Neural Mechanisms of Reward-guided Learning and Irrational Decision-making

Georgios K. Papageorgiou

Linacre College

University of Oxford

Thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy at the University of Oxford

Hilary Term 2016

Ἢ τὰν ἢ ἐπὶ τᾶς
[Plutarch, *Moralia*, 241]

Neural Mechanisms of Reward-guided Learning and Irrational Decision-making

Georgios K. Papageorgiou

Abstract

The ability to take effective decisions is fundamental for successful environmental adaptation and survival. In this thesis, I investigated situations in which decisions appear irrational, at least from certain standpoints. I conducted a behavioural decision-making experiment in two groups of macaques: controls and a group with ventromedial prefrontal cortex/medial orbitofrontal cortex (vmPFC/ mOFC) lesions. Some choices lead to compound outcomes composed of different constituent parts. Control macaques' decisions suggested their estimates of the value of the compound were biased away from the sum of the values of the constituents and towards their mean. Lesions of vmPFC/mOFC diminished the size of the effect so that macaques in some ways appeared to make more rational decisions. Based on the results of this experiment I devised a similar Functional Magnetic Resonance Imaging (fMRI) paradigm with the control animals. This demonstrated strong vmPFC/mOFC activity when similar decisions were made and suggested a value comparison process.

In addition, I investigated the role of dopamine in learning using Fast-Scan Cyclic Voltammetry (FSCV), while rats performed a simple decision-making task. Theories about the role of dopamine in learning have focused on the possibility that it codes scalar reward value prediction errors. Less consideration has been given to the possibility that dopamine might reflect prediction errors about reward identities regardless of value. I measured dopamine in the nucleus accumbens when unexpected changes in reward value or identity occurred while rats executed a two-choice two-reward instrumental task. Dopamine levels in the nucleus accumbens reflected reward value prediction errors. In addition, however, they also reflected some information about reward identity under some circumstances. Further investigation suggested that this might be due to differences in the nutritional value of different reward types that did not have clear measurable impacts of behaviour in the tasks that I used.

Thesis submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
Linacre College, University of Oxford
Hilary Term 2016

Acknowledgements

If someone would ask me at the beginning of my undergraduate studies what I would like to do later in life, I would have definitely never answered that I would like to obtain a doctoral degree. I considered it a waste of time and a very dreadful choice. I had never seen myself becoming a Doctor and I was envisioning a chilled out lifestyle next to the sea. As a child the only thing I remember constantly doing was playing in the neighbourhood and setting traps for birds. I would find it so hard to study seriously to the extent that my parents doubted whether I would be able to get a place to study anything in any university. If I had to mention a single dream, this would have nothing to do with science, but that one day I would free my island from the Turkish invaders who unfortunately still occupy a significant part of it.

But life does play strange games sometimes and growing up my mindset started changing. I started seeing what I previously considered as meaningless the best way to give a special meaning to my life. I think one of my main motivations for becoming a neuroscientist was that I always felt fascinated about the human brain and intrigued observing and analysing the human behaviour. Most importantly it was the hope that with my research I would have an impact on improving the quality of human life. Hopefully when all is set and done I will feel peaceful inside of me that I did sincerely try for this purpose.

I have to thank many people that supported me throughout my doctoral studies.

First of all, I would like to thank my labmates for their help, collegiality and support, and especially Jerome Sallet, Marco Wittmann, Bolton Chau, Urs Schufelgen, Franz Neubert, Rhyanne Dale, Anna Huber and Flavia Cucca.

I would like to thank my second supervisor, Mark Walton, for all his help, meticulous feedback and for being there whenever I had a question or needed any help. He taught me many things as well as how to be careful and cautious in my analyses and experiments. I also had the honour to be his first doctoral student publishing results from his newly formed lab.

I would like to give special thanks to my first supervisor, Matthew Rushworth, who gave me the opportunity to pursue a doctoral degree in Oxford. I think I could have written many pages about how great of a human being Matthew is, but I am afraid this thesis would have become twice as big. He showed me so much trust and tremendous support whenever I needed it and through our numerous conversations and interactions he allowed me to develop my own ideas and evolve as a new scientist. It is rare that you meet such individuals that are on a different level both as scientists and as humans. Such people give to this world an example and a hope.

I would also like to thank my friends for their encouragement and joyful conversations. I would like to thank my partner for her care and precious support. I must have been quite boring constantly speaking about neuroscience and work-related topics during our numerous coffees at Costa. I would also like to thank my sisters for their support, love and that extra push when needed. I could not skip thanking my four little friends Puck, Pringle, Pugsley and Puzzle. Without them, this thesis would not have been possible.

Finally, and most of all, I would like to thank my parents. I was blessed to have parents that showed me enormous support and unconditional love during my unusually long studies - although they never understood why I left my beautiful island of Cyprus.

I am devoting this thesis to my parents and to my brother who has not even for a moment left my thoughts.

Table of Contents

1 General Introduction 1
1.1 A general historical review: Phineas Gage 2
1.2 Neurobiology of Decision-Making 3
1.2.1 Frontal Cortex 3
1.2.2 Basal ganglia 6
1.3 The neuroanatomy of the rat, macaque and the human brain 10
1.4 The neuropsychology of decision-making in animals (macaques & rodents) 13
1.4.1 Cognitive flexibility 13
1.4.2 Outcome encoding 14
1.5 The role of dopamine in learning 19
1.6 Frontal striatal dopaminergic systems: learning & information representation……..... 22
1.7 Neurophysiology of reward 24
1.7.1 Reward identity 26
1.7.2 Contiguity 28
1.7.3 Contingency 28
1.7.4 Prediction error 30
1.8 Learning theories and models 33
1.8.1 Decision Theory 33
1.8.2 Dual-process accounts of reasoning theory 34
1.8.3 Incentive Theory 35
1.8.4 Reinforcement Learning 36
1.8.5 Rescorla-Wagner model 37
1.8.6 Temporal Difference Learning model 39
1.9 Techniques 40
1.9.1 Functional Magnetic Resonance Imaging 40
1.9.2 Fast-Scan Cyclic Voltammetry 41
1.9.3 Lesion studies 43
1.10 Thesis experiments 43

2 The role of ventromedial prefrontal cortex in suboptimal decision-making…….....46
2.1 Introduction 47
2.2 Methods 49
2.2.1 Subjects.....49
2.2.2 Surgery and Histology 50
2.2.3 Behavioural training 52
2.2.4 Stimulus-reward learning 55
2.2.5 Main task 56
2.2.6 Devaluation.....58
2.3 Results 59
2.3.1 Preliminary training.....59
2.3.2 Main task 60
2.3.3 Devaluation.....62
2.4 Discussion 65

3 Less is more: Ventromedial prefrontal cortex and evaluation of compound options 72
3.1 Introduction 74
3.2 Methods 76
3.2.1 Subjects.....76
3.2.2 Behavioural training 76
3.2.3 Imaging data acquisition.....79
3.2.4 FMRI data preprocessing.....79
3.2.5 FMRI data analysis.....80
3.2.6 Timecourse analysis 85
3.2.7 Chosen/Unchosen value analysis.....85
3.3 Results 87
3.3.1 Behavioural Results.....87
3.3.2 Neuroimaging Results 92
3.4 Discussion 104

4 The influence of reward identity on mesolimbic dopamine release 112
4.1 Introduction 113

4.2 Methods 115
4.2.1 Surgical Procedures 115
4.2.2 Behavioural Paradigm 116
4.2.3 Fast-scan cyclic voltammetry 121
4.2.4 Data analysis 121
4.2.5 Histology 124
4.3 Results 126
4.3.1 Behavioural performance during Baseline food-restricted sessions 126
4.3.2 Dopamine release in response to outcome surprise in a food-restricted state 128
4.3.3 Dopamine release in response to predictive cues in a food-restricted state 133
4.4 Discussion 137

5 Mesolimbic dopamine reflects state-dependent prediction errors 140
5.1 Introduction 142
5.2 Methods 144
5.2.1 Data analysis 145
5.3 Results 146
5.3.1 Behavioural results 146
5.3.2 Dopamine release at reward delivery following sensory-specific satiation procedures 148
5.3.3 Rapid updating of cue-elicited dopamine release after changes in state 151
5.4 Discussion 156

6 General Discussion 161

Bibliography 168

List of Figures

Figure 1.1: 3-D computer generated view of the scull and lesions on Phineas Gage brain 2
Figure 1.2: Human and macaque vmPFC/mOFC regions 6
Figure 1.3: Example of NAc neurons' dendrites 7
Figure 1.4: Prefrontal cortex homologue regions in the human, macaque monkey and rat brain 12
Figure 1.5: The main dopaminergic pathways projections 23
Figure 1.6: The reward circuit 25
Figure 1.7: Behavioural functions of rewards: contiguity, contingency and prediction error 28
Figure 1.8: Neuronal behaviour under different contingencies 30
Figure 2.1: Histology 51
Figure 2.2: Stimuli information 53
Figure 2.3: Experimental setting 54
Figure 2.4: Choices between pairs across all sessions 62
Figure 2.5: Choices between pairs during the vegetable devaluation session 64
Figure 2.6: Choices between pairs during the fruit devaluation session 65
Figure 3.1: Experimental setting 77
Figure 3.2: Macaque behaviour before and during scanning 88
Figure 3.3: Median Reaction Times 89
Figure 3.4: Choices and reaction times for CS+s versus blank side or CS- stimuli 92
Figure 3.5: VmPFC/mOFC activity at cue onset time 93
Figure 3.6: Critical comparison 95
Figure 3.7: Decision difficulty 96
Figure 3.9: VmPFC/mOFC value transitivity 98
Figure 3.10: Chosen minus Unchosen stimulus value 100
Figure 3.11: VmPFC/mOFC stimulus value information 101
Figure 3.12: VmPFC/mOFC activity at reward delivery 102
Figure 3.13: NAc activity at cue onset 104
Figure 4.1: Task design 118
Figure 4.2: Representation of recording sites 125

Figure 4.3: Proportion of food choices 127
Figure 4.4: Individual example MORE (left) and SWITCH (right) trials 128
Figure 4.5: Reward delivery during Baseline sessions 129
Figure 4.6: Dopamine signals in Baseline A 130
Figure 4.7: Comparison of naive versus experienced rats 131
Figure 4.8: Reward delivery dopamine data 133
Figure 4.9: Forced versus Choice trials dopamine release 134
Figure 4.10: Cue-elicited dopamine data 135
Figure 4.11: Forced versus Choice trials 136
Figure 5.1: Preference change 147
Figure 5.2: Effect size during devaluation 148
Figure 5.3: Dopamine release in MORE trials 149
Figure 5.4: Dopamine release on Expected and SWITCH trials 150
Figure 5.5: Dopamine release during Devaluation and Baseline sessions 151
Figure 5.6: Dopamine release during Devaluation sessions 152
Figure 5.7: Dynamic changes in cue-evoked dopamine after selective satiation 154
Figure 5.8: Cue-elicited dopamine release on 1st trials of the session 155

List of Tables

Table 2.1: Experimental timeline.....56
Table 3.1: General Linear Model regressors for GLM-1 and GLM-2.....82
Table 3.2: General Linear Model contrasts for GLM-1 and GLM-2.....84
Table 3.3: Decision-related activity.....94
Table 3.4: Reward-related activity.....103

1 General Introduction

Human behaviour is fundamentally based on decision-making and every day decisions shape our lives. These decisions vary in their complexity, frequency and importance, but most critically they define who we are and to a certain extent our personal wellbeing. In order to get a better understanding on how these decisions are made, we need to dig down to the neuronal level. Francis Crick (1916-2004) in his "astonishing hypothesis" proposed that:

> *"A person's mental activities are entirely due to the behaviour of nerve cells, glial cells, and the atoms, ions, and molecules that make them up and influence them".*

As all mental activities incorporate a decision-making component, it becomes obvious that our decisions represent the end result of what is happening both at neural and neurochemical levels. A vital aspect of rational decision-making however is the ability to accurately evaluate different options. Accurate anticipation of the value of future events that may result from the choices we take is critical for guiding decision-making. It is only in this way that we can take choices that avoid harm and maximise gain. It is therefore important to gain a better understanding of the brain regions and brain chemicals that are involved in those decisions. This is the focus of this doctoral thesis and through experiments in non-human primates and rats I have tried to shed a ray of light onto the decision-making process.

1.1 A general historical review: Phineas Gage

A famous example that demonstrates the role of frontal regions in decision-making is that of an American railroad construction foreman, Phineas Gage. Phineas' job concerned drilling boulders and filling them with dynamite. One day, while he was preparing an explosion by compacting a bore with dynamite using a tamping iron, a spark from the tamping iron ignited the explosive powder. This resulted in the iron being propelled at high speed through his skull, following a trajectory through his left cheekbone and exiting through the top of his head (Ratiu et al., 2004) (Fig. 1.1). As a consequence, Phineas suffered extensive damage to the orbital and midline prefrontal regions and exhibited inflexible, erratic, stimulus-bound behaviour (Damasio et al., 1994; Schoenbaum et al., 2009). A recent study by Van Horn et al., (2012) focusing on the effects of injury to white matter connections, showed that while the damage was mostly in the left frontal cortex, connections were lost between this brain region and left temporal cortex, right frontal cortex, as well as left limbic structures.

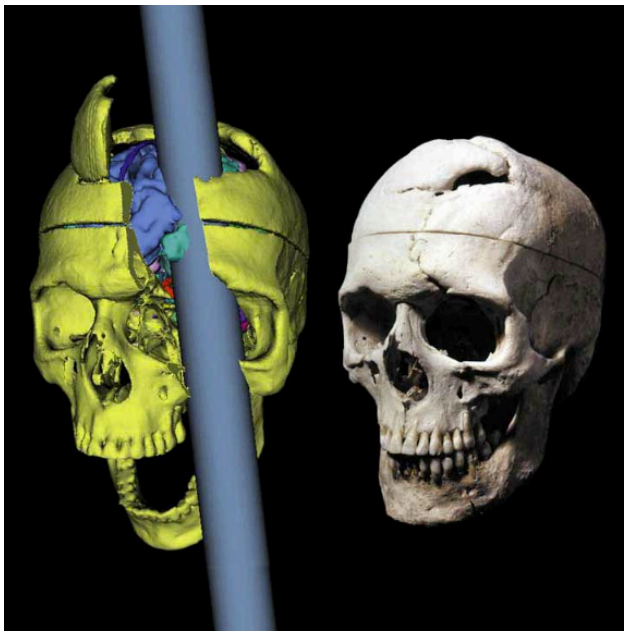


Figure 1.1: 3-D computer generated view of the scull and lesions on Phineas Gage brain (Adapted from Ratiu et al., 2004)

Remarkably, he was conscious and able to walk just minutes after the accident and within a few months he was able to lead an apparently normal life. But it did not take long until he started exhibiting significant changes in his personality. He lost his prior social inhibitions and was fired from his job. He switched to other jobs and finally went on to live with his mother. In general, his behaviour and his decision-making ability changed dramatically. Around thirteen years after the accident, he died as a result of complications arising from epileptic convulsions (Ratiu et al., 2004). Discussions of Phineas Gage often emphasise the damage that he sustained to the orbitofrontal cortex (OFC) but as the reconstructions of the tamping iron's trajectory demonstrate it is clear that the damage was more widespread and many brain regions were disconnected from one another. In the next section, however, I review the evidence implicating more specific brain regions in reward-guided decision-making.

1.2 Neurobiology of Decision-Making

Multiple brain regions are implicated in decision-making. Of particular interest for my work are the frontal cortex (FC) and the ventral striatum, including the nucleus accumbens (NAc).

1.2.1 Frontal Cortex

FC plays a major role in reward-guided learning and decision-making (Rushworth et al., 2011; Takahashi et al., 2011). Four distinct anatomical regions make important contributions: the ventromedial prefrontal cortex (vmPFC) and adjacent medial orbitofrontal cortex (mOFC), the lateral orbitofrontal cortex (lOFC), the anterior cingulate cortex (ACC) and the anterior prefrontal cortex (aPFC)

(Rushworth et al., 2011). The most studied of these regions is the vmPFC/mOFC, a region in which activity co-varies with the value of potential choices (Plassmann et al., 2007; Rushworth et al., 2011).

OFC has been thought to play a significant role in flexibility in decision-making (Schoenbaum et al., 2009; Walton et al., 2010), inhibiting previously chosen actions (Clarke et al., 2008), processing negative outcomes (Fellows, 2007) and signalling outcome expectancies (Schoenbaum et al., 2009). Walton et al., (2010) have argued that some of these disparate processes are linked to the process of assigning credit for decision outcomes to a particular preceding event.

Lesion experiments in monkeys have shown that distinct FC regions have a unique role to play in terms of choosing between actions and stimuli. Rudebeck et al. (2008) examined the effect of OFC and ACC lesions in macaques in four experiments and showed that the lesions were associated with different patterns of impairment (see also Kennerley et al., 2006). Two of the experiments investigated stimulus-reward association learning and two investigated action-reward association learning. ACC was particularly important whenever decisions were based on the associations of actions with reinforcement whereas OFC was essential when decision-making depended on the associations of stimuli with reinforcement.

1.2.1.1 Ventromedial prefrontal cortex

The vmPFC/mOFC is part of the prefrontal cortex. Many studies have indicated that the vmPFC/mOFC plays a critical role in the evaluation of choices (Boorman et al., 2009; Fellows & Farah, 2007; Padoa-Schioppa & Assad, 2006; Plassmann et al., 2007; Rushworth et al., 2011; Strait et al., 2014). The vmPFC/mOFC is also activated during the presentation of sensory cues such as

pleasant odours that signal specific reward (Gottfried et al., 2003), during simple evaluations when subjects have to rate single options (Barron et al., 2015; Hare et al., 2009; Lebreton et al., 2009). Its activity, in addition to reflecting value information, seems to reflect confidence estimates, regardless of whether the participants had to explicitly report how confident they were about their decisions or not (Lebreton et al., 2015).

Estimating the correct value of future outcomes and assigning the correct value to different options is essential for optimal decision-making. However, in the case of patients with vmPFC/mOFC lesions this becomes problematic. Patients with vmPFC/mOFC lesions are more inconsistent in their preferences (Fellows & Farah, 2007), had difficulties in differentiating between optimal and suboptimal decisions (Young et al., 2010) and expressed reduced reward priming behaviour (Vaidya & Fellows, 2015). A similar effect was found in macaque monkeys with vmPFC/mOFC lesions. For example, it has been repeatedly shown that such lesions impair visual discrimination reversal learning tasks in which the stimulus-reward associations are changed periodically (Izquierdo et al., 2005). Animals with OFC lesions cannot stop responding to a previously rewarded object when it is no longer rewarded. More recently Rudebeck et al (2013) have suggested that such OFC deficits may actually reflect damage to white matter subjacent to the OFC; they found that fibre-sparing excitotoxic lesions of OFC do not impair discrimination reversal learning. It is, however, still recognised that even excitotoxic OFC lesions disrupt the ability to respond in an adaptive way to changes in outcome value in reward devaluation task (Rudebeck et al., 2013).

What is more intriguing about this area though is the fact that although it is probably the region most frequently studied in investigations of decision-making, its

cytoarchitectural boundaries are still not well defined. Anatomical research from Mackey and Petrides (2010, 2014) has attempted to clarify the architectonic features that characterise areas in this region for both the macaque and the human brain. They argue that area 25, ventral portions of areas 14m, 24 and 32, and medial portions of 10 are homologous in the two primate species and constitute the vmPFC/mOFC region (Fig. 1.2).

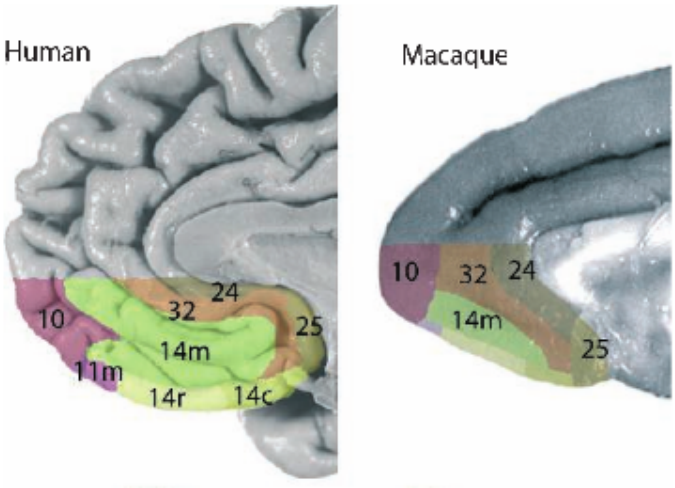


Figure 1.2: Human and macaque vmPFC/mOFC regions
(Adapted from Mackey & Petrides, 2010)

1.2.2 Basal ganglia

The basal ganglia have been implicated in motivation and action selection. The basal ganglia consist of four nuclei: the striatum, the substantia nigra pars reticulata, the globus pallidus and the subthalamic nucleus. Of particular interest for this thesis is the striatum. The striatum is divided into three regions in primates: the putamen, the caudate nucleus and the ventral striatum or nucleus accumbens (NAc) – which correspond to dorsolateral, dorsomedial and ventral striatum in the rodent. The

ventral striatum, and particularly the NAc core region is the focus of this thesis and will be discussed in more detail below.

