Theories of bounded rationality. *Decision and organization*, *1*(1), 161-176.

Simon, H. A. (1986). Rationality in psychology and economics. *Journal of Business*, S209-S224.

Sinha, N., Manohar, S., & Husain, M. (2013). Impulsivity and apathy in Parkinson's disease. *Journal of neuropsychology*, *7*(2), 255-283.

Skinner, B. F. (1938). The behavior of organisms: an experimental analysis. New York: Appleton-Century.

Skinner, B. F. (1974). About behaviorism. New York: Alfred Knopf.

Smith, C. U. M. (2014). The 'Hard Problem' and the Cartesian Strand in British Neurophysiology: Huxley, Foster, Sherrington, Eccles. In *Brain, Mind and Consciousness in the History of Neuroscience* (pp. 255-272). Springer, Dordrecht.

Smith, S. M. (2002). Fast robust automated brain extraction. Human brain mapping, 17(3), 143-155.

Smith, T. R., Beran, M. J., & Young, M. E. (2017). Gambling in rhesus macaques (Macaca mulatta): The effect of cues signaling risky choice outcomes. *Learning & behavior*, *45*(3), 288-299.

Smith, A. T., Williams, A. L., & Singh, K. D. (2004). Negative BOLD in the visual cortex: evidence against blood stealing. *Human brain mapping*, *21*(4), 213-220.

Stauffer, W. R., Lak, A., Bossaerts, P., & Schultz, W. (2015). Economic choices reveal probability distortion in macaque monkeys. *Journal of Neuroscience*, *35*(7), 3146-3154.

Stehling, M. K., Turner, R., & Mansfield, P. (1991). Echo-planar imaging: magnetic resonance imaging in a fraction of a second. *Science*, *254*(5028), 43-50.

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory*. Princeton University Press.

Strait, C. E., Blanchard, T. C., & Hayden, B. Y. (2014). Reward value comparison via mutual inhibition in ventromedial prefrontal cortex. *Neuron*, *82*(6), 1357-1366.

Summerfield, C., & Tsetsos, K. (2012). Building bridges between perceptual and economic decision-making: neural and computational mechanisms. *Frontiers in neuroscience*, *6*, 70.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT press.

Thaler, D., Chen, Y. C., Nixon, P. D., Stern, C. E., & Passingham, R. E. (1995). The functions of the medial premotor cortex. *Experimental Brain Research*, *102*(3), 445-460.

Thrun, S. B. (1992). The role of exploration in learning control. In: D. A. White and D. A. Sofge (eds.) *Handbook of Intelligent Control: Neural, Fuzzy and Adaptive Approaches*.

Thura, D., & Cisek, P. (2014). Deliberation and commitment in the premotor and primary motor cortex during dynamic decision making. *Neuron*, *81*(6), 1401-1416.

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

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

Tomasello, M., & Call, J. (1997). *Primate cognition*. Oxford University Press, USA.

Tremblay, L., & Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. *Nature*, *398*(6729), 704.

Trepel, C., Fox, C. R., & Poldrack, R. A. (2005). Prospect theory on the brain? Toward a cognitive neuroscience of decision under risk. *Cognitive brain research*, *23*(1), 34-50.

Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences*, *109*(24), 9659-9664.

Tsetsos, K., Moran, R., Moreland, J., Chater, N., Usher, M., & Summerfield, C. (2016). Economic irrationality is optimal during noisy decision making. *Proceedings of the National Academy of Sciences*, 201519157.

Tufail, Y., Matyushov, A., Baldwin, N., Tauchmann, M. L., Georges, J., Yoshihiro, A., ... & Tyler, W. J. (2010). Transcranial pulsed ultrasound stimulates intact brain circuits. *Neuron*, *66*(5), 681-694.

Tufail, Y., Yoshihiro, A., Pati, S., Li, M. M., & Tyler, W. J. (2011). Ultrasonic neuromodulation by brain stimulation with transcranial ultrasound. *nature protocols*, *6*(9), 1453.

Turner, B. M., Van Maanen, L., & Forstmann, B. U. (2015). Informing cognitive abstractions through neuroimaging: the neural drift diffusion model. *Psychological review*, *122*(2), 312.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and uncertainty, 5(4), 297-323.