1.2.2.1 Nucleus Accumbens

The NAc is classically divided in two subdivisions: the core and the shell (Baliki et al., 2013; Pothuizen et al., 2005). The NAc core, which is the focus of some of my own experiments described in later chapters, is made up mainly of medium spiny neurons (MSNs), a type of GABA-ergic inhibitory cells (Shirayama & Chaki, 2006). MSNs are often classified into being either D1-type or D2-type, depending on which type of dopamine (DA) receptors they express. Neurons in the core, in contrast to neurons in the shell, have a higher density of dendritic spines as well as more dendritic branch segments and terminal segments (Meredith et al., 1992). See also Fig. 1.3.

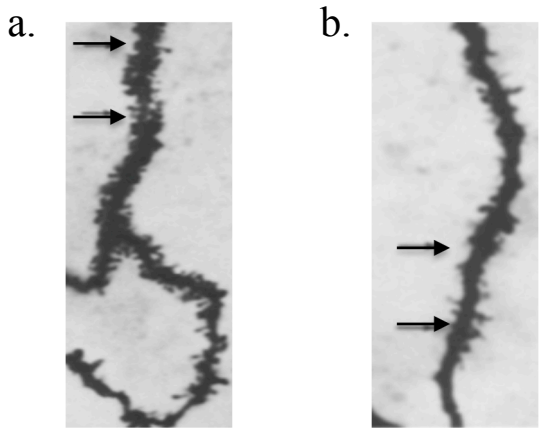


Figure 1.3: Example of NAc neurons' dendrites
(a) NAc core neuron dendrite. (b) NAc shell neuron dendrite. The area between the two arrows (10μm) in each figure demonstrates the difference in dendritic spines between the two neuron types (Adapted from Meredith et al., 1992).

NAc is a key structure for reward, motivation, action and learning (Corbit & Balleine, 2011; Everitt, 1991; Mogenson et al., 1980). Dysfunction of NAc circuits has also

been linked with depression (Shirayama & Chaki, 2006) and drug addiction (Di Chiara, 2002). However, more fine-grained distinctions between the core or shell region have also been described.

For instance, a recent study by Saddoris et al., (2015) investigated patterns of dopamine release in the two compartments of the NAc during a goal-directed appetitive behavioural task. For this purpose, rats had to learn an instrumental chain schedule task in which pressing one lever led to the extension of a second lever and pressing that second lever led to a reward. The dopamine release from both the shell and the core was recorded using fast-scan cyclic voltammetry. After the last recording session an extinction session was run, that was identical to previous schedules, but with no reward delivery following the second lever press. It was found that core dopamine was higher during the presentation of the first lever presentation but not during the second lever or at the delivery of reward. Shell dopamine on the other hand, was high in all task events. Core dopamine was constant throughout the sessions where shell dopamine showed a decrease through the course of the session. During the extinction session, core dopamine decreased during the initial lever presentation whereas the shell dopamine decreased mainly during the second lever presentation. These results indicate that core dopamine release encodes reward prediction errors – in other words, the discrepancy between expected and actual events. In this paradigm the presentation of the first lever, in contrast to the later fully predictable events, enhances motivation and at the same time it constitutes a prediction error as it is associated with the highest dopamine release. Shell dopamine however seems to play an important role in motivation as its dopamine release gradually decreases through the course of the sessions and during extinction. It seems therefore that there are essential differences in the dopamine signal between core and

shell subregions and even within a given behavioural task there is no single, universal, and homogeneous dopamine signal in all parts of the striatum. The simultaneous presence of such signals may provide motivational and predictive information for learning purposes and action execution. The NAc core and shell are also differentially involved in various aspects of decision-making. Inactivation of the NAc core but again not of the shell, led rats to prefer low effort small rewards instead of larger ones that were however associated with higher workload demands (Ghods-Sharifi & Floresco, 2010).

In a study by Day et al., (2011) the role of the NAc neurons in cost-benefit decision-making was assessed. Rats had to learn distinct visual cues predicting either high or low value rewards, during two different tasks: an effort-based and a delay-based task. A reward was considered as 'high value' if it was associated with immediate delivery or low effort for obtaining it whereas a 'low value' reward was associated with high effort or a delay to its delivery. At a behavioural level animals showed a preference for the high value rewards irrespective of the cost for obtaining these. Multi-neuron electrophysiological NAc recordings showed that different sub-populations of neurons exhibited strong phasic firing increase to reflect either the cost-discounted value of future responses, cue presentations, the relative cost of gaining the reward, high-effort requirements, delayed rewards, response initiation or reward delivery.

An important study that demonstrates the importance of this region in stimulus-reward learning comes from a NAc lesion study from Everitt et al., (1991). In this study, rats had to execute a conditioned place preference test before and after NAc lesions. In particular, they had to learn to associate specific environments (places) comprised of distinctive stimuli with the delivery of rewards. Once the

associations were established the rats were tested in extinction i.e. in the absence of any reward. Before the NAc lesions, rats exhibited a strong preference for places that previously were associated with rewards and therefore during extinction they were much more keen on spending more time in those places. After the NAc lesions however, such learning was abolished. They showed decreased exploratory behaviour and overall problems in motor responses towards previously valuable environmental stimuli. NAc core lesions were also found to cause persistent impulsive choice in rats (Pothuizen et al., 2005; Cardinal et al., 2001) and Pavlovian contextual conditioning (Levita et al., 2002).

In primates, the ventral striatum / NAc has also been implicated in reward and motivation (Haber & Knutson, 2010). For instance, fMRI studies of reward-guided decision-making frequently show that activation in the ventral striatum correlates with the value of an option (e.g., Croxson et al., 2009; Knutson et al., 2001) and tracks individual's subjective value of monetary rewards (Kable & Glimcher, 2007). More recently, single unit recording studies in primates have also found that cells here encode the value of presented offers, the difference in value between two options, and the value of whichever is chosen (Strait & Hayden, 2015). It is currently unclear whether the same clear anatomical and functional distinctions between the core and shell observed in rodents are present in primates (though see Baliki et al., 2013).

1.3 The neuroanatomy of the rat, macaque and the human brain

The size and location of main FC structures vary between species. Clear homologies can be established between many areas in human and macaque FC but, while similarities can be identified, there are important differences between primates

and rodents. The rat has a much more developed olfactory system but its visual cortex is less developed in comparison with primates. The neocortex, an area that is crucial for many higher cognitive and behavioural functions, has been extensively developed in primates under evolutionary pressure. A number of FC structures such as ACC, prelimbic, infralimbic, agranular insular, and agranular orbital cortex are present in all three species. However, the largest part of the primate prefrontal cortex has no clear homologue in other mammals. This part includes the granular prefrontal areas, areas with dense granule cell layers, and consists of the dorsolateral prefrontal cortex (area 46), ventral prefrontal cortex (areas 12, 44, 45, 47) frontal pole cortex (area 10), rostral orbitofrontal cortex (parts of area 11, 13, 14), medial and dorsal prefrontal areas (parts of area 9) (Wise, 2008). Some of the most posterior regions of the orbital surface of the FC in the primate, on the border between OFC and anterior insula and in the anterior insula may correspond to areas typically labelled OFC in the rat and parts of posterior agranular medial frontal cortex in the primate may also correspond to prelimbic and infralimbic cortex in the rat (Fig. 1.4).

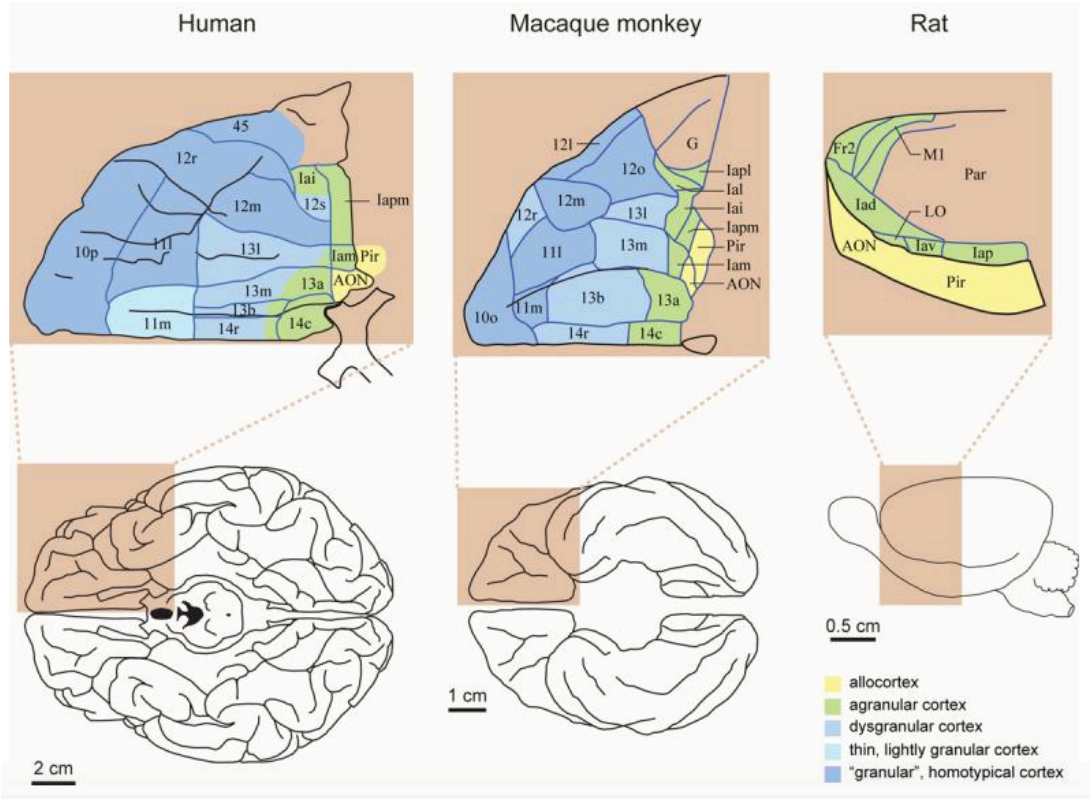


Figure 1.4: Prefrontal cortex homologue regions in the human, macaque monkey and rat brain
(Adapted from Wise, 2008)

As discussed above, the ventral striatum / nucleus accumbens is structurally similar in rodents and primates (Haber & Knutson, 2010). The dopaminergic midbrain, located rostral to the brain stem and the pons and including the substantia nigra and ventral tegmental area (VTA) provides important neurochemical input to the striatum and FC. In the rat brain, substantia nigra pars compacta (SNc) and ventral tegmental area (VTA) are sometimes designated areas A9 and A10 respectively (Dahlstrom & Fuxe, 1965). Dopamine projections are considered to be largely conserved in the mammalian brain (Garcia-Cabezas, 2009).

1.4 The neuropsychology of decision-making in animals (macaques & rodents)

Two of the skills most affected by OFC dysfunctions include cognitive flexibility and outcome encoding.

1.4.1 Cognitive flexibility

OFC has a major role in cognitive flexibility in response to unexpected outcomes (Butter, 1969; Paton et al., 2006; Roberts, 2006; Rudebeck & Murray, 2008). Damage to OFC is thus related to deficits in optimal decision-making. For example, animals may be taught to respond to one cue in order to receive a reward while withholding a similar response to a non-rewarding stimulus. Although animals can effectively learn such a pattern of behaviour, when the meaning of the cues is reversed (by reassigning reward to the previously unrewarded stimulus and no reward to the previously rewarded stimulus), animals with OFC lesions have great difficulty in executing the task correctly, failing to explore the alternative option. Such research has been taken to demonstrate the significance of the OFC in flexible decision-making (Schoenbaum et al., 2009; Schoenbaum et al., 2007). However, a recent paper from Rudebeck et al., (2013) has challenged this conclusion by showing that when the monkeys have excitotoxic OFC lesions, which selectively destroy cell bodies but spare adjacent white matter pathways, there is no impairment in their ability to perform an object-reward association reversal learning task.

A study by Walton et al. (2010) re-examined the role of OFC in decision-making. More specifically, macaque monkeys performed dynamically changing three armed-bandit tasks. During one set of conditions, the reward associations of the three options changed gradually and independently of one another, whereas in the other conditions, reward assignments remained stable. In agreement with previous research,

it was found that OFC lesions impair flexible decision-making, but further analyses revealed that this behaviour was not due to insensitivity to negative outcomes or failure to inhibit previously rewarding responses; both lesioned animals and controls were able to respond to local changes in reward likelihood when the best stimulus remained the same. It was also demonstrated that the overall pattern of impairments could have been explained if the OFC was concerned with specific contingent learning – the ability to link a specific outcome to a specific stimulus choice. This learning mechanism allows rewards received for a certain choice among many alternatives to be assigned to that option alone. Monkeys with OFC lesions were unable to associate an outcome with the specific choice that had caused it and showed a tendency to relate outcomes with all choices that had occurred close in time.

A vital role for the OFC is therefore in learning and updating contingent relationships between specific choices and their outcomes. A failure in this may account for the pattern of impaired decision-making seen after lesions to parts of primate OFC. For example, failure in object discrimination reversal learning seen after aspiration lesions of OFC (Izquierdo et al., 2004) may reflect a failure of the credit assignment process identified by Walton and colleagues that prevents the animals from learning that the reward should be credited to a different stimulus after the reversal.

1.4.2 Outcome encoding

Multiple brain regions are implicated in outcome encoding but one basic distinction that can be drawn among decision-making systems is between striatal and frontal based systems. Two discrete corticostriatal networks drive distinct decision processes. Certain frontal cortical (FC) regions as well as their projections in

dorsomedial striatum are considered responsible for 'goal-directed' decision-making (Balleine et al., 2008). 'Goal-directed' decision-making is based on predicting and comparing potential future specific outcomes using a model of the environment (Botvinick & An, 2008; Daw et al., 2005). In the second instance, dorsolateral striatum and sensorimotor cortices are thought to be responsible for 'habitual' or automatic actions (Balleine et al., 2008). Such habitual decision-making is based on stored or "cached" action values (Botvinick & An, 2008).

A prerequisite for acquisition and expression of stimulus-outcome (also called "Pavlovian") and instrumental action-outcome associations is outcome encoding. Ostlund and Balleine (2007) assessed OFC's role in outcome encoding. Rats with bilateral excitotoxic OFC lesions made either before or after training were tested in Pavlovian and instrumental conditioning tasks. In the first experiment, rats with OFC or sham lesions were trained in both a Pavlovian and an instrumental task. During Pavlovian conditioning sessions two auditory cues that served as conditioned stimuli (CSs) were paired with distinct rewards that served as unconditioned stimuli (USs); in the instrumental conditioning sessions each of two levers (right and left) were also paired with one of these same two reward outcomes (USs). OFC lesions made before training had no influence in learning rates on either the Pavlovian or instrumental task.

After the associations had been acquired, rats were given two outcome devaluation-testing sessions where, prior to the session, the animals were given 1 hour's free access to one of the two reinforcers to induce selective satiety for this option. In these sessions, both levers were inserted into the operant chamber but no reward was delivered (extinction test). This is a sensitive test of the nature of associations between actions and outcomes as the animals have to use the change in

value of one reinforcer to selectively adjust their responding in the absence of direct experience of that reinforcer in the testing context. It was found that instrumental outcome devaluation was unaffected OFC-lesioned rats. Two days later, rats returned to the operant chamber to investigate whether the CSs facilitated instrumental performance on the basis of a shared outcome-representation, a phenomenon better known as Pavlovian-instrumental transfer. Each lever was inserted again into the operant chamber and again no reward was delivered. At points during the session, each CS was presented. While sham lesioned animals or animals that had received OFC lesions pre-training selectively increased responding during CS presentation on the lever associated with the same reward type as the CS, post-lesion OFC lesions abolished this transfer effect.

The results therefore suggest a role for OFC in encoding specific stimulus-reward associations. This conclusion was supported by a second experiment that manipulated the strength of the stimulus-reward association by using a procedure called contingency degradation; this reduced the predictive information of each cue by presenting the USs with equal probability at times when the CS was not present. OFC lesioned rats, in contrast to controls, failed to direct their responding to stimuli as a function of whether the stimuli were predictors of reward or not.

It has also been suggested that OFC signals outcome expectancies related to the sensory and contextual properties as well as the unique value of the expected reward (Schoenbaum et al., 2009; Takahashi et al., 2011). Evidence for this comes from neuronal recording studies that demonstrate anticipatory firing in OFC neurons in both the rat (Schoenbaum et al., 1998) and macaque (Tremblay & Schultz, 1999) when animals are expecting particular types of reward. Similarly, reinforcer devaluation studies, in which animals devalue a previously rewarded cue, as a result

of the cue being paired with illness, satiety, or simply absence of any reward. Both rats and monkeys with OFC lesions fail to devalue such cues in the normal way (Balleine et al., 2008; Rudebeck & Murray, 2011).

Besides the OFC, another brain region considered to be critical in outcome encoding and reward-based choice is the ventral striatum. A recent study by Strait et al., (2015) attempted a direct comparison of the neural activation in the ventral striatum and the vmPFC/mOFC while macaque monkeys were choosing between gambles. In their experiments monkeys had to perform a two-option gambling task during which they had to choose between high-stakes and low-stakes options. Similar neural activity patterns were observed in both brain regions in terms of encoding expected value, value difference, the chosen relative to the unchosen option value and the outcome value. After accounting for other factors that influence value they both showed a significant correlation between residual firing rate variance and choices. It therefore seems that these two brain regions have an overlapping role in various aspects of value encoding. This finding is very intriguing considering recent interest in the primate vmPFC (particularly human vmPFC in neuroimaging studies) and the NAc on both rodents and primates. The recent results of Strait and colleagues suggest ways in which vmPFC and NAc may have related functions in decision-making in primates while, in the absence of vmPFC, NAc may be critical for reward-guided decision-making in rodents.

Corbit and Balleine (2011) assessed the role of NAc in specific outcome encoding, learning and motivation. For this purpose, three groups of rats – sham lesioned rats or rats with excitotoxic core or shell lesions – were initially trained on a Pavlovian schedule to associate specific auditory cues with different types of reward (a food pellet, sucrose solution or polycose solution). After learning was established,

they continued to an instrumental task in which they had to learn to press two levers each associated with one of the reward types. The critical test in the experiment concerned a Pavlovian-instrumental transfer (PIT) test that rats had to perform in extinction. During that test, only one of the two levers was available and periodically one of the previously learnt auditory cues was presented. Core-lesioned and sham rats showed significantly increased responding to the lever that was paired with the auditory cue that predicted the same outcome ("outcome specific motivational" transfer). In contrast, shell lesions eliminated outcome-specific PIT. Shell-lesioned and sham rats showed significant increase on lever responses during the presentation of the third auditory cue although this was not paired with either of the two levers ("general motivational" transfer). In contrast, core lesions eliminated general PIT. A second experiment, with reversible inactivation of the same brain regions but only right before PIT testing, confirmed the same pattern of results. These results demonstrate that the NAc and instrumental performance are directly linked but core and shell have separate roles. The NAc core seems to mediate general excitatory effects of reward-related cues whereas the shell mediates outcome-specific reward-related information.

In a crossed-disconnection study by Izquierdo and Murray (2010), the role of the NAc and the mediodorsal thalamic nucleus (MDm) in reward processing was assessed. This study was based on the assumption that if the NAc or the MDm and their respective interactions with the OFC and the amygdala are critical for value-based decisions then a lesion could cause a cognitive deficit. For this purpose, macaque monkeys, with amygdala and OFC lesions in one of the two hemispheres had an excitotoxic lesion of either the NAc or the MDm in their contralateral hemisphere. After the lesions were made, they were trained on discriminating

between different objects that were baited with two different foods. Four critical sessions then commenced; two of these were preceded by selective devaluation of one of the two foods. Monkeys with an MDm disconnection were impaired during the devaluation sessions, failing to adjust their choices in favour of the more valuable option. Monkeys with a NAc disconnection however did not show any impairment. These results show that the MDm but not the NAc is an important component of a broader circuit responsible for goal-directed actions.

A similar effect was observed by Lex and Hauber (2010) who also attempted to investigate the role of the NAc in outcome encoding but from a dopamine perspective. Rats with NAc core dopamine depletion ran both a Pavlovian and an instrumental task after selective outcome devaluation. They were found to be impaired only in the Pavlovian but not in the instrumental task. These findings suggest that dopamine may not be essential for encoding action-outcome associations, but instead to allow environmental cues to amplify motivational drive.

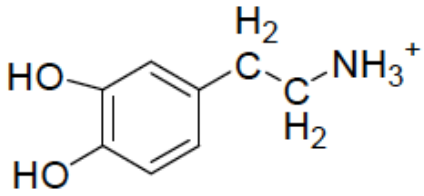
1.5 The role of dopamine in learning

Both humans and animals have the ability to associate neutral stimuli with rewards (Flagel et al., 2011; Pavlov, 1928). It is important to be able to represent the associations between neutral stimuli and predicted rewards in order to determine whether or not the actual rewards when delivered meet these expectations. As described above, the difference between predicted and received rewards produces a reward prediction error signal.

The activity of many dopamine neurons, and dopamine release in the NAc core correlate quantitatively with reward prediction error signals (Bayer & Glimcher,

2005; Cohen et al., 2012; Day et al., 2007; Hart et al., 2014; Montague et al., 1996; Schultz et al., 1997). As described above, dopamine is thought to play an important role in reward-related learning and in mediating the motivating effects of reward on behaviour. Its role in aversive learning is more controversial (Fiorillo et al., 2013; Matsumoto & Hikosaka, 2009; Ungless et al., 2004). Therefore, for simplicity and as all the experiments described in this thesis used appetitive rewards, discussion will be limited to the appetitive domain.

Dopamine is a monoamine neurotransmitter. Its chemical structure is:



Dopamine neurons in subcortical brain regions fire in two modes: phasic and tonic. Phasic or transient dopamine release is caused by dopamine neuron firing in response to relevant stimuli. Tonic or sustained dopamine release is regulated by prefrontal cortical afferents and it regulates the intensity of phasic dopamine release through its effect on extracellular dopamine levels (Grace, 1991). Tonic dopamine levels are set both by slow, spontaneously occurring baseline spike activity of VTA dopamine neurons (Bilder et al., 2004) but also potentially by the number of spontaneous phasic transients (Hamid et al., 2016; Owesson-White et al., 2012).

While the activity of midbrain dopamine neurons has been extensively studied, less is known about the patterns of phasic dopamine release in terminal regions. There are phasic increases in dopamine in the nucleus accumbens core (though not always in other structures) following the unexpected delivery of reward

and in response to a learned cue predicting reward (Day et al. 2007; Brown et al. 2011). However, this seems to depend on the cues gaining incentive properties and therefore promoting behavioural responses ("sign tracking") (Flagel et al., 2011). Generally, dopamine neurons are activated by events that provide reward at levels higher than predicted and exhibit the opposite pattern for non-rewarded events, particularly when such an outcome is worse than predicted (Schultz et al., 1997). Interestingly many models of dopamine function fail to link dopamine to any other changes in the environment, such as changes in reward identity (Balleine et al., 2008; Niv & Schoenbaum, 2008).

An optogenetics study in rats by Steinberg et al., (2013) provided strong evidence for a causal role for dopamine reward prediction errors in learning. In this study, two groups of rats were trained in a "blocking" procedure during which they learned to associate an auditory cue with the delivery of a sucrose solution. The first group ('Blocking group') was subsequently presented with the same auditory cue but paired with a visual cue that led to the delivery of the sucrose solution. The second group ('Control group') followed the same procedure but with the important difference that the auditory cues during the initial and subsequent presentation were not the same. The idea behind the blocking procedure is that if a cue can fully predict the delivery of a specific reward then any other cue present at the time of that cue is 'blocked' from acquiring any reward-related associations. One day later the visual stimulus was presented to the two groups and conditioned responding was assessed. The behavioural results confirmed the blocking effect in the first group of animals but not in the second group as these animals were responding significantly less at the presence of the visual stimulus.

During a subsequent testing session, optogenetically activating the VTA during the reward delivery period showed increased learning for the visual stimulus (or blocked cue) in a group of rats with an excitatory opsin in VTA dopamine neurons but not in two control groups either lacking the opsin or where optical stimulation was performed during the inter-trial interval period. After the completion of this testing procedure, the three groups of rats were tested in two different type of extinction. The first type concerned receiving water instead of sucrose during the last 1-day testing and the second type concerned the absence of any reward delivery. In both cases, rats with Cre recombinase expression spent significantly more time in the reward port during the cue presentation. These results demonstrate that artificially activating VTA at the time of reward, regardless of its predictability, supports learning.

1.6 Frontal striatal dopaminergic systems: learning & information representation

The dopamine system projects strongly to two broad regions: the frontal cortex and the striatum. Moreover, dopamine neuronal activity can be regulated by projections from the striatum, frontal cortex and amygdala.

There are four main dopaminergic tracts in the brain: the mesolimbic and mesocortical pathways, the tuberoinfudibular pathway and the nigrostriatal (or mesostriatal) pathway (Fig. 1.5). The mesolimbic and mesocortical pathways transmit dopamine from the VTA to NAc, amygdala, hippocampus, olfactory bulb, and the prefrontal and cingulate cortex (shown with solid arrows). The tuberoinfudibular pathway originates from the arcuate nucleus of the hypothalamus and transmits dopamine to the pituitary gland (shown with downward dashed arrows). The

nigrostriatal pathway originates from the SN (especially the pars compacta of the SN: SNc) and projects to the dorsal striatum (shown with upward dashed arrows). Deterioration and morphological changes of this pathway can result to Parkinson's disease (Nyberg et al., 2015).

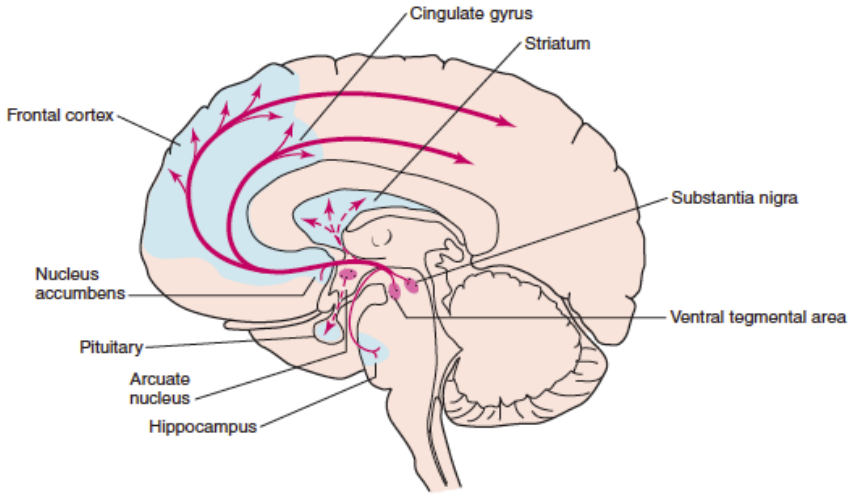


Figure 1.5: The main dopaminergic pathways projections
(Adapted from Nestler et al., 2009)

In a related study by Takahashi et al. (2011) the role of OFC in expectancy-related changes in firing of dopamine neurons, was investigated. Dopamine activity was recorded in rats with OFC lesions during a simple 2-option decision-making task, which revealed significant changes in the dopamine signal when rewards were different from those expected (either different size or occurring at a different time). More specifically, in the lesioned animals, the dopamine cells showed a retarded decrease in prediction-error evoked dopamine activity with learning in spite of there being no direct change in behaviour. These same neurons also failed to suppress firing when an expected reward was omitted, showed less obvious firing differences between non-equal valued rewards, and did not exhibit different firing patterns based on expected rewards during free-choice trials.

1.7 Neurophysiology of reward

When talking about reward it is important to understand what actually constitutes a reward. First of all, the term 'reward' is inextricably linked with pleasantness, as it is something you are happy or keen to work for. This can be a pleasant event such as getting an increase in your salary, a high-impact factor publication etc. Similar processes operate in simpler, daily tasks. For example, as a child you might not care about salaries or publications but perhaps what would be of more interest to you is the approval you might receive from your parents when you finish your homework. Primary reinforcers, such as food are often important rewards for non-human animals. Animals do not have the complicated lifestyle that their civilised relatives have but one fundamental concern they share is that same need for food or reward. They have to forage to find their food and foraging can be quite problematic and risky especially when the available resources are scarce and various predators lurk. This essential need for primary reward is what is often exploited in scientific laboratories.

Reward in most situations elicits two forms of behaviour: approach and consumption. The repetition of reward-seeking behaviour is directly related to reinforcement; if an action or behaviour is reinforced, the higher the probability becomes that it will be repeated in the future (White, 2011). The need for reward therefore supports incentive-based learning (Haber & Knutson, 2010).

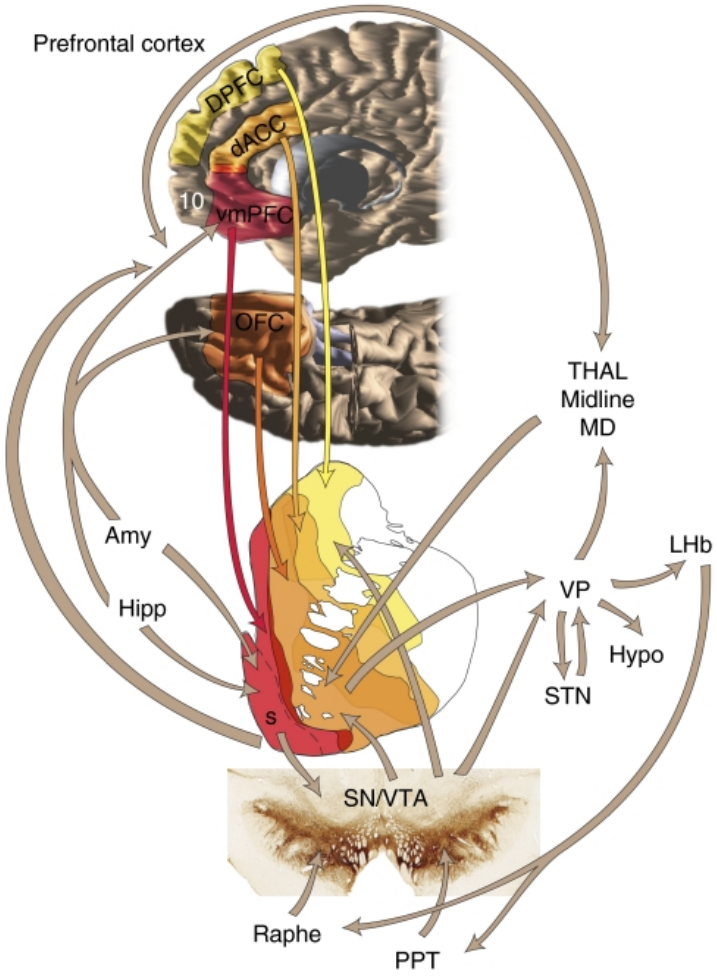


Figure 1.6: The reward circuit
(Adapted from Haber & Knutson, 2010)

As already discussed, many FC areas project to the striatum but there is variation in which striatal regions FC areas project to. In Fig. 1.6 the arrows indicate the main projections. The red arrow denotes the striatal input from the vmPFC (this is the same region that I referred to as vmPFC/mOFC earlier in this chapter), the dark and light orange arrows respectively indicate the striatal input from the OFC and the dACC while the brown arrows denote the projections from the ventral striatum to the VTA which in turn, projects back to the striatum and to the FC through the medial dorsal (MD) nucleus. Some other brain structures such as the amygdala (Amy), the hippocampus (Hipp), the raphe nuclei, the SN and the lateral habenula (LHb) are also interconnected with some of the same brain areas.

Famous experiments conducted more than 60 years ago were some of the first to implicate the dopamine system in reward coding (Olds, 1958; Olds and Milner, 1954). They showed that animals are happy to work not only for food but also for stimulation in certain brain regions. Food deprived rats were found to run faster and to respond up to 2000 times per hour and for 24hrs non-stop for brain stimulation instead of food – one rat even ignored food altogether for the pleasure of stimulation (Olds, 1958; Olds & Milner, 1954). Many of the stimulation sites in these experiments were in the mesolimbic pathway, a dopaminergic pathway that begins in the VTA and connects with NAc. In a more recent experiment by Hernandez et al., (2008) it was shown that electrical stimulation of the lateral hypothalamus led to an increase in dopamine levels in the NAc. Critically, however, changes in dopamine following intracranial self-stimulation only occur if the stimulation is unpredictable, suggesting that dopamine release is not encoding the rewarding aspect of the stimulation (Garris et al., 1999). In the following sections I explain how contiguity, contingency, and prediction errors affect both learning and dopamine neuron activity.

1.7.1 Reward identity

A large majority of decision-making studies that focus on reward value and the dopamine system do not examine the influence of reward *identity*. Conventional reinforcement learning theory implies that we should only learn about things that are better or worse than expected, regardless of the identity of the outcome (Niv & Schoenbaum, 2008). However, reward identity is clearly essential in many circumstances for guiding appropriate decision-making and for aiding learning about causes and outcomes (Izquierdo et al., 2004; Rushworth et al., 2011).

In fact, many regions innervated by dopamine do use reward identity

information to guide learning. In a study by McDannald et al. (2011), three groups of rats (controls, OFC-lesioned, VS-lesioned) were initially trained to associate visual cues (CSs) with food stimuli (USs). USs differed either in their identity (e.g., grape versus banana flavoured pellets) or in their quantity (e.g., one grape pellet versus three grape pellets). In the next sessions, the visual cues were presented in compound with novel auditory stimuli and in conditions in which either the reward size ("value unblocking") or reward type ("identity unblocking") was changed whereas in others the reward was the same as during learning ("blocking"). By comparing response rates following presentation of the novel auditory stimulus in extinction sessions, it was found that the OFC was necessary for learning that is influenced by changes in the rewards' identity but not value. In contrast, VS was found to hold an important role for both changes in reward identity and value.

A recent study by McDannald et al., (2014) showed that OFC neurons do take into account the identity of rewards. To demonstrate this, McDannald et al., (2014) examined OFC neuron responses to stimuli that predicted either a change in the value of an outcome (for example, because its size was changing) or a change in the identity of an outcome that involved no change in the outcome's value. OFC neurons reflected such differences in reward identity even in the absence of changes in reward value. Both studies provide evidence regarding the role of the OFC in encoding information about the identity of different rewards that goes beyond abstract value.

In animal learning theory three important features of reward determine the impact they will have on learning: contiguity, contingency and prediction error (Fig. 1.7). Each of these aspects of reward is also reflected in the firing of dopamine neurons.

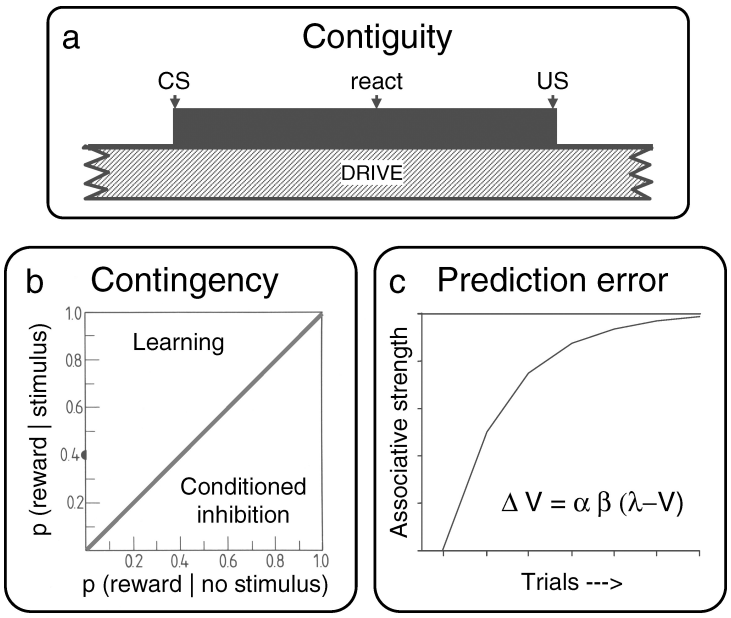


Figure 1.7: Behavioural functions of rewards: contiguity, contingency and prediction error
(Adapted from Schultz, 2006)

1.7.2 Contiguity

Contiguity is based on the idea of two events happening close in time (Fig. 1.7a). This can also be determined by the temporal proximity of a conditioned stimulus (CS), for example, an auditory stimulus and an unconditioned stimulus (US), for example a food reward. However, contiguity alone does not mean that there is any causal relationship between CS and US. CS-US contiguity occurs when a US is presented after a previously untrained CS for a number of trials. Dopamine neurons that initially respond to the US will start to respond to the untrained CS after some paired trials (Schultz, 2006).

1.7.3 Contingency

Contingency is defined in terms of whether or not a US is delivered as a function of whether or not a CS also occurred or not (Murphy and Baker, 2004). It

refers to the conditional relationship between the CS and US (Fig. 1.7b). Optimal decision-making is often dependent on contingency-based learning because the presence of CS-US contingency determines the ability of animals to discern causal relationships between CSs and USs.

There is a CS-US contingency if the occurrence of the CS indicates that there is a higher probability for the US to happen than in the absence of the CS (Fig. 1.7b). Rescorla (1967) was one of the first to show that simply pairing a CS and US in time (contiguity) was insufficient to always ensure learning occurred and therefore suggested that it is vital for the CS to indicate a contingency with the US.

Neurons mediating learning should thus be able to discriminate between three types of CSs: conditioned exciters after which reward occurs at a higher frequency than when no CS is present, conditioned inhibitors after which reward occurs less frequently than when CS happens, and neutral stimuli that are not influenced by changes in reward frequency compared with no stimulus. Dopamine neurons appear to have these characteristics. They increase their firing rates in response to reward-predicting CSs, decrease their activity in response to conditioned inhibitors and maintain a constant activity level in response to neutral stimuli (Schultz, 2006). See Fig. 1.7.

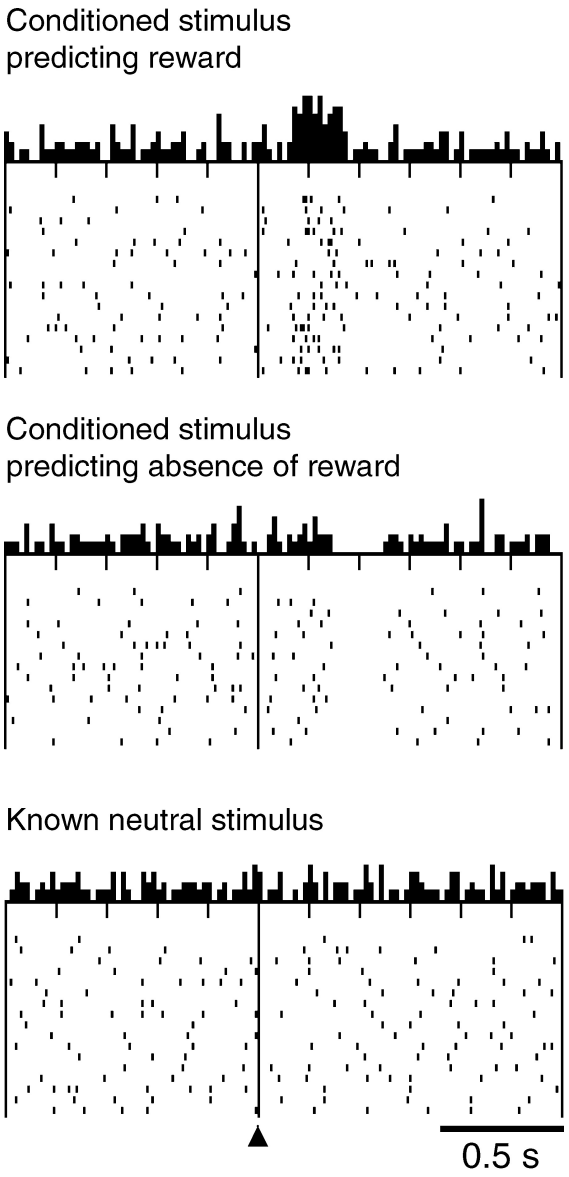


Figure 1.8: Neuronal behaviour under different contingencies
(Adapted from Schultz, 1998)

1.7.4 Prediction error

Prediction errors describe the difference between the reward that is expected when a choice is made and the reward that is actually experienced when the outcome is received. An agent should adjust its future expectations about rewards after an outcome is received in proportion to the size of the prediction error and the learning rate. A large prediction error indicates that the agent either knew little about what would follow the choice or that the agent's environment has changed significantly,

meaning that it must considerably revise its future expectations about the consequence of its choice. By revising its expectation, the agent becomes more likely to have more accurate expectations on future trials. Future prediction errors are also likely to be smaller. A small prediction error means that the agent was already able to predict the consequences of the choice and has little left to learn. Therefore, prediction error-based learning provides a way of maximising reward and minimising punishments (Niv & Schoenbaum, 2008).