Vanduffel, W., Fize, D., Mandeville, J. B., Nelissen, K., Van Hecke, P., Rosen, B. R., ... & Orban, G. A. (2001). Visual motion processing investigated using contrast agent-enhanced fMRI in awake behaving monkeys. *Neuron*, *32*(4), 565-577.

Van Essen, D. C. (2002). Surface-Based Atlases of Cerebellar Cortex in the Human, Macaque, and Mouse. *Annals of the New York Academy of Sciences*, *978*(1), 468-479.

Velling, V. A., & Shklyaruk, S. P. (1988). Modulation of the functional state of the brain with the aid of focused ultrasonic action. *Neuroscience and behavioral physiology*, *18*(5), 369-375.

Verhagen, L., Gallea, C., Folloni, D., Constans, C., Jensen, D., Ahnine, H., ... & Klein-Flugge, M. (2018). Offline impact of transcranial focused ultrasound on cortical activation in primates. *bioRxiv*, 342337.

Walton, M. E., Bannerman, D. M., Alterescu, K., & Rushworth, M. F. (2003). Functional specialization within medial frontal cortex of the anterior cingulate for evaluating effort-related decisions. *Journal of Neuroscience*, *23*(16), 6475-6479.

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

Wang, X. J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, *36*(5), 955-968.

Watson, J. (1913). Psychology as a Behaviorist Views It. *Psychological Review*, 20: 158–77.

Weber, B. J., & Chapman, G. B. (2005). Playing for peanuts: Why is risk seeking more common for low-stakes gambles?. Organizational Behavior and Human Decision Processes, 97(1), 31-46.

Weber, E.H. (1851). *Der tastsinn und das gemeingefühl.* Translated in English by D.J.Murray *On the sense of touch and common sensibility,* London.

Wernicke, C. (1874). *Der aphasische Symptomencomplex: eine psychologische Studie auf anatomischer Basis.* Breslau: Cohn. Published in English: (1970) The aphasic symptom-complex: a psychological study on an anatomical basis. *Archives of Neurology*, *22*(3), 280-282.

Wilson, B., Kikuchi, Y., Sun, L., Hunter, D., Dick, F., Smith, K., ... & Petkov, C. I. (2015). Auditory sequence processing reveals evolutionarily conserved regions of frontal cortex in macaques and humans. *Nature communications*, *6*, 8901.

Wilson, B., Marslen-Wilson, W. D., & Petkov, C. I. (2017). Conserved sequence processing in primate frontal cortex. *Trends in neurosciences*, *40*(2), 72-82.

Wise, S. P. (2008). Forward frontal fields: phylogeny and fundamental function. *Trends in neurosciences*, *31*(12), 599-608.

Wittmann, M. K., Kolling, N., Akaishi, R., Chau, B. K., Brown, J. W., Nelissen, N., & Rushworth, M. F. (2016). Predictive decision making driven by multiple time-linked reward representations in the anterior cingulate cortex. *Nature communications*, *7*, 12327.

Wittmann, M. K., Lockwood, P. L., & Rushworth, M. F. (2018). Neural mechanisms of social cognition in primates. *Annual review of neuroscience*, *41,* 99-118.

Woolrich, M. (2008). Robust group analysis using outlier inference. *Neuroimage*, *41*(2), 286-301.

Woolrich, M. W., Behrens, T. E., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *Neuroimage*, *21*(4), 1732-1747.

Worsley, K. J. (2001). Statistical analysis of activation images. *Functional MRI: An introduction to methods*, *14*(1), 251-70.

Wundt W.M. (1873-4). *Grundzüge der physiologischen Psychologie*, 2 volumes. Leipzig: Engelmann. Translated in English by E.B. Titchener (1904) *Principles of physiological psychology*, London: Allen.

Wyart, V., & Koechlin, E. (2016). Choice variability and suboptimality in uncertain environments. Current Opinion in Behavioral Sciences, 11, 109-115.

Yamada, H., Tymula, A., Louie, K., & Glimcher, P. W. (2013). Thirst-dependent risk preferences in monkeys identify a primitive form of wealth. Proceedings of the National Academy of Sciences, 201308718.

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience, 19(3),* 356.

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