The firing rate of dopamine neurons reflects prediction errors; it is quantitatively proportional to the difference between the "predicted reward" and the "experienced reward" associated with a particular event. More specifically, these neurons exhibit stronger activation when the US is unpredictable in time or amount as opposed to when the US was already predicted. In contrast, dopamine activity is depressed when the predicted reward fails to occur (Bayer & Glimcher, 2005; Cohen et al., 2012; Montague et al., 1996; Schultz, 2006; Schultz et al., 1997).

Caplin and Dean (2007) attempted to provide a more robust mathematical basis for defining what conditions should be met for a brain signal to qualify for a reward prediction error (RPE) categorization. The following three axioms were proposed:

Axiom 1. Consistent prize ordering

$$\delta(z, p) > \delta(z', p) \Rightarrow \delta(z, p') > \delta(z', p')$$

where δ is the RPE signal as a function of a prize z received from a lottery p. The first axiom suggests that if the RPE that follows a specific reward expectation (which Caplin and Dean refer to using the economic term "lottery") p which yields a specific

prize z is bigger than the RPE that derives from the same lottery which leads to a different prize z' then the RPE of a different lottery p' that yields a prize z should also be bigger than the RPE of that same second lottery p' when it yields the prize z'

Axiom 2. Consistent lottery ordering

$$\delta(z, p) > \delta(z, p') \Rightarrow \delta(z, p') > \delta(z', p')$$

This second axiom is similar to the previous one but the difference is that now the focus is on expectation or lottery consistency across different conditions. In this axiom, as well as in the previous one, if either of the two parts of the equation does not hold then we cannot have prize or lottery consistency across conditions.

Axiom 3. No surprise equivalence

$$\delta(z') = \delta(z)$$

The third axiom suggests that the response to the same RPE in the context of two different prizes should be equivalent. If for example an animal receives either one pellet or four pellets after a fully predictive CS the RPE should be zero in both cases because in both cases the reward (US) was fully expected and predicted.

Hart et al., (2014) investigated whether dopamine encodes the full range of RPEs in a manner consistent with these axioms. By using Fast-Scan Cyclic Voltammetry (FSCV) they recorded dopamine release in the nucleus accumbens (NAc) during a deterministic and a probabilistic decision-making task. In the deterministic task, rats had to choose between two levers each assigned the delivery of either one or four pellets. During the probabilistic task, rats again had to choose

between two levers but each lever was assigned to one of two lotteries. A 75% lottery led to four pellets in three quarters of trials and to one pellet otherwise whereas the 25% lottery let to four pellets in one quarter of trials and to one pellet otherwise. The results from the deterministic task showed that reward value (or prize) did not modulate dopamine release at reward delivery, fulfilling Axiom 3. The results from the probabilistic task provided strong evidence that the NAc dopamine signals are modulated both by reward (or prize) and condition (or lottery) meeting both Axiom 1 and 2. Additional analyses showed that the RPE signals are most robust in specific time windows when linear relationship can be demonstrated between RPE and extracellular DA.

The uncertainty of reward is also an important factor that influences learning. The reward becomes more uncertain when its probability approaches 0.5. In such a case it is unclear whether a reward will be delivered or not. Some aspects of dopamine neuron activity increase when the relation between CS and US is at p=0.5 (Fiorillo et al., 2003).

1.8 Learning theories and models

1.8.1 Decision Theory

Decision theory has an interdisciplinary background having close links with psychology, economics, computer science, philosophy, mathematics and statistics. Nowadays, it is also inextricably linked with a new, exciting field of neuroscience research better known as Neuroeconomics. It addresses many different aspects of decision-making, varying from simple perceptual decisions, such as when having to choose between two t-shirts of different colours to more complex decisions that

integrate probabilities, losses and uncertainty. A case of the latter is for example when having to choose between partners, jobs, university studies etc.

This theory makes 3 basic assumptions: First, there are multiple options to choose from. If there is only one option and you are forced to choose that, then the decision element is no longer evident. Second, the choices are made in a non-random way. If there is no logical pattern in the way you choose between option and the decisions are taken randomly then we cannot talk about decision-making but only about random behaviour. Third, these choices should be goal-directed. There are of course differences between accounts of how decisions should be made (normative decision theory) and accounts of how they are actually made (descriptive decision theory) (Anand, 1993).

1.8.2 Dual-process accounts of reasoning theory

According to 'dual-process' accounts, decisions are taken on the basis of two distinctively separate cognitive systems. System 1 is a non-logical, implicit and automatic system that operates effortlessly with no sense of voluntary control. System 2 is logical, explicit, controlled and responsible for more complex calculations where mental focus and computations are necessary (Stanovich & West, 2000).

A classic paradigm that demonstrates the prevalence of System 1 over System 2 comes from the 1980s when Amos Tversky and Daniel Kahneman ran their most famous and controversial experiment. This experiment, involved a fictitious lady called Linda and the participants were given some information about her age, attributes, studies, interests and beliefs. They were then given a list of eight possible scenarios and were asked to rank the scenarios based on their likelihood. Surprisingly, subjects ranked a scenario that provided more but inaccurate information (i.e. Linda

being a bank teller and a feminist) as more probable than a scenario that provided less but accurate information (i.e. Linda being a bank teller) (Tversky & Kahneman, 1983). An alternative version of the experiment, with a smaller number of scenarios helped subjects in another study to overcome the fallacy. This experiment marked the onset of a series of experiments both in humans and animals.

1.8.3 Incentive Theory

This theory claims that both humans and animals are often motivated to make specific actions when they expect to receive some rewards. People, for example, working in academia may be motivated to work for a new publication for the reward of gaining more recognition. In this theory, concepts such as reinforcement (reinforcement learning) and association (associative learning) play a vital role. In simple words, we choose based on the consequences (Balleine et al., 2010). Both intrinsic and extrinsic motives are therefore essential for action. In a classic experiment by Deci (1971) two groups of participants i.e. experimental and control group, had to reproduce specific puzzle configurations during three successive sessions. In the first session, both groups had to play freely i.e. without any reward. In the second session, participants were divided into two groups: an experimental group that received payment if they solved the puzzle and a control group that did not. In a third session the participants were again left to solve the puzzle but without any reimbursement. It was found that the experimental group spent less time playing with the puzzle configurations during the third and last session. He concluded that the intrinsic motivation is influenced by the presence of reward.

Incentive theory further suggests that motivational characteristics influence performance of certain actions. This is a consequence of changes in the

outcome representation. Any change in the reward or the stimuli affects the choice because they both influence the neuronal activation patterns that determine the choice (Balleine et al., 2010).

1.8.4 Reinforcement Learning

According to stimulus-response reinforcement learning, rewarding events reinforce direct relationships between contiguous events. This means that responses that are reinforced in the presence of a particular stimulus will be repeated when that stimulus is present on subsequent occasions. After a stimulus-reward association is reinforced over an extended period of time, the stimulus comes to directly elicit the response in a way that is no longer dependent on the response-outcome contingency or value (Niv & Montague, 2008). At this point the response is said to have become habitual.

This particular type of learning allows an agent to learn about the best choices to make even in the absence of explicit instruction and solely via interaction with the environment. It can guide an agent to the optimal decision even if the agent does not have any prior knowledge of the best choice on first entering the environment. A reinforcement-learning (RL) agent can learn to explore the environment, find where the rewards are and then learn an optimal policy for maximizing reward gain (Sutton & Barto, 1998). More specifically, in RL we have a sequence of actions and state transitions (s, a, s′,…) and in some states we have some reward associated with these. Without knowing the reward function R(s, a) and the transition model (P) of the world that describes the transition probability in going from state s to s′ when performing action a, the agent can learn to respond appropriately by monitoring the feedback that is obtained by interacting with the world (Sutton & Barto, 1998). A key component of

reinforcement learning is the prediction error. I have already discussed prediction errors in the context of dopamine and neural activity in the dopaminergic midbrain. In the following section, however, I review it briefly again from the perspective of reinforcement learning theory.

1.8.5 Rescorla-Wagner model

According to the Rescorla-Wagner (RW) model, learning can be explained by the following rule:

$$V_{new} = V_{old} + \eta(outcome - prediction) = V_{old} + \eta(R - V_{old})$$

where V is the prediction that is made on the basis of the observed stimuli and as a result of past experience with the same stimuli (e.g. an auditory stimulus that precedes food etc.). The learning rate η ($0 < \eta \leq 1$) indicates the extent that each experience influences future prediction and R denotes an abstract value of the outcome (which can, for example, be a pellet of food etc.) (Niv & Schoenbaum, 2008).

This model describes the circumstances under which Pavlovian conditioning occurs. The strength of a conditioned response (CR) is closely related to the likelihood of an unconditioned stimulus (US) taking place. CR is related to whether or not a conditioned stimulus (CS) has any predictive power over whether the US happens. Animals just like humans, seem to be sensitive to these contingencies. Contingency learning is described with the following formula:

$$\Delta p = P(US|CS) - P(US|{\sim}CS)$$

The delivery of USs during the intertrial interval (ITI) has deleterious effects on conditioning (Balsam & Fairhurst, 2006; Delamater, 1995; Rescorla, 1968). This can be explained because such reward deliveries, which are not predicted by CSs, reduce $\mathrm{p}(US|CS)$ and increase $P(US|{\sim}CS)$ so that the net effect reduces Δp. In a study by Delamater (1995) rats were trained in Pavlovian predictive associations, in which certain sounds predicted certain outcomes, and in this case sucrose pellets. After these associations were established, selective degradation of the contingent predictive relationship was attempted by providing reward both during sound and non-sound periods. It was shown subsequently that the degraded stimulus did not exhibit Pavlovian instrumental transfer. In other words, presentation of this stimulus no longer led to a selective increase of specific lever presses that had been previously linked with sucrose. The result is very important since it reveals that it is not just the pairing of a stimulus with an outcome that mediates its effects on choice, but also the context in which such pairings occur.

A significant disadvantage of this model though is that it does not take into consideration events that may take place during a trial, but only looks at what was the initial prediction, what actually happened, and whether or not the two agree. The prediction error is only apparent upon completion of the trial. Taking these disadvantages into account, Temporal Difference (TD) learning models provide a more "complete" perspective by taking into consideration all events happening between the initial prediction and the final outcome. Any new information from these events can lead to prediction errors if they change the expectations of when or what outcome will be reached (Niv & Schoenbaum, 2008).

1.8.6 Temporal Difference Learning model

This model explains that the difference between the expected and actual values of a certain state in a sequence of behaviours is used as a teaching signal to facilitate the acquisition of associative rules or policies to direct future behaviour (Schoenbaum et al., 2009; Sutton & Barto, 1987). Based on this model, the prediction at a timepoint *t* is as follows:

$$prediction(t)=Expected[outcome(t+1)] + Expected[outcome(t+2)] + \ldots$$

This equation says that if our prediction at a timepoint *t* is correct then this should be equal to the sum of all expected outcomes after that timepoint scaled by a temporal discount factor that allows for the fact that future rewards are not valued the same as current rewards. If our prediction is wrong, then the discrepancy is between what we originally predicted and what has actually happened in the new state together with the new prediction that can be made from the new state. The prediction error then is defined as:

$$\delta(t+1)=outcome(t+1) + prediction(t+1) - prediction(t)$$

Going back to the RW model this can now be transformed as:

$$Vnew = Vold + \eta[outcome(t+1) + prediction(t+1) - prediction(t)]$$

The TD learning model is more informative from the RW learning model as it takes into account all the details that happen during an event, their timing, and transitions between states (Niv & Schoenbaum, 2008; Schultz, et al., 1997).

1.9 Techniques

Three major neuroscience techniques were used for the purposes of this thesis: i. Functional Magnetic Resonance Imaging (fMRI), ii. Fast-Scan Cyclic Voltammetry (FSCV) and iii. Lesion studies.

1.9.1 Functional Magnetic Resonance Imaging

FMRI is a neuroimaging technique that measures brain activity based on changes to blood flow. It is based on the fact that when neurons in a brain area are active, even when a person or animal appears to be at rest, then the flow of oxygenated blood to that part of the brain is also increased. The main form of fMRI uses the blood-oxygen-level dependent (BOLD) contrast to provide an indirect measure of the neural activity; what we see is an increase in the inflow of oxygenated hemoglobin to the brain that correlates with activity of the neurons (Huettel et al., 2009). FMRI is not invasive and it has a relatively good 3D resolution although typically the way that fMRI scans are taken and processed means that it is used to inform us about activity changes occurring over several millimetres of neural tissue.

1.9.2 Fast-Scan Cyclic Voltammetry

Dopamine concentration in rodents has been traditionally measured with microdialysis. Microdialysis integrates the dopamine signal over space and time but with relatively low temporal and spatial resolution. Microdialysate samples are collected over 5 to 15 min, although environmental events tend to happen on a much faster timescale. This major limitation has been addressed by FSCV a powerful electrochemical technique that provides a high, subsecond temporal resolution (Clark et al., 2010). It also samples from a smaller extracellular region than microdialysis since a typical electrode is only 7μm x 150μm dimensions (Robinson & Wightman, 2007).

FSCV has become an essential tool for behavioural neuroscience that allows the investigation of the kinetics of molecules in the extracellular space of the rodent brain (Garris et al., 1997; Johnson & Wilson, 2007) especially in freely moving rodents (Robinson & Wightman, 2007). By using carbon-fiber microelectrodes analytes can be detected, and especially changes in the extracellular concentration of electroactive molecules can be monitored. These molecules represent the three major monoamine neurotransmitters, dopamine, serotonin and norepinephrine (John & Jones, 2010). Of particular interest for this thesis is the monoamine dopamine neurotransmitter.

Typically, for detection of dopamine release, the potential at the electrode's surface is driven against an Ag/AgCl reference electrode from a negative (-0.4V) to positive potential (+1.3V) and back every 100ms. Dopamine around the electrode is oxidised during the positive sweep at ~0.65V to form dopamine-o-quinone, with the loss of 2 electrons, and dopamine-o-quinone is reduced back to dopamine at ~-0.2V, with the gain of 2 electrons. This flux of electrons can be measured as current and it is

linearly proportional to the concentration of dopamine at the electrode (Phillips et al., 2003). The recorded current change at particular applied voltages is called a cyclic voltammogram. FSCV is a differential technique - i.e., it records *relative* changes in dopamine levels – as the background current which comes from the charging of the electrode's double layer, is subtracted to reveal finer changes in current due to dopamine's oxidation (Robinson & Wightman, 2007).

The voltammogram profile of dopamine is different from other electrochemical compounds such as serotonin (Heien et al., 2003) but similar to noradrenaline. While this similarity can be problematic in brain regions such as the prefrontal cortex that is rich in both dopaminergic and norepinephrinergic input, it is not an issue in the NAc core, for example, where there is only a very limited noradrenaline projection. To determine whether an electrode is sufficiently sensitive to detect dopamine, a cyclic voltammogram time-locked to the unexpected delivery of reward is extracted and compared against a "template" dopamine voltammogram. Unsignalled reward delivery reliably drives dopamine neuronal activity and dopamine release in the NAc core. If there is a high correlation between the two voltammograms ($r^2 > 0.75$), this implies that the electrode can detect dopamine release. A more robust overall method, however, is to use a principal components regression algorithm. Such an approach groups together differences in current that are due to dopamine from differences in current that are due to pH changes or noise (Robinson & Wightman, 2007). This approach is used in the experiments reported in some of the later chapters.

1.9.3 Lesion studies

A lesion is defined as any abnormality in the tissue of an organism that can be caused by disease or trauma. Of particular interest for this thesis are brain lesions. In humans, these may be the result of strokes, tumours, traumatic brain injuries or even of multiple sclerosis. Lesions in specific parts of the brain could lead to important differences in our decision-making. Studies with brain lesioned patients bring very important insights but studies with human subjects present certain obvious limitations. Most importantly, lesions in different patients are not identical and are usually not very focal.

A different type of lesion is the focal lesions in animals that is made intentionally to investigate the role of the lesioned part of the brain. These lesions fall into two categories: The first involves infusion of excitotoxin into specific brain regions. The second category entails direct removal of brain tissue often by aspiration. Such lesions disrupt fibers of passage leading often to additional cognitive deficits to those seen after excitotoxic lesions (Malkova et al., 1997; Rudebeck et al., 2013).

1.10 Thesis experiments

The following Chapters and Experiments are included in this thesis:

Chapter 2: The role of the ventromedial prefrontal cortex in decision-making

I will be presenting the results from a behavioural study in macaque monkeys, which aims to reveal or even challenge the current literature regarding the role of the vmPFC in decision-making. There is a general consensus that the brain of every living

organism aims to facilitate its survival. One way to accomplish this is through optimal decision-making. But could it be the case that in certain situations your brain fails to guide you correctly and instead of taking optimal decisions you end up being suboptimal? Can any suboptimal decision-making processes be identified within a particular brain region? If so will lesions to such areas actually make you a better decision-maker? I shall be presenting experimental evidence for just such a paradoxical facilitation in decision-making behaviour after vmPFC/OFC lesions.

Chapter 3: Less is more: Ventromedial prefrontal cortex and evaluation of compound options

Chapter 3 attempts to better characterise the key brain areas and neural signals involved in reward-guided decision-making when multiple facets of an outcome have to be taken into consideration in order to ascertain the best course of action. An fMRI version of the task used in Chapter 2 was therefore devised and performed by four macaques. The inability of the monkeys to take optimal decisions was once again evident. Most importantly, the neural correlates of this suboptimality were revealed. In addition, the anatomical boundaries of the vmPFC/mOFC in the macaque brain during reward-guided decision-making were identified as well as intriguing activation differences with the human brain.

Chapter 4: The influence of reward identity on mesolimbic dopamine release

Theories about dopamine suggest that it should not signal a change in the identity of an expected reward so long as there is no difference between the value of the expected and received rewards: it should simply care about whether something is better or worse than expected. To test this hypothesis, I used FSCV in rats to record

sub-second dopamine release in the NAc core during a 2-option / 2-reward instrumental task in situations where animals either received the reward they expected or where there was a surprising change in reward size or reward identity. I found that dopamine release was selectively increased in food-restricted rats on trials where they expected sucrose solution but received food, even though the animals had no overall preference for the food over the sucrose solution. By contrast, there was a small decrease in dopamine when they expected food but received sucrose solution. This implies that NAc core dopamine release does conform to a classical reward prediction error, but only if the reward signals are encoded in a frame of reference more closely aligned to physiological state than to subjective preference.

Chapter 5: Mesolimbic dopamine reflects state-dependent prediction errors

Based on the findings of Chapter 4, the same group of rats performed an identical task but now they did so after selective satiation in order to investigate in greater detail the influence of current physiological state on dopamine release. I found that dopamine signals rapidly updated following selective satiation in a manner consistent with a state-dependent prediction error.

2 The role of ventromedial prefrontal cortex in suboptimal decision-making

Summary

Decision-making research has shown that under certain conditions humans are quite irrational in their judgements. The use of heuristics often plays an important role in their irrationality. One of the most well studied heuristics is the so-called "affect heuristic" which can lead humans to prefer options that are less valuable or less optimal than others. I have tested a group of macaque monkeys to investigate whether the same behaviour is evident in non-human primates (NHPs). Macaques had to make choices between options that led to less valuable single rewards and more valuable options that led to multifaceted rewards. They exhibited an irrational decision-making behaviour, as they did not choose the multifaceted options in many cases. I also investigated the importance of the ventromedial prefrontal cortex/medial orbitofrontal cortex (vmPFC/mOFC) in generating the unusual pattern of behaviour. A second group of macaque monkeys with vmPFC/mOFC lesions was tested. These monkeys were less irrational than the control group, choosing the multifaceted options, which should objectively have been more valuable, more often. These results suggest that the vmPFC/mOFC plays an important role in representing values of multifaceted options.

2.1 Introduction

The behaviour of decision makers suggests that they assign values to the options that they choose between. However, the values that are assigned do not always appear to be rational suggesting that subjectivity is an important factor in decision-making. Philosophers from ancient times, such as Aristotle and Porphyry, and more recently behavioural psychologists, and economists have attempted to describe the limits of rationality in human decision-making (e.g. Kahneman & Tversky, 1979; Samuelson, 1937) using a variety of theories (e.g. Utility theory, Prospect theory) attempted to address this issue. There has been particular interest in the possibility that part of the reason why decisions are sometimes irrational is because people are following various simplifying heuristics when they are confronted by novel scenarios that require decision-making (Tversky & Kahneman, 1974).

Heuristics, although helpful under certain circumstances, can lead to biases and poor decision-making habits. One of the most well studied heuristics is the so-called affect heuristic (see Denes-Raj & Epstein, 1994; Ganzach, 2001; Hsee & Kunreuther, 2000). This is often used when having to choose between different options of varying value. What happens in such cases is that feelings of happiness, satisfaction, pleasantness or unpleasantness that specific stimuli, events or scenarios generate influence our ability to make accurate judgements. Studies with human primates have repeatedly demonstrated the prevalence of this heuristic in decision-making but relevant research with non-human primates (NHPs) is limited and rather inconclusive. My first aim in this experiment was to design a paradigm that could investigate the presence (or absence) of the same phenomenon in NHPs.

My hypothesis was that NHPs and specifically macaque monkeys might use a similar heuristic to guide their decisions. My second aim was to identify which brain

regions play a key role in the specific example of suboptimal decision-making. Activity in human ventromedial prefrontal cortex (vmPFC) co-varies with value (Abitbol et al., 2015; Howard et al., 2015; Lebreton et al., 2015; Li et al., 2016; Noonan et al., 2011). Analyses of the time course of such activity and of the relation between inter-subject variability in signal size and variation in decision-making behaviour suggest it may guide decision-making (Boorman et al., 2009; Chau et al., 2014; Hunt et al., 2012; Jocham et al., 2014; Jocham et al., 2012; Kolling et al., 2012). It was additionally found to be more active while experiencing positive emotions (Winecoff et al., 2013) and during evaluation processes in social tasks (Smith et al., 2014a). Single-unit recording studies in monkeys (Monosov & Hikosaka, 2012; Padoa-Schioppa & Assad, 2006) also showed that the vmPFC and adjacent medial orbitofrontal cortex (mOFC) is more active during reward processing.

At the same time, however, when lesions are made in vmPFC and adjacent orbitofrontal cortex (OFC) in macaques they do not impair all value-guided decisions (Rudebeck and Murray, 2011), casting doubt on the general role of vmPFC in reward-guided decision-making. For example, even when lesions are restricted to central OFC areas 11 and 13 they impair performance in reward devaluation tasks in which animals have to update their expectations about the value of a specific type of reward that a choice will deliver on the basis of their current levels of satiety (Rudebeck and Murray, 2011), but even large lesions of OFC do not disrupt decisions in other simpler stimulus-reward association tasks (Rudebeck et al., 2013).

In many situations human patients with vmPFC/mOFC lesions can be suboptimal in their decisions, opting for choices that yielded immediate gains even if these were smaller and failing to take into account higher future losses (Bechara et al., 2000). In general they were more prone to be misguided in their decisions (Asp et al.,

2012). In addition, such patients have important difficulties when making choices that involve uncertain or ambiguous outcomes (Fellows & Farah, 2007). In general research with human vmPFC/mOFC lesions patients has demonstrated that such lesions make you worse at a number of aspects of decision-making. But what happens, when you use a paradigm in which normal healthy control subjects make the wrong decisions? Could it be the case that the absence of this region could switch your decisions into a better direction? In order to investigate this intriguing hypothesis, macaques with vmPFC/mOFC lesions were used in an attempt to gain a better understanding of the role of this critical decision-making brain region. I have used a two-choice six-option decision-making task, in which macaque monkeys had to choose between two alternative options of different value each time.

2.2 Methods

2.2.1 Subjects

Six adult macaque monkeys (Macaca mulatta, 3 males and 3 females) weighing between 6.2 and 13.2kg and between 8 and 9 years of age participated in the experiment. Two of the three female monkeys had bilateral vmPFC/mOFC lesions. They were housed either in groups of two or groups of four and kept on a 12-hr light cycle, with access to water 12-16h on testing days and with ad libitum access to water on non-testing days. All procedures were conducted under licences from the United Kingdom (UK) Home Office in accordance with the UK The Animals (Scientific Procedures) Act 1986.

2.2.2 Surgery and Histology

Both groups of monkeys were treated with a steroidal anti-inflammatory [20mg/kg methylprednisolone, intramuscularly (i.m.)] and an antibiotic (8.75 mg/kg amoxicillin, i.m.) a minimum of 12hrs 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 6hr intervals. At the morning of the surgery, they 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 (0.2 mg/kg meloxicam), atropine (0.05 mg/kg) and opioid (0.01 mg/kg buprenorphine). These aimed to reduce secretions and provide analgesia. They were further treated with an H2 receptor antagonist (1 mg/kg ranitidine) for gastric ulceration protection due to the administration of both steroidal and non-steroidal anti-inflammatory treatments. They were later moved to the operating theatre where they were intubated, switched onto isoflurane inhational anaesthesia (1-2% in 100% O_2) and placed in a head holder. Their head was shaved and cleaned using alcohol and antimicrobial scrub. A midline incision was made, the tissue was retracted in anatomical layers and a bilateral bone flap was removed. All lesions were made by aspiration with a fine-gauge sucker.

Throughout the surgery, respiration rate, heart rate, body temperature, blood pressure and expired CO_2 were continuously monitored. Once the lesion was made, the wound was closed in anatomical layers. Non-steroidal anti-inflammatory analgesic (0.2. mg/kg meloxicam, orally) and antibiotic (8.75 mg/kg amoxicillin, orally) were administered for a minimum 5 days after the procedure. The surgeries were carried out under sterile conditions and with the aid of a binocular microscope and the wound was closed in anatomical layers.

After the end of their testing, both mOFC/vmPFC lesioned animals were anaesthetised with sodium pentobarbitone and perfused with 90% saline and 10% formalin. The brains were then removed and placed in 10% sucrose formalin. The brains were blocked in the coronal plane at the level of the most medial part of the central sulcus. Each brain was cut in 50-µm coronal sections. Every tenth section was retained for analysis and stained with Cresyl violet. Eight coronal sections through the frontal lobes and eight coronal sections through the temporal lobes are shown in Fig. 2.1.

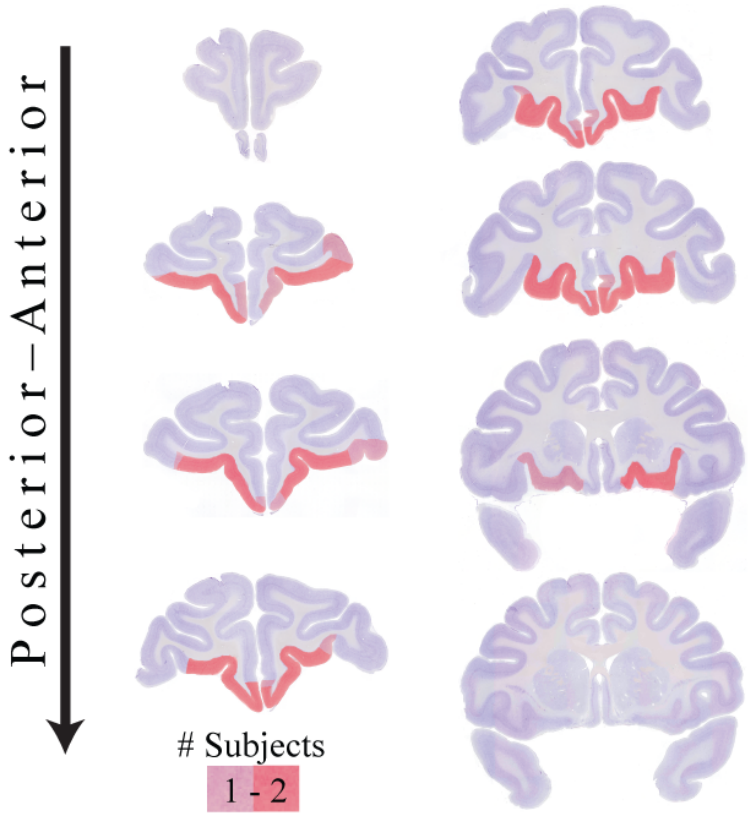


Figure 2.1: Histology
MOFC/vmPFC lesion locations represented on an unoperated control, with redness indicating actual lesion overlap in two animals.

Examination of the histology confirmed that the lateral limit of the vmPFC/mOFC was the lateral orbitofrontal sulcus while the medial limit was the

inferior bank of the rostral sulcus dorsal to the gyrus rectus. The anterior limit was an imaginary line between the anterior tips of the lateral orbital sulcus and the medial orbital sulcus and extending onto the medial surface to the vicinity of the anterior tip of the rostral sulcus. The posterior limit was an imaginary line between the posterior tips of the lateral orbital sulcus and medial orbital sulcus and the extending onto the medial surface in the vicinity of the posterior tip of the rostral sulcus.

2.2.3 Behavioural training

At the initial phase of the experiment all monkeys were separated from their group mates in a smaller part of their home cage and in separate but consecutive days they were given freely either 30 pieces of fruit or the same amount of vegetable. The aim was to find appropriate food items that the monkeys were happy to consume. If the monkeys were happy to consume the offered food then the next phase of the experiment commenced; otherwise a different food was used. The experiments involved choices between stimuli associated with either a piece of fruit, a piece of vegetable, or a "Compound" comprising the same amounts of both the same fruit and the same vegetable. The fruit option used for all monkeys was grape but a range of vegetables was used (carrot, sugar snap peas and cucumber: although only one type was used with a given animal) because preferences for vegetables differed between monkeys. These specific foods were used because they are not often included as standard parts of the animals' daily diet. All fruit/vegetable pieces were prepared so as to be of the same weight and approximately of the same shape. Four of the monkeys that weighed between 6.2 and 9.1kg were given food pieces that weighed 4-5g each and the other four monkeys that weighed between 9.9 and 13.2kg were given food pieces weighing 7-8g each.

All monkeys had previous experience on decision-making tasks that involved the use of wooden targets. For this experiment, they were trained on a new set consisting of 14 targets in total that differed in shape and colour pattern (Fig. 2.2a). Each target was assigned different reward contingencies across monkeys (see example stimulus-reward contingencies in Fig. 2.2b).

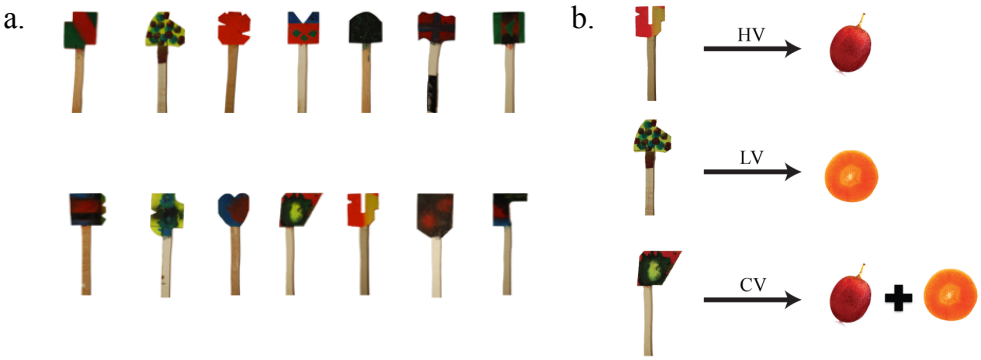


Figure 2.2: Stimuli information
(a) All 14 stimuli used during the experiment. **(b)** Example Stimulus-Reward associations (HV: High Value option, LV: Low Value option, CV: Compound Value option).

During training, the experimenter stood across the testing table, opposite to the monkey that was sitting inside a transport box (62 x 52 x 45cm) and presented the monkey simultaneously with a pair of targets. One target led to reward (CS+) and the other target was not associated with any reward (CS-). See Fig. 2.3.

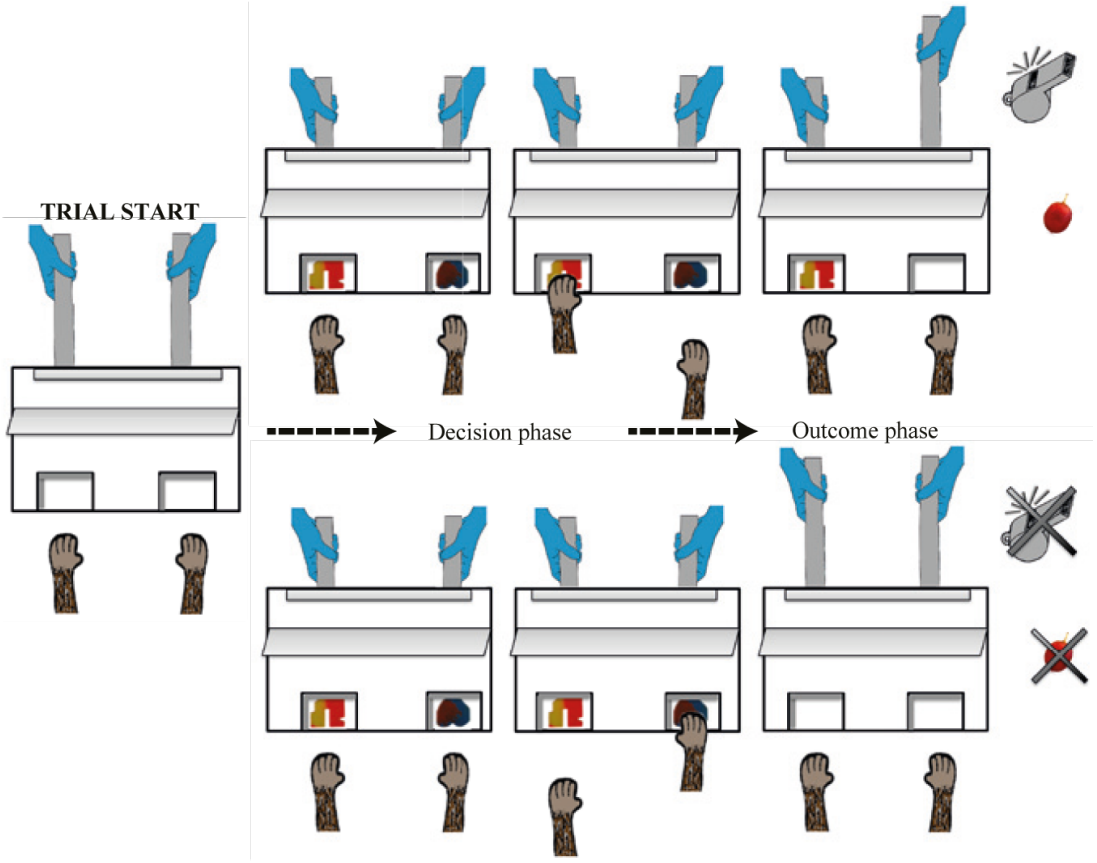


Figure 2.3: Experimental setting
The top row demonstrates an example sequence of events when an animal chose the rewarding stimulus (CS+); the bottom row demonstrates a similar sequence of events, but when an animal chose the non-rewarding stimulus (CS-).

The reward was either a piece of fruit or vegetable. The order of learning these contingencies was counterbalanced between monkeys. The presentation side of the targets (right/left) was assigned pseudorandomly and no target could be presented more than two consecutive times on the same side. The monkey made a decision by touching one of the two targets. In the case of choosing a CS+ target, the experimenter blew on a whistle, then the unchosen option was withdrawn and the reward associated with the CS+ was offered. Once the monkey took the reward the CS+ target was instantly withdrawn as well. If the monkey chose the CS- target, then both targets were withdrawn, there was no whistling, and an ITI period of ~10s was starting. The ITI period in all CS+ trials was based on the consumption time.

2.2.4 Stimulus-reward learning

Each stimulus-reward learning session during the learning phase (S1) consisted of 30 rewards. A monkey needed to complete 30 correct trials to receive all rewards. If the CS- target was chosen, the trial was repeated until the monkey picked the CS+ target. A criterion was set for a minimum 80% (or 30 out of 37 trials) of trials to be performed correctly within a single session for determining whether a monkey had learned the stimulus-reward contingencies well. Half of the animals were taught the stimulus-fruit associations first and then the stimulus-vegetable associations, and the other half of the animals were taught in the opposite order. In each case a new CS+ was used which was paired with a CS-. Once a monkey learned the stimulus-reward contingencies for the fruit and the vegetable, a third pair of targets was introduced for the Compound option. A Compound option consisted of both the fruit and the vegetable delivered together. If a monkey looked reluctant to take the less valuable option during the Compound reward, the experimenter held it in front of the monkey until it was taken. Because all monkeys preferred the fruit option to the vegetable option we referred to these two options as the "High Value" option and the "Low Value" option respectively. Thus, in total, all animals learned about a High Value (HV), a Low Value (LV), and a Compound Value (CV) option composed of the same amounts of the HV and LV options. Between trials all food rewards were placed on a plastic bowl in front of the experimenter but hidden from the view of the monkey.

To minimise the possibility of inadvertent cuing of the monkeys by the experimenter, the experimenter was trained to perform stereotyped movements during testing. He stood behind the testing table so as that the height of his eyes was ~50cm higher than the height of the monkey's eyes, preventing his gaze from meeting with

the monkey's gaze when the latter had to make a choice. The experimenter looked downwards at the centre of the testing table and observed the monkey's choices via peripheral vision.

After learning the different stimulus-reward contingencies, revision sessions (S2, S4, S7, S9, S11) with all pairs were given to ensure that the monkeys remembered the different stimulus-reward contingencies. When the same learning accuracy criterion was reached (80% correct performance in a single session) the animal moved onto the main task. See Table 2.1.

S1	S2	S3	S4	S5	S6
Learning phase	**Revision sessions**	**Main task**	**Revision sessions**	**Main task**	**Learning phase**
Set A	**Set A**	**Set A**	**Set A**	**Set A**	**Set B**
1-4 days	*1-2 days*	*2 days*	*1-2 days*	*1 day*	*1-4 days*
S7	S8	S9	S10	S11	S12
Revision sessions	**Main task**	**Revision sessions**	**Main task**	**Revision sessions**	**Devaluation phase**
Set B	**Set B**	**Set A**	**Set A&B**	**Set A|B**	**Set A|B**
1-2 days	*3 days*	*1-2 days*	*3 days*	*1-2 days*	*2 days*

Table 2.1: Experimental timeline
From S1-S12 the order of the different Stages of the experiment is presented.

2.2.5 Main task

Once the animals learned the associations between the various CSs and the HV, LV, and CV rewards during the Learning phase (S1) and reached the criterion during the Revision sessions (S2), their choices between the CSs were assessed. In the main task (S3), for the first time, the macaques were presented with a choice between: i. CS+s associated with the HV and LV rewards, ii. CS+s associated with the CV and

LV rewards, iii. CS+s associated with the HV and CV rewards and iv. CS+s associated with HV, LV, and CV juxtaposed against CS-s associated with the non-delivery of HV, LV and CV reward respectively. As in the previous sessions (S1, S2) if a monkey chose the CS- target over its CS+ pair, the trial was repeated until the monkey corrected its error. A revision session (S4) was also given in between the first two days (S3) and the last day (S5) in the main task with the first set of stimuli (Set A) to ensure that the monkeys remembered all the Stimulus-Reward associations.

After the third day on the main task, a different vegetable was introduced, and the monkeys had to learn the new stimulus-reward associations for the new CV and LV option (Set B) during a new Learning phase (S6). As before, once the monkeys learned these associations, revision sessions (S7) were given and once the learning accuracy criterion was met, the monkeys moved into the next three days of the main task (S8). After the sixth day, revision sessions (S9) were given for the LV and CV stimuli-reward pairings that had been used in the first three sessions (S3, S5) of the main task, and once the learning accuracy criterion was again met the monkeys moved into the last three days of the main task (S10). These testing days included all the stimuli and reward types (Set A&B) learned during the previous days and presented under all possible combinations. The full set of stimuli therefore comprised one HV option (grape), two LV options (two different vegetables), and two CV options (grape combined with the first vegetable and grape combined with second vegetable).

After the completion of the main task, all monkeys were given revision sessions (S11) that included only the fruit, one of the two vegetables as well as their respective CV option. Once they reached the learning accuracy criterion they continued with the Devaluation phase (S12).

2.2.6 Devaluation

Devaluation procedures are often used in the context of investigations of goal-based decision-making and the role of the OFC. In such procedures animals are allowed to feed to satiety on one food type so that its value to the animal decreases. In other variants on the procedure, the food item is devalued by pairing with nausea. Critically, when devaluation is used to assess goal-based decision-making, it is imperative that animals make choices between CS+s associated with each food type without experiencing the food itself subsequently; this is necessary to ensure that the animals are making choices on the basis of internal representations of the expected outcomes that have been revalued in the absence of direct experience of the particular choice and the outcome in new sated state. In order to ensure that this is the case, studies with rodents often examine decision-making during extinction (when rewards are not actually delivered to the animals) (Gallagher et al., 1999; Pickens et al., 2005; Schoenbaum and Roesch, 2005; Stalnaker et al., 2015). Studies of goal-based decision-making conducted in macaques avoid giving macaques repeated experience of a given CS+ choice in the context of the food type by training the monkeys on multiple pairs of stimuli and ensuring that any given CS+ is only experienced once during the critical decision-making test when CS+s associated with each of the two foods are paired against one another (Baxter et al., 2000; Malkova et al., 1997; Rudebeck et al., 2013).

By contrast, here the aim of the devaluation procedure was quite distinct. It was not intended to investigate goal-based decision-making on the basis of internal representations of reward expectations. Instead the aim was quite simply to decrease the value of the vegetable/fruit option. We were interested in investigating the possibility that the compound stimuli that were comprised of the vegetable option

would become even less valuable than a high value fruit in isolation. Each devaluation session was conducted ~24h after the previous feeding opportunity (so that more than one food type was not inadvertently devalued). A food box (22 x 10 x 14.5cm) was placed in the monkeys' home cage filled with 250g of the vegetable/fruit option that was to be devalued. The monkey was free to consume the food for 25 minutes without being directly observed. The experimenter then entered the room and if most of the food was consumed, an additional 150g of the same food was added. After five minutes the experimenter started observing the animal through the monkey's housing room window until the monkey refrained from consuming any food for five minutes. The food box was then removed from the home cage and the remaining food was weighed at the end of the session. In all sessions, 35 minutes were sufficient for the monkeys to complete this procedure. The monkey was then taken from its home cage and moved into the testing room in a transport box. The main phase of the testing started within five minutes from the completion of the selective satiation procedure.

During this phase, a task similar to the one described in *Main task* was given. The devaluation sessions were conducted with a 4 to 6 days gap in between and the order of these was counterbalanced across monkeys. See also Table 2.1 for a schematic representation of the different stages of this experiment.

2.3 Results

2.3.1 Preliminary training

All the monkeys were content to consume the grape and at least two different vegetables when these were freely available in their home cage. When given a choice, nearly all the animals preferred the grape to either of the vegetables. But when they

had to work to obtain the same rewards, by learning to touch wooden targets associated with the food, one of the control monkeys refused to work for the vegetable option of sugar snap peas. Therefore, an alternative vegetable was used. In a single case a control monkey showed no preference for the fruit over the vegetable and so an alternative vegetable of a lower relative value was used.

During the devaluation task, the control group consumed on average more fruit and more vegetable than the vmPFC/mOFC group (309.25 ± 70.25g compared to 247.5 ± 53.03g and 295 ± 35.37g compared to 224.5 ± 38.9g, respectively) but these differences were not statistically significant $F_{1, 4} = 2.282$, $p = 0.205$.

An independent samples t-test showed that the control monkeys and the vmPFC/mOFC monkeys did not differ in the number of sessions they needed initially to learn the stimulus-reward associations for grape ($t_4 = 1.044$, $p = 0.355$), the first vegetable ($t_4 = -1.633$, $p = 0.178$), the second vegetable ($t_4 = 0.795$, $p = 0.471$), the first Compound ($t_4 = -2$, $p = 0.116$) or the second Compound reward ($t_4 = -1.633$, $p = 0.178$). Also there was no difference between the two groups in the number of revision sessions they needed during the different parts of the task ($F_{1, 4} = 0.267$, $p = 0.633$).

2.3.2 Main task

A repeated measures ANOVA with a Huynh-Feldt correction showed that the control monkeys responded differently across the three different choice types (i.e. HV versus LV, CV versus HV and CV versus LV) ($F_{1.063, 3.189} = 385.402$, $p < 0.0005$) during the 10 testing days. There was also a marginally significant interaction between the three different choices types and the days of testing ($F_{8.338, 25.015} = 2.276$, $p = 0.054$). A one-sample t-test against 50% on the mean scores from the 10 days of

testing showed that the control monkeys were choosing the HV more often than the LV option (97.5 ± 4%, $t_3 = 23.718$, $p < 0.0005$), the CV than the LV option (92.5 ± 4.6%, $t_3 = 18.304$, $p < 0.0005$), but surprisingly, the CV less often than the HV option (17.76 ± 4.1%, $t_3 = -15.709$, $p = 0.001$; Fig. 2.4).

A repeated measures ANOVA across the 10 testing days with the CV versus HV option pair with the group (vmPFC/mOFC versus control animals) as a between-subjects factor, showed a statistically significant effect of group ($F_{1, 4} = 21.734$, $p = 0.01$). A repeated measures ANOVA with a Huynh-Feldt correction showed that the monkeys responded in a different way during the 3 different choice types ($F_{1.372, 5.489} = 288.930$, $p < 0.0005$), and there was a statistical significant interaction between the pair types (HV versus LV; LV versus CV; HV versus LV) and the group ($F_{1.372, 5.489} = 5.921$, $p = 0.049$).

Although none of the monkeys appeared satiated during the sessions and consumed most of the food, a separate analysis was run without the last 10 trials from each session because we reasoned that any general satiety effects that might have confounded results would have affected those trials most. A repeated measures ANOVA across the 10 testing days showed an even stronger statistically significant effect of group on the CV versus HV option pair ($F_{1, 4} = 35.272$, $p = 0.004$).

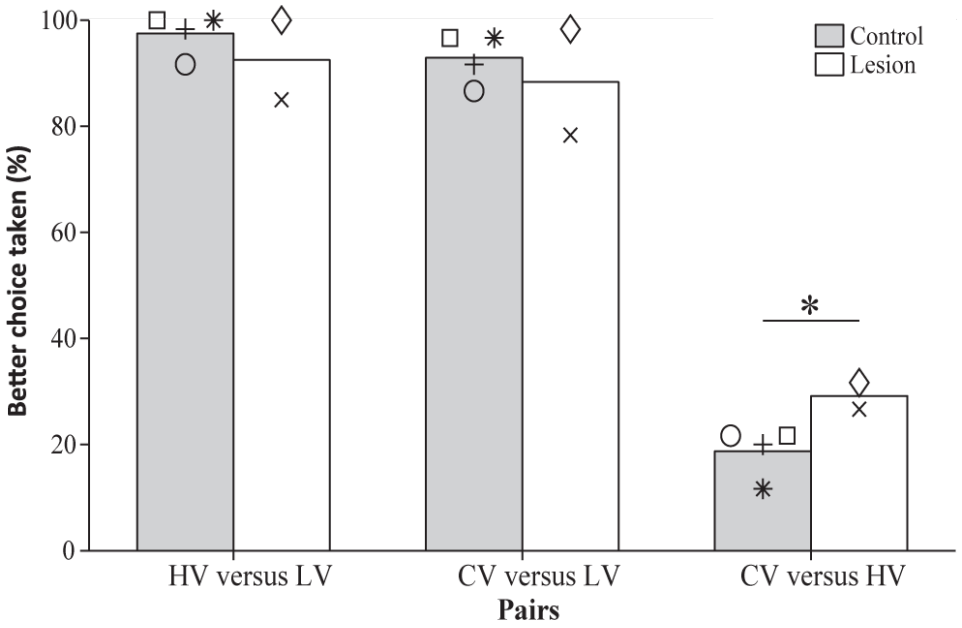


Figure 2.4: Choices between pairs across all sessions
Frequency of choosing the objectively better of the two options. Controls (grey bars) as well as lesioned animals (open bars) preferred HV- to LV- stimuli and CV- to LV-stimuli. For the HV vs CV decision, strikingly, controls preferred HV-stimuli, receiving only a subset of the rewards that CV-stimuli would have offered. However, macaques with vmPFC/OFC lesions did not prefer CV- to HV-stimuli as much as controls (*$p < 0.05$).

2.3.3 Devaluation

A repeated measures ANOVA with a Huynh-Feldt correction during the two devaluation sessions (Vegetable devaluation, Fruit Devaluation) showed that there was a statistically significant difference in the way that the animals made their choices across the different pairs (i.e. HV versus LV, CV versus HV and CV versus LV), ($F_{2, 8}$ = 16.839, p = 0.001). There was also a statistically significant interaction between those pairs and the two sessions ($F_{2, 8}$ = 9.915, p = 0.007) and a significant interaction between pairs, days and group ($F_{4, 8}$ = 4.036, p = 0.044). The pattern of results can be better understood by looking more closely at performance on each of the sub-tests of the devaluation procedure (vegetable and fruit devaluation). While it is difficult to know what pattern of results to expect in the fruit devaluation procedure we hypothesised that devaluation of the vegetable might lead to a CV option that was

even less attractive to control animals than the HV option. This is indeed what we found.

2.3.3.1 Vegetable devaluation

An independent samples t-test during the vegetable devaluation session, showed that the vmPFC/mOFC lesioned monkeys were choosing the CV over the HV option (45 ± 21.21%) more frequently than were the control monkeys (5 ± 5.77%), t_4 = -3.939, p = 0.017. In summary, we replicated the difference in vmPFC/mOFC lesion versus control behaviour although now the effect was numerically stronger. In contrast, during the same devaluation session, a one-sample t-test against 50% showed that the control monkeys were choosing the HV over the LV option (97.5 ± 5%, t_3 = 19, p < 0.0005), the CV over the LV option (87.5 ± 15%, t_3 = 5, p = 0.015) and less often the CV over the HV option (5± 5.77%, t_3 = -15.588, p = 0.001). See Fig. 2.5.

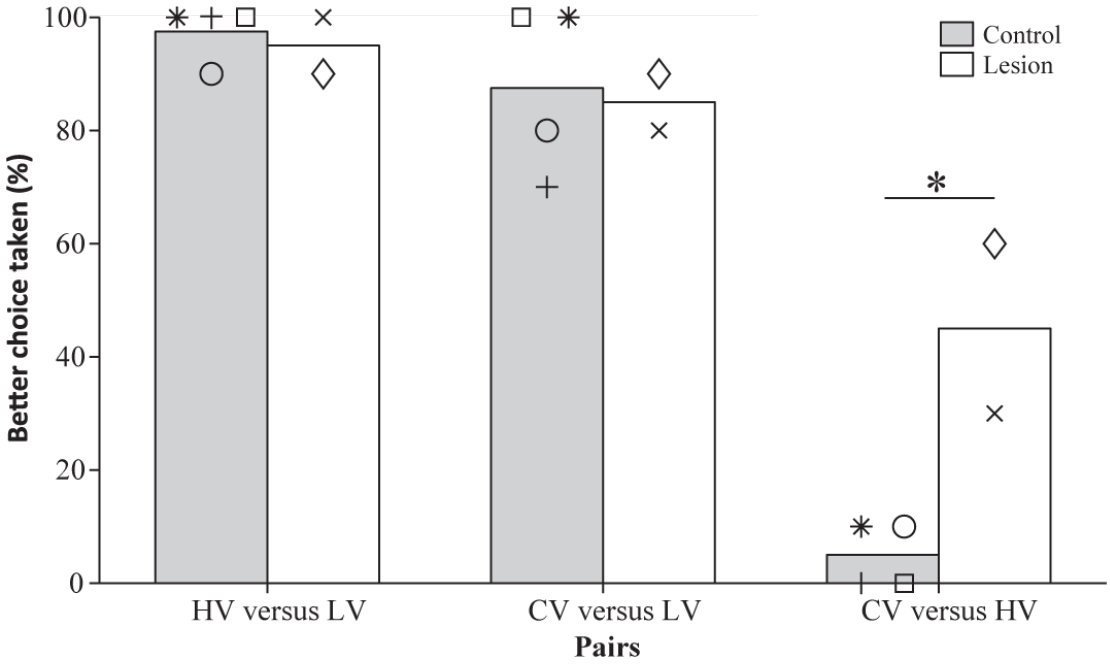


Figure 2.5: Choices between pairs during the vegetable devaluation session
Frequency of choosing the objectively better of the two options, during the vegetable devaluation session. The pattern of results was replicated, including group difference in HV versus CV decisions, even after HV versus CV decisions were made easier by satiating animals with LV rewards prior to testing (*$p < 0.05$).

2.3.3.2 Fruit devaluation

During the fruit devaluation session only the CV versus HV option pair for the control group reached statistical significance but with a reversed pattern, as the CV option was chosen more often than the HV option (80 ± 14.14%, $t_3 = 4.243$, $p = 0.024$) demonstrating optimal behaviour on behalf of the control animals. On average there was a decrease in the number of times the animals chose the fruit related options (i.e. HV, CV) over the LV option but these differences were not statistically significant. In those pairs there was an overall shift of the choices towards the mean (Fig. 2.6).

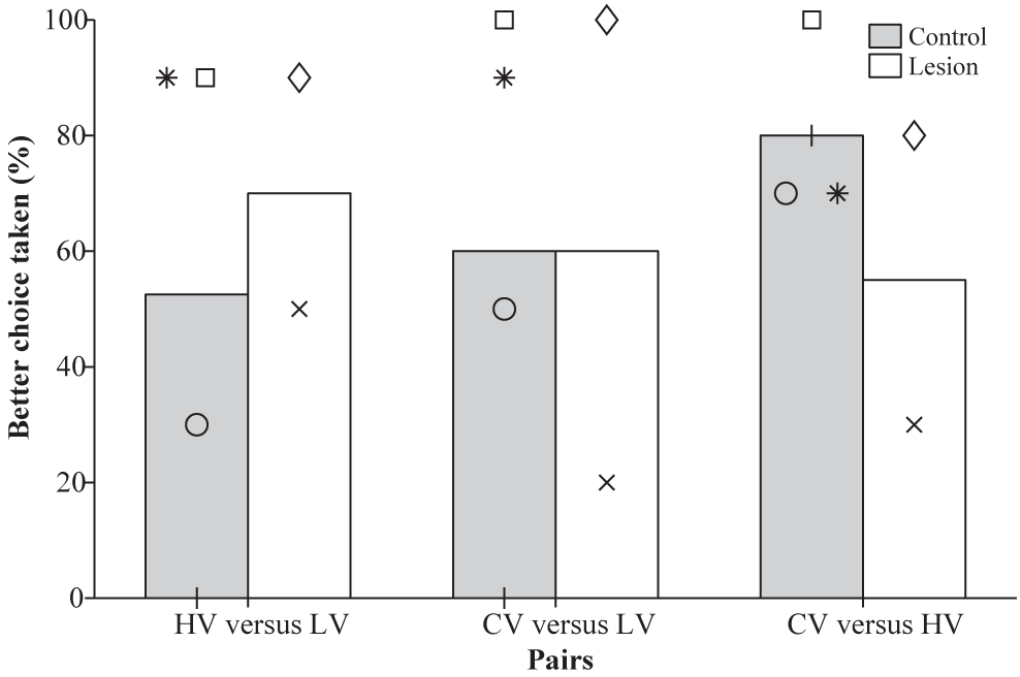


Figure 2.6: Choices between pairs during the fruit devaluation session
An overall shift of the choices towards the mean was observed, as decisions were made harder by satiating animals with HV rewards prior to testing.

2.4 Discussion

When macaques evaluate CS+s that are associated with multi-component reward outcomes then it seems that their evaluations are biased away from the sum of the value of the components and towards an average estimate of the components' values. This was apparent in the striking behaviour of the four control animals that fairly consistently chose a CS associated with a high value (HV) food option (a piece of fruit) in preference to a CS associated with an identical outcome plus a less preferred outcome (a piece of vegetable - I refer to this option as the compound value option or CV). The fact that, on interleaved trials, the same control monkeys chose a CS that was associated with just the vegetable outcome (a low value or LV choice) when was pitted against a CS- with no reward association demonstrated that although the LV option was less preferred its value was still positive. In addition, the control

animals appeared content to eat the LV option. When the value of the vegetable component was made even lower, by devaluing it by letting animals feed to satiety on the option prior to testing, then the control animals' preference for the HV option over the CV became even more pronounced.

The evaluations of the multi-component compound outcomes depended on vmPFC/mOFC. When lesions were made in vmPFC/mOFC in two macaques they exhibited a significantly different pattern of behaviour when choosing between HV and CV options. Instead of exhibiting a consistent preference for the HV option, like the control macaques, their choice pattern suggested that they were more indifferent between the CS+s.

Humans have the tendency to average the value of groups of items instead of summing them up. Although this may be beneficial in a number of circumstances it may sometimes lead to suboptimal decision-making. More specifically, during evaluations of options comprising a bundle of items human subjects were willing to pay more for an option comprised of fewer components than a similar option comprising the same components plus a few extra, but low value components (Hsee et al., 1998; List et al., 2002). For example, collectors were inclined to pay more for a set of high value baseball cards than they were for an identical set of high value cards together with some additional low value cards (List, 2002).

There has been some evidence for similar irrationality in chimpanzees that were given choices between two visible food options that consisted of one highly preferred option and a mixture of this highly preferred option with a less preferred option. The effects, however, were inconsistent and appeared to depend on reward size (Beran et al., 2009).

Previous experiments with macaque monkeys have also provided some equivocal and inconsistent evidence for a similar effect in macaques (Kralik et al., 2012). The strength, consistency, and sustained nature of the effects found in my experiment may be due to fact that we assessed choices based on reward expectations elicited by arbitrary CSs, whereas Kralik et al. (2012) examined choices that were made directly between the food items themselves. Kralik et al., (2012) showed that this effect was evident but not quite as resilient when tested in a laboratory environment. More specifically, during their experiment three monkeys were trained to press buttons to choose between two food options. Once a button was pressed, the respective reward was given to the animal. After being tested for three sessions the results showed that: a) a similar effect to the one that I showed was evident for one monkey; b) a second animal showed the effect during the first two thirds of the sessions; c) for the third animal the effect disappeared during a third testing session.

The vmPFC/mOFC has been widely considered to signal the value of future rewards (Abitbol et al., 2015; Bouret & Richmond, 2010; Gottfried et al., 2003; Howard et al., 2015; Lebreton et al., 2015; Li et al., 2016; Monosov and Hikosaka, 2012; Noonan et al., 2011; O'Doherty et al., 2001; O'Doherty, 2011; Plassmann et al., 2007). Analyses of the time course of such activity and of the relation between inter-subject variability in signal size and variation in decision-making behaviour suggest it may guide decision-making (Basten et al., 2010; Boorman et al., 2009; Chau et al., 2014; Hunt et al., 2012; Jocham et al., 2014; Jocham et al., 2012; Kolling et al., 2012; Philiastides et al., 2010; Rushworth et al., 2011). At the same time, however, when lesions are made in vmPFC/mOFC in macaques they do not impair all value-guided decisions (Rudebeck and Murray, 2011) casting doubt on a general role for vmPFC/mOFC in reward-guided decision-making. For example, even when lesions

are restricted to central OFC areas 11 and 13 they impair performance in reward devaluation tasks in which animals have to update their expectations about the value of a specific type of reward that a choice will deliver on the basis of their current levels of satiety (Rudebeck & Murray, 2011) but even large lesions of OFC do not disrupt decisions in other simpler stimulus-reward association tasks (Rudebeck et al., 2013).

The discrepancy between the results of human neuroimaging studies and macaque lesion studies might be explained in a number of ways. First, there is uncertainty about the anatomical identity of the vmPFC region in which reward-guided decision-making activity arises in humans and its relation to vmPFC and OFC areas in other primates such as macaques (Wallis, 2012). It is therefore possible that macaque lesion studies might not have investigated the homologue of the critical region identified in human neuroimaging studies. Second, it is possible that the behavioural tasks used in the two species do not assess comparable cognitive processes. It is notable that several human neuroimaging studies that have reported vmPFC activity during value-guided decision-making tasks that employ multi-faceted choice options. For example, several studies have employed choices that vary both in terms of their associated reward probability and reward magnitude but vmPFC activity is only prominent when decision-making depends on consideration of both facets of each choice (Hunt et al., 2012) and the complementary abilities to integrate information about multiple aspects of a choice outcome into a unified representation (Barron et al., 2013) and to focus on the most relevant aspect of each choice when required (Hunt et al., 2014). One reason why the decision-making behaviour of animals with vmPFC/mOFC lesions was significantly different to control behaviour in

the present experiment may well have been because the outcomes in the present task were multicomponent outcomes.

Because the value of both the HV and LV option is above zero then logically one might expect the value of the CV option to be highest of all because it was composed of the same amounts of food as the HV and LV options but in conjunction. Surprisingly, however, control monkeys consistently preferred the HV to the CV option. Interestingly, monkeys with vmPFC/mOFC lesions exhibited a less strong preference for the HV over the CV option. In other words, they were more indifferent between the HV and CV options. During the vegetable devaluation session, when the value of the vegetable component of the CV option was devalued by pre-feeding, the control animals were even more likely to choose the HV over the CV option but vmPFC/mOFC lesion animals chose at chance between the two options. Because the control monkeys' choices between HV and CV options were arguably suboptimal then an interesting feature of the vmPFC/mOFC lesions is that they could be described as making monkeys less suboptimal. In contrast, during the fruit devaluation session when the value of the fruit component of the CV option was devalued by pre-feeding, the control animals exhibited optimal behaviour by choosing the CV over the HV option, whereas the vmPFC/mOFC animals kept choosing at chance between the two options. This may imply that when the value of the components of the CV are closer to each other, the control animals, in which vmPFC/mOFC is intact, are able to estimate the value of the CV option in a more adaptive manner by summing the value of the additional components.

Despite the surprising nature of the control animals' preference for the HV option over the CV option, there may be a sense in which the behaviour might be thought of as adaptive and possibly even optimal. For a foraging animal it is

sometimes possible to maximise the rate of food intake by foregoing opportunities to take a low value food, if it entails a handling cost, and just focus on opportunities for a higher value food (Stephens and Krebs, 1986). It is possible that a fuller understanding of vmPFC function, might be achieved by adopting such an ecologically grounded approach (Pearson et al., 2014).

While the reward expectation elicited by the stimulus paired with the CV option might have been expected to be the sum of the reward expectations associated with the HV and LV options, the pattern of behavioural results suggest instead that in control animals the reward expectation actually lies in between that associated with the HV and LV options. Rather than being equal to the sum of the value expectations associated with HV and LV options, the value expectation elicited by the CV option appeared biased towards the mean of the expectations associated with the HV and LV options. This effect was less obvious in the vmPFC/mOFC lesion monkeys and in some circumstances the vmPFC/mOFC lesion monkeys were indifferent between the HV and CV options suggesting that they found it more difficult to distinguish their values.

A possible taste interaction effect between the fruit and the vegetable that could lead to distaste for the CV option is considered unlikely for a number of reasons. First of all, macaque monkeys are omnivores and typically include a variety of fruits and vegetables in their diet. Second, animals ate both the HV and LV foods when these were delivered freely into their home cage and therefore both appeared to have a positive value for the animal. Third, ate both the fruit and the vegetable when they were choosing the CV option. Fourth, the CV option had a higher value than the LV option as macaques consistently chose the former over the latter. Fifth, the monkeys were tested with different vegetables demonstrating that the effect is

generalisable. Sixth, during the fruit devaluation, when the value of the HV option decreased, the control monkeys switched their preferences towards the CV option. Seventh, the monkeys were food controlled when they were tested and so they should have been motivated to maximise reward.

3 Less is more: Ventromedial prefrontal cortex and evaluation of compound options

Summary

Activity in the human ventromedial prefrontal cortex (vmPFC) reflects the value of potential choices (Boorman et al., 2009; Chau et al., 2014; Hunt et al., 2012; Kable & Glimcher, 2007) and so it may have a role in value-guided decision-making. Because lesions that included vmPFC/mOFC disrupted the ability of macaques to decide between choices associated with compound outcomes in Chapter 2 an fMRI version of the same paradigm was designed in order to identify the precise region within vmPFC/mOFC that was associated with task performance. This also made it possible to investigate whether activity in other brain areas was also associated with the same type of decision-making. A strong vmPFC/mOFC signal was found at the time that the cues were presented on the screen. Activity in vmPFC/mOFC at the decision time was higher when animals had to make a choice between an option that led to a highly valued reward (HV option) and an option that led to a compound comprising the same option together with a lower value option (CV). Disruption of such activity might explain the behavioural impact of the vmPFC/mOFC lesions reported in the previous chapter and corroborate the notion that under certain circumstances vmPFC activity is linked to an apparently suboptimal pattern of decision-making. In addition, activity in an adjacent vmPFC/mOFC region reflected decision difficulty, the relative values of the choices available on each trial, and the value of the chosen stimulus. These results provide evidence regarding the vmPFC/mOFC areas involved in value-guided decision-making in the macaque and

demonstrate its role in decision-making including in apparently suboptimal decision-making when the HV option is chosen in preference to the CV option.

3.1 Introduction

The ability to take good decisions is central to successful living and well-being. An optimal decision maker should choose the highest value option regardless of the complexity of the situation in which he finds himself. In some situations (Hsee, et al., 2001; Kahneman, 2011; Kahneman & Tversky, 1979), however, humans are surprisingly poor decision-makers. In several cases these situations have involved participants choosing between options with multifaceted outcomes. In such situations a choice that leads to a compound outcome comprising both high valued and low valued components may be valued less than an outcome simply comprising the same high value element in isolation. For example, baseball card collectors may value a choice leading to a single rare card more than they value a choice leading to a composite comprising the high value and low value elements (List, 2002).

A general consensus is that the brain takes decisions by assigning values to available options and consequently makes choices as a result of comparing these values (Bossaerts et al., 2008; Kable & Glimcher, 2009). The ventromedial prefrontal cortex/medial orbitofrontal cortex is a brain region in which activity reflects the value of potential choices (vmPFC/mOFC) (Bouret & Richmond, 2010; Litt et al., 2011). Surprisingly this region of the human brain is less active when people choose between different options on the basis of their values. However, despite this decision-related decrement in activity, vmPFC/mOFC's activity also covaries with the difference in value between chosen and the unchosen options; when the value difference vmPFC activity increases too (Boorman et al., 2009; FitzGerald et al., 2009; Rudorf & Hare, 2014; Rushworth et al., 2011). VmPFC/mOFC activity does not only reflect the value of expected gains but it also reflects the value of expected losses (Tom et al., 2007).

Activity also decreases when delays are introduced between decision and reward delivery (Kable & Glimcher, 2007).

Decision-making often involves weighing the benefits and the costs of choosing or rejecting an option. Successful cost-benefit integration could therefore lead to reward maximization. The vmPFC/mOFC as well as the anterior-medial prefrontal cortex (amPFC), anterior cingulate cortex (ACC), and dorsolateral prefrontal cortex (dlPFC) have been considered candidate regions for cost-benefit comparisons (Basten et al., 2010; Croxson et al., 2009; Hunt et al., 2015; Kennerley et al., 2009; Rudebeck et al., 2006; Rushworth et al., 2011). Choices that are closer together in terms of value are more difficult and take longer (Fitzgerald et al., 2009) and are associated with higher dlPFC activation (Rudorf & Hare, 2014). DlPFC is also recruited when choices are more time-consuming when there are additional cognitive demands as a result of the need for value integration (Jocham et al., 2014).

Important insights into the role of vmPFC/mOFC in reward-guided decision-making come from lesion studies (for example, Fellows, 2011; Malkova et al., 1997; Rudebeck & Murray, 2011, 2014; Noonan et al., 2010). In Chapter 2 of this thesis, I presented data from two groups of animals: controls and vmPFC/mOFC lesioned animals, while performing a decision-making experimental paradigm. Such studies can provide information about the causal role of the lesioned areas in behaviour. However, it is also desirable to examine more precisely how, where, and when the activity in the brain area arises. In order to link the results to both human fMRI studies and experiments with macaques I therefore devised an fMRI version of the behavioural task used in the previous chapter in which macaques learned associations between stimuli and multiple reward outcomes while in the MRI environment. Our aim was again to investigate decisions more similar to the multiple component

decisions that engage human vmPFC using the same neuroimaging technique that is used in human research.

I expected that it might be possible to relate variations in activity on different types of decision-making trials to the manner in which decision-making impairments arose after lesions. For instance, a choice between two highly valued options might lead to a different pattern brain activation than a choice between a high value and a low value option.

3.2 Methods

3.2.1 Subjects

Four adult male macaque monkeys weighing 10.8 - 13.4kg and 8 years of age participated in the experiment. Three of the monkeys had participated also in Experiment 1.

3.2.2 Behavioural training

During this experiment all monkeys were trained on a modified version of the task described in Chapter 2, but inside the fMRI scanner. The wooden targets were replaced with clipart pictures displayed on a screen placed ~30cm in front of the head-fixed animal and the food outcomes were replaced with juice outcomes. A stimulus could be presented on either or both sides of the screen and the monkey could press one of the two infrared sensors placed in front of his left and right hand to make a choice of the spatially adjacent stimulus (Fig. 3.1).

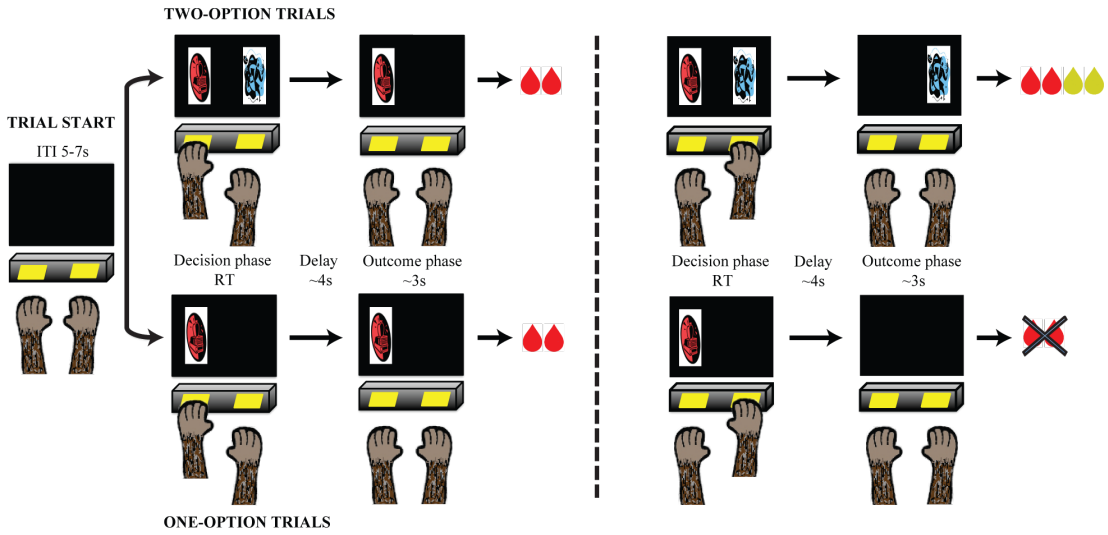


Figure 3.1: Experimental setting
Two option trials: During those trials (top row) an animal could choose between two stimuli, each one associated with a single reward (left) or a compound reward (right).
One-option trials: On these trials (bottom row) a single stimulus was presented on one side of the screen and an animal could either choose it by touching the button placed in front of that or reject it by touching the button in front of the blank side of the screen.

Six stimuli were used in total: animals learned about three pairs of stimuli where one of the stimuli was associated with the High Value (HV), Low Value (LV), or Compound Value (CV) reward options. In each case choosing the other stimulus in the pair led to no reward (thus there were three unique CS- stimuli including the CS- that was used as the foil when animals learned which CS+ predicted the HV outcome, the CS- that was used as the foil when animals learned which CS+ predicted the CV outcome, the CS- that was used as the foil when animals learned which CS+ predicted the LV outcome). Stimuli were counterbalanced across animals but the stimulus-reward associations were constant for a given animal across all sessions.

In order to find appropriate juices for this experiment, all monkeys were separated from their group in a smaller part of their home cage. In the first stage, the experimenter extended a syringe containing 50ml of a juice towards the animal's mouth and delivered a small volume of juice. A second juice was delivered subsequently in the same way. In both cases the experimenter carefully observed

whether the animal was happy to consume the offered juices. At the second stage, the experimenter gave the animal the opportunity to choose the juice that he most preferred, by extending both syringes at approximately equal distances from the left and right side of the head of the monkey. The presentation side of these syringes was changed every 1 to 2 trials. The monkey could discriminate the juices not only by their taste and smell but also by the colour of the syringes.

Three distinct juices were used: grapefruit, strawberry and blackcurrant. One juice served as the HV option and a second juice as the LV option. In both cases juice delivery lasted approximately 500ms. The CV option consisted of identical amounts of each of the two juices (in other words, each of the two juices that comprised the outcome were delivered for 500ms separated by a 300ms interval). In other words, just as in Chapter 2, the outcome of choosing the CV option corresponded to the sum of the outcomes of HV and LV choices. The order of the juice delivery in the CV option was counterbalanced so that in the half of the trials juice A was delivered first and then juice B and in the other half the opposite order of presentation was used.

The task consisted of 120 trials, from which 75% (or 90 trials in a given session) were one-option trials: only one stimulus was presented on either the left or right side of the screen. The other 25% (or 30 trials) were two-option trials: two stimuli were presented on either side of the screen. As a result of the extensive training prior to scanning, all monkeys learned that they could skip a given one-option trial by touching the sensor that was placed in front of the blank side of the screen. From the two-option trials, 50% of the trials consisted of choices between the CS+ and CS- target-pairs and the remaining 50% of the trials, consisted of choices between the CS+ targets.

Each trial began with a blank screen [inter-trial interval (ITI): 5-7s] and at the end of the ITI one or two stimuli were presented on the screen. During both one-option and two-option trials, if the CS+ stimulus was chosen, all stimuli disappeared from the screen but ~4s later the chosen stimulus re-appeared and the juice was delivered (outcome phase: 3s). Each reward was composed of two 0.5ml drops of juice delivered by a spout placed near the monkey's mouth. If the blank side of the screen or a CS- stimulus was chosen, during one-option and two-option trials respectively, then all stimuli disappeared from the screen, and an additional delay of 4s was added to the ITI period.

3.2.3 Imaging data acquisition

Imaging data were collected using a 3T MRI scanner and a four-channel phased-array receive coil in conjunction with a radial transmission coil (Windmiller Kolster Scientific, Fresno, CA). FMRI images and reference images for artefact corrections were collected while awake monkeys were head-fixed in a sphinx position in an MRI compatible chair. FMRI data were acquired using a gradient-echo T2* echo planar imaging (EPI) sequence with 1.5x1.5x1.5mm^3 resolution, TR = 2.28s, TE = 30ms, flip angle = 90°. Proton-density-weighted images using a gradient refocused echo (GRE) sequence (TR = 10ms, TE = 2.52ms, flip angle = 25°) were acquired as reference for body motion artefact correction. T1-weighted MP-RAGE images (0.5x0.5x0.5mm^3 resolution, TR = 2500ms, TE = 4.01ms) were acquired in separate anesthetised scanning sessions (for details, see Sallet et al., 2013).

3.2.4 FMRI data preprocessing

FMRI data were corrected for body motion artefact by an offline-SENSE

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

3.2.5 FMRI data analysis

Whole-brain analysis was conducted using a univariate General Linear Model (GLM) approach with FMRIB's Software Library (FSL: *11*). We searched for brain regions that exhibited activity at the cue onset and at the reward delivery time. To do this we applied two GLMs to every testing session.

The first GLM (GLM-1) included 24 regressors and the second GLM (GLM-2) included ten regressors (Table 3.1). In, GLM-1 we used nine constant regressors, each time-locked to onset of a decision type of interest (HV versus LV decisions, HV versus CV decisions, etc.) and, in addition, a nuisance regressor for discarded choice trials. A further set of these ten regressors were used in the analysis that were time-locked to the reward delivery period as opposed to the onset of cue presentation. Four additional regressors indexed events at the time of response. Two regressors indexed whether responses were made with either the left or the right hand and both were, like nearly all other standard regressors, convolved by the standard macaque

hemodynamic response function (HRF): a gamma function with 3s mean and 1.5s variation (Chau et al., 2015). Two other regressors indexing left and right responses were not HRF-convolved as part of an additional attempt (alongside the other procedures described above) to capture movement related artifacts that might have an immediate impact on the BOLD signal rather than an impact mediated by the HRF convolution.

GLM-2 employed a complementary parametric analysis approach. One regressor, time-locked to cue presentation, indexed the sum of the values of the options offered to the animals and another one, again time-locked to cue presentation, indexed the difference in value between the choice taken and the choice rejected. Both these regressors were only used on trials on which a CS+ stimulus was chosen (respectively these regressors are listed below as 'Chosen + Unchosen value' and 'Chosen - Unchosen value'). In addition to the parametric value regressors, another two regressors were used to index all cue trials on which a CS+, a CS with a specific reward association, was chosen ('CS+ Chosen') and all trials on which a CS-, a CS with no reward association, or the blank screen was chosen ('CS- /Blank side Chosen'). Two versions of each of these regressors were employed. One was time-locked to onset of cue presentation and one was time-locked to the reward-delivery period.

GLM-1 regressors	GLM-2 regressors
Cue Onset time	
HV versus Blank Screen	CS+ Chosen
CV versus Blank Screen	CS-/Blank side Chosen
LV versus Blank Screen	
HV versus CS-	Chosen + Unchosen value
CV versus CS-	Chosen - Unchosen value
LV versus CS-	
HV versus CV	
HV versus LV	
CV versus LV	
Discarded trials (i.e. choices with RT > 1SD)	
Reward delivery time	
HV versus Blank Screen	CS+ Chosen
CV versus Blank Screen	CS-/Blank side Chosen
LV versus Blank Screen	
HV versus CS-	
CV versus CS-	
LV versus CS-	
HV versus CV	
HV versus LV	
CV versus LV	
Discarded trials (i.e. choices with RT > 1SD)	
Response time	
Left Hand	Left Hand
Right Hand	Right Hand
Left Hand without HRF	Left Hand without HRF
Right Hand without HRF	Right Hand without HRF

Table 3.1: General Linear Model regressors for GLM-1 and GLM-2

GLM-2, like GLM-1, contained four motion-related nuisance regressors capturing the effects of the monkey responding by either a right or a left button press either with or without HRF-convolution. As in GLM-1 these regressors control for motion-related neural activity and for motion-related image artifacts, respectively.

Additional contrasts were constructed for both GLM-1 and GLM-2. In GLM-1 a contrast for the 'Critical comparison' reflected the difference between activity associated with the critical decision pair 'HV versus CV' that had been most affected by the lesion in Chapter 2 and the 'Non-critical pairs' that included the 'HV versus LV' and the 'CV versus LV' pairs that were less affected by the lesion in Chapter 2. Other contrasts were used to capture activity related to the taking of 'Hard decisions', 'Easy decisions'. Hard decisions were ones on which there was a small difference in value between the options that were available and monkeys' preferences were unclear and inconsistent. In other words, these were decisions on which one choice was taken on some trials but the other choice was taken on other trials (LV versus its corresponding CS-; LV versus the blank screen). Easy decisions were ones on which there was a large difference in the values associated with the options available and animals had consistent preferences (HV versus its corresponding CS-; CV versus its corresponding CS-; HV versus the blank screen; CV versus the blank screen). A 'Difficulty' contrast compared activity on difficult and easy trials. An additional set of contrasts related to the transitivity of decisions ('Transitivity A', 'Transitivity B', and 'Transitivity C') was employed that will be explained in more detail in Results below (Table 3.2). Finally, two contrasts captured activity related to the making of decisions when only when CS+ was available on a trial ('One-option trials') or two CS+s were available ('Two-option trials').

For GLM-2, a contrast captured neural activity related to the onset of cue presentation on decisions ('Decision') and another two contrasts captured the difference in activity on trials on which a CS+ stimulus was chosen rather than a CS- stimulus both at the cue onset ('Stimulus value') time and at the reward delivery time ('Reward') (Table 3.2):

GLM-1 contrasts	GLM-2 contrasts
Cue Onset time	
Non-Critical pairs = HV versus LV + CV versus LV	Decision = CS+ Chosen + CS- Chosen
Critical comparison = HV versus CV – Non-critical pairs	Stimulus value = CS+ Chosen – CS- Chosen
Hard decisions = LV versus CS- + LV versus Blank side	
Easy decisions = HV versus CS- + CV versus CS- + HV versus Blank side + CV versus Blank side	
Difficulty = Hard decisions – Easy decisions	
Transitivity A = HV versus CV – CV versus LV	
Transitivity B = HV versus CV – HV versus LV	
Transitivity C = HV versus LV – CV versus LV	
One-option trials = HV versus Blank side + CV versus Blank side + LV versus Blank side	
Two-option trials = HV versus CS- + CV versus CS- + LV versus CS- + HV versus CV + HV versus LV + CV versus LV	
Reward delivery time	
	Reward = CS+ Chosen – CS- Chosen

Table 3.2: General Linear Model contrasts for GLM-1 and GLM-2

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

and $p < 0.05$ cluster-corrected (Worsley et al., 1992).

3.2.6 Timecourse analysis

To illustrate the significant activity in vmPFC/mOFC, we placed a region of interest (ROI) over the peak of the vmPFC/mOFC signal at the time that the visual stimuli were presented on the screen and extracted the timecourse of the BOLD signal from two-voxel radius spherical masks. For statistical analyses we have extracted the time course of activity from the period corresponding to the full-width-half-maximum (FWHM) of the peak established using a leave-one-out procedure to avoid bias. A one-sample t-test and paired-sample t-test were used (two-tailed).

3.2.7 Chosen/Unchosen value analysis

For the purposes of this analysis a single value was assigned to each CS+, CS-, and Blank screen. The value of each CS+ was based on the following calculations:

i. The percentage of occasions that each of the 3 CS+ (HV, CV and LV option) were chosen during the one-option trials where the animals could choose between a CS+ or forgo the trial by touching the sensor in front of the blank side of the screen. This percentage was then multiplied by the number of trials on which the specific choice was presented in this manner in each day's testing session (i.e. 30 times).

ii. The percentage of occasions that each of the CS+ were chosen during two-option trials when an animal could choose between a CS+ and a CS-. This percentage was then multiplied by the number of trials on which this specific choice was presented in this manner in each day's testing session (i.e. 5 times).

iii. The percentage of occasions that each of the CS+ were chosen during two-option trials on which an animal could choose between two CS+ (for example trials on which they could choose between. HV versus LV options). This percentage was then multiplied by the number of trials on which the specific choice was presented in each day's testing session (i.e. 5 times with each of the other two CSs+).

In the final stage of value calculation each stimulus' value for an individual animal was calculated by summing the outputs of the three calculations described above and dividing by the total number of trials on which the choice had been offered.

For example, if an animal chose the HV stimulus in 95% of the HV versus Blank side of the screen trials, 90% in HV versus CS- (5 trials), 80% in HV versus CV (5 trials) and 85% in HV versus LV (5 trials), then the HV stimulus was assigned the parametric value of 0.92 for this animal. The relevant equation and example calculations are presented below:

$$HV = ((HV \text{ versus Blank side}) * \#trials + (HV \text{ versus CS-}) * \#trials + (HV \text{ versus CV}) * \#trials + (HV \text{ versus LV}) * \#trials)/Total\#trials$$

In the example given above the HV option's parametric value was calculated as follows:

$$HV = ((0.95*30) + (0.9*5) + (0.8*5) + (0.85*5))/45 \Rightarrow$$

$$HV = 0.92$$

The value assigned to the blank side of the screen corresponded to the average frequency, from all three one-option trials, on which an animal chose the blank side of the screen over the CS+. Similarly, the value for each of the CS- corresponded to the frequency with which each was chosen in preference over its respective CS+.

3.3 Results

3.3.1 Behavioural Results

All four monkeys had a strong preference for one juice (HV juice) over the other juice (LV juice) during both pre-scanning (t_3 = 19, $p < 0.0005$) and scanning sessions (t_3 = 17.301, $p < 0.0005$; Fig. 3.2). First, we assessed whether we could replicate the "less is more" effect that we had seen in the experiment in Chapter 2 in the control group; we examined whether macaques preferred the HV to the CV option even though the CV option was composed of the same amounts of each of the HV and of the LV juice rewards. A one-sample t-test against 50% during the last two pre-scanning sessions showed that the monkeys had, on average, a statistically significant preference for the HV over the CV option (t_3 = -3.873, $p = 0.030$). In addition, the macaques preferred the HV over the LV option t_3 = 19, $p < 0.0005$) and the CV over the LV option (t_3 = 8.878, $p = 0.003$). Again, during the subsequent scanning sessions there was, on average, strong preference for the HV over the CV option (t_3 = -13, $p = 0.001$), for the CV over the LV option (t_3 = 5, $p = 0.015$) and for the HV over the LV option (t_3 = 19, $p < 0.0005$) (Fig. 3.2 – upper panel).

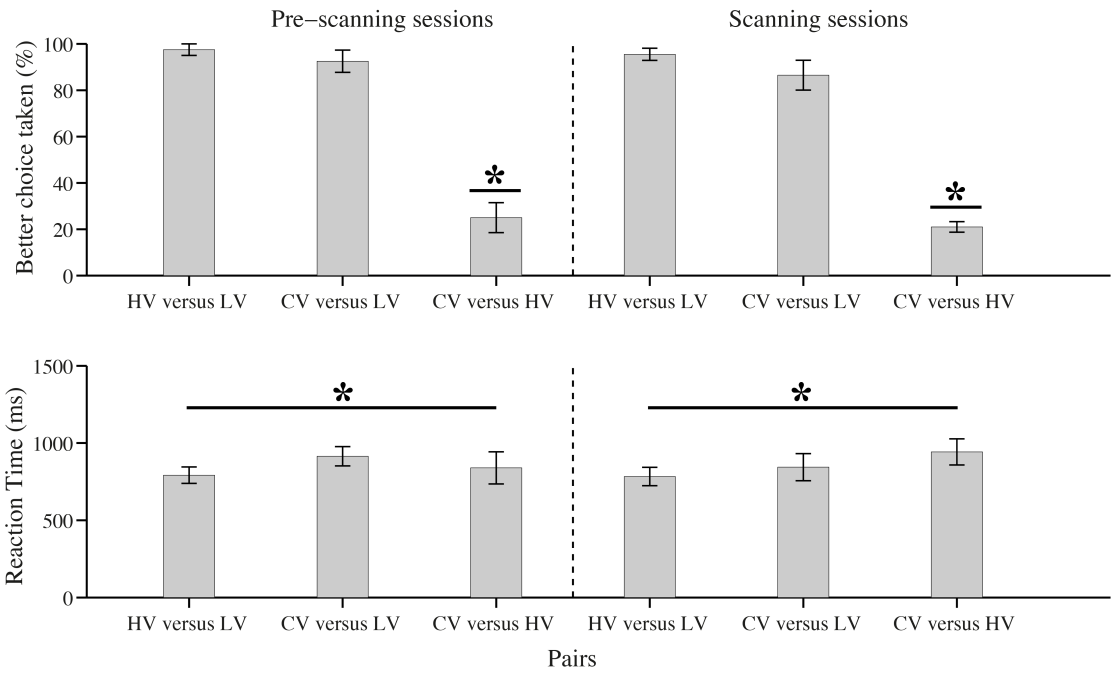


Figure 3.2: Macaque behaviour before and during scanning
Upper panel: Choice behaviour when choosing between CS+ stimuli.
Bottom panel: Reaction time when choosing between CS+ stimuli.

Analysis of the reaction times (RTs) on the same CS+ choices showed that the animals were significantly quicker on taking decisions when having to choose between the HV and the LV option than when having to choose between the HV and the CV option, in both the pre-scanning (t_3 = -5.741, p = 0.01) and scanning (t_3= -4.320, p = 0.023) sessions (Fig. 3.2 – bottom panel).

As well as in their choices, we also found evidence for the "less is more" effect in the animals' RTs on the trials when there was only a single CS+ present on the screen (the only alternative choice was pressing the sensor under the blank side of the screen). A repeated measures ANOVA during the scanning sessions showed that there was a statistically significant effect of the expected reward type on RTs ($F_{1.710, 30.787}$ = 15.348, p < 0.0005). On these single-option trials there was a statistically significant difference in the median RT taken to respond to the single HV stimulus (920.5 ± 117.96ms) compared to the single CV stimulus (1003.87 ± 133.25ms) (t_3 = -

5.833, p = 0.01), to the single HV (920.5 ± 117.96ms) stimulus compared to the single LV stimulus (1121.5 ± 103.36ms) (t_3 = -5.653, p = 0.011), and to the single CV stimulus (1003.87 ± 133.25ms) compared to the single LV stimulus (1121.5 ± 103.36ms) (t_3 = 3.853, p = 0.031). There was also a statistically significant correlation between the HV and CV RTs across monkeys (r = 0.981, p = 0.019) indicating that the animals responded similarly in both types of trials. Nevertheless, each option type was associated with approximately the same level of motivation throughout the testing session; there were only small average changes in RT during testing sessions. A repeated measures ANOVA with a Huynh-Feldt correction in the median RTs across the four animals between the beginning, middle and end of the task did not show any differences in the way in which animals responded to the HV ($F_{1.154, 3.461}$ = 0.474, p = 0.561), the CV ($F_{2, 6}$ = 0.972, p = 0.431), or the LV option ($F_{1.002, 3.007}$ = 0.998, p = 0.391) over the course of the testing session. See also Figs. 3.3 and 3.4.

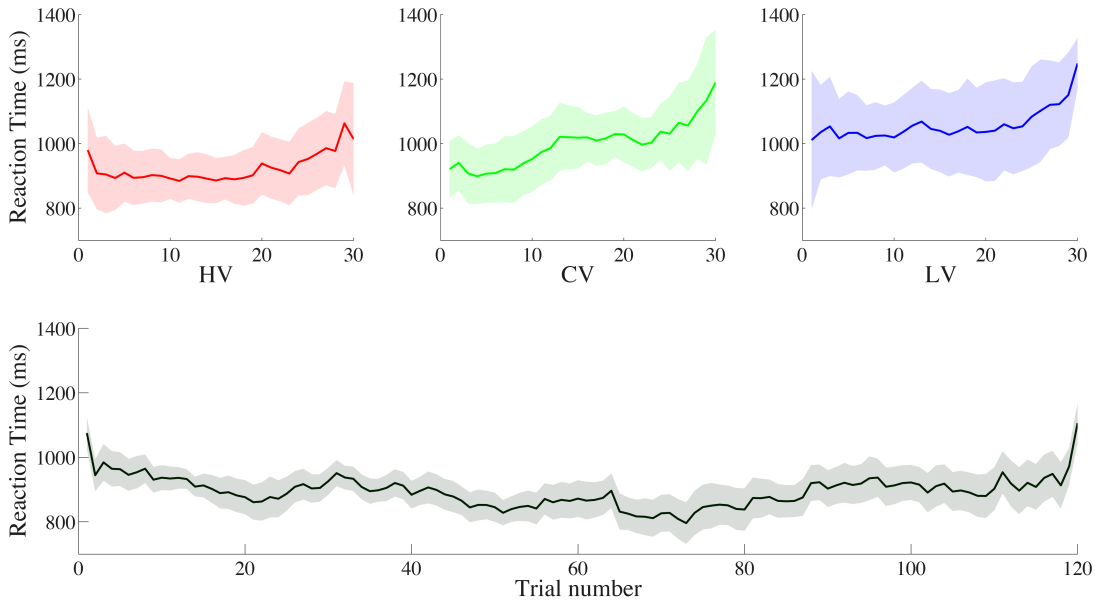


Figure 3.3: Median Reaction Times
Top row: Median RT in HV (red), CV (green) and LV (blue) one-option trials respectively.
Bottom row: Median reaction time across a session.
(Moving window =10 trials)

Even though both choice patterns and RTs indicated that monkeys preferred the HV option to the CV option there was still evidence that all options had positive values. The largest part of each session included choices between CS+ and forgoing the trial by pressing the sensor next to the blank side of the screen (75% of the trials). During those trials, an animal could choose to reject a potential reward and proceed to the next trial by touching the sensor associated with the blank side of the screen. For the choices between CS+ (HV, CV and LV) and the blank side of the screen the HV option was chosen on average 90.26 ± 13.7% of occasions, the CV option was chosen on 87.01 ± 14.07% of trials, and the LV option was chosen on 77.32 ± 12.66% of the trials. A one-sample t-test against 50% showed that all stimuli were chosen significantly more times than they were rejected ($p < 0.0005$) suggesting that all options, HV, CV, and LV had a positive value because they were taken in preference to moving onto the next trial.

The different choices that the animals had to make were further divided into two sub-categories: Easy Decisions (EDs) and Hard Decisions (HDs). EDs included decisions between options that were associated with a big value difference (mean parametric value difference = 0.69). Such decisions included the choices: HV versus the corresponding CS- stimulus that had been presented with the HV CS+ during learning, CV versus its corresponding CS- stimulus, the HV option versus the blank side of the screen, and the CV option versus the blank side of the screen. HDs included decisions between options that had a relatively small value difference (mean parametric value difference = 0.18). Such decisions included choices between: LV versus the corresponding CS- stimulus; LV versus the corresponding blank side of the screen.

More specifically, in the case of the one-option choices between a CS+ (HV, CV and LV) versus the blank side of the screen there was a statistically significant correlation, across animals, between the frequency with which HV and CV options were chosen ($r = 0.946$, $p = 0.046$). In other words, the animals tended to make their choices in a similar manner when choosing between either the HV stimulus and the blank screen or between the CV stimulus and the blank screen but not when they were making choices between the LV stimulus and the blank side of the screen.

Other aspects of choice behaviour also suggested similarities between the way that monkeys evaluated and made choices involving the HV and CV options. For the choices between each CS+ option versus CS-, a paired-samples t-test showed that there was a statistically significant difference in the way that the animals responded when choosing between the HV option versus the corresponding CS- and when choosing between the LV option versus the corresponding CS- ($t_3 = 5.547$, $p = 0.012$) and a difference when choosing between the CV option versus the corresponding CS- and the LV option versus the corresponding CS- ($t_3 = -3.800$, $p = 0.032$) but not when choosing between the HV option versus the corresponding CS-and the CV option versus the corresponding CS- ($t_3 = 1$, $p = 0.391$) (Fig. 3.4 – upper panel).

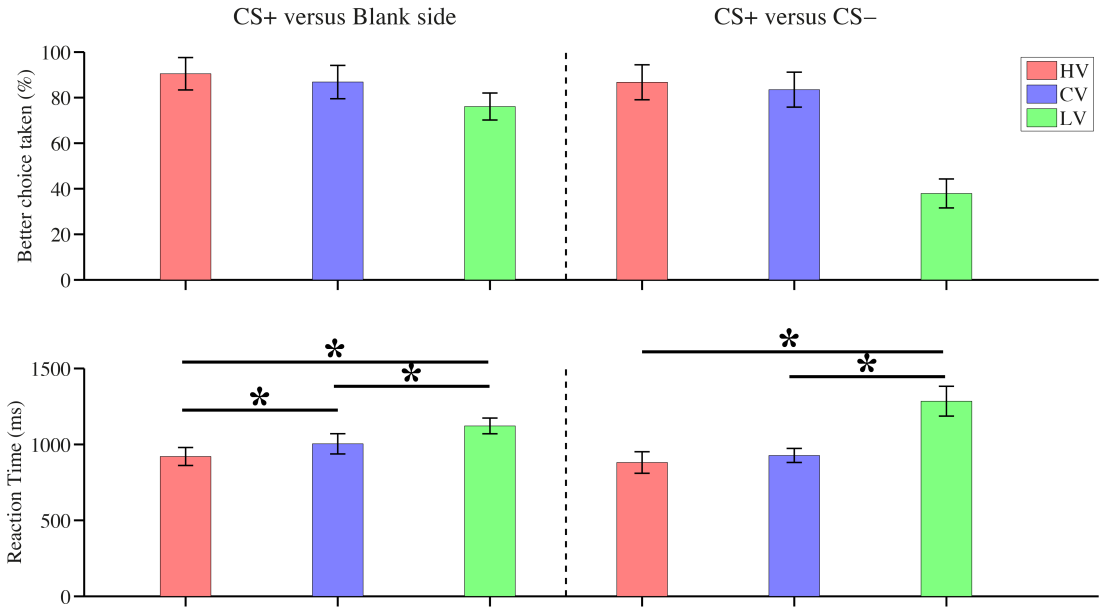


Figure 3.4: Choices and reaction times for CS+s versus blank side or CS- stimuli
Upper panel: Choice behaviour when choosing between a CS+ stimuli (left hand side) and either the blank side of the screen or a CS- stimulus (right hand side).
Bottom panel: RTs on trials on which only a single CS+ was presented either paired with a blank screen or with a CS-.

Also (Fig. 3.4 – bottom panel), there was a statistically significant difference in the RTs associated with decisions between the HV option versus the corresponding CS- option and decisions between the LV option versus the corresponding CS- option (t_3 = -3.998, p = 0.028) and decisions between the CV option versus the corresponding CS- option and decisions between the LV option versus the corresponding CS- option (t_3 = 3.866, p = 0.031).

3.3.2 Neuroimaging Results

3.3.2.1 Region of Interest

A key question is whether there is activity in the vicinity of macaque vmPFC/mOFC area in which a lesion was made in Chapter 2 when decisions involve the HV option as opposed to the CV option. Moreover, we wanted to investigate how

the same brain region encodes different aspects of choice such as the value of the options offered on a given trial and the difficulty of making different types of decisions.

In our first attempt to identify decision-related neural activity, we sought areas throughout the brain in which there was activity at the cue onset time for all decisions. The strongest signal over the whole brain was observed in the vmPFC/mOFC (GLM-2; CARET macaque F99 coordinates (McLaren et al., 2008; Saleem & Logothetis, 2006): x = 3, y = 21, z = 3; peak Z = 3.99; Fig. 3.5a, b; Table 3.3). Note that the decision-related activity is positive rather than negative as is the case in many human brain-imaging experiments. The peak coordinates of this signal were used for most of the ROI analyses described below. In some of the subsequent analyses a two-voxel radius spherical mask centered at this location is referred to as ROI-1.

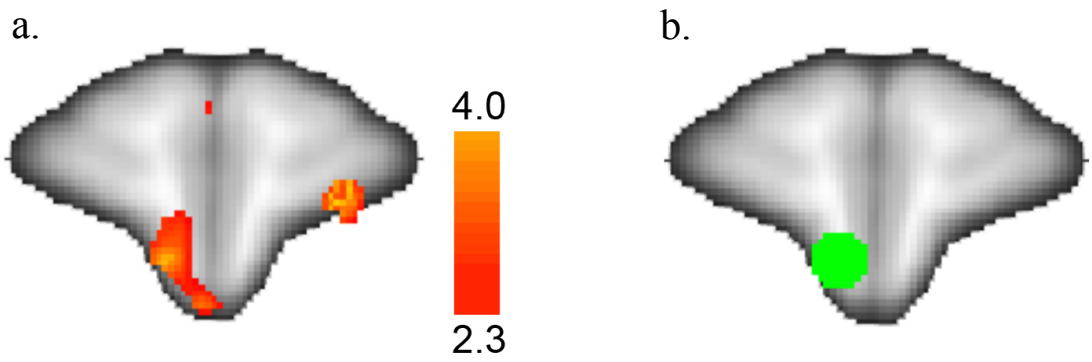


Figure 3.5: VmPFC/mOFC activity at cue onset time
(a) VmPFC/mOFC decision-related activity, once stimuli were presented on the screen. **(b)** Mask from which the signal was extracted for timecourse analyses (GLM-2).

The brain activity observed in these two clusters of the vmPFC/mOFC is located in areas 14 and 11 (Carmichael & Price, 1994; Mackey & Petrides, 2010; Petrides & Pandya, 1994; Pandya & Yeterian, 1998) and it overlaps with the site of the brain lesions of the vmPFC/mOFC in the animals described in Chapter 2.

Event	Brain Regions	MNI coordinates			z-score
Cue Onset		**x**	**y**	**z**	
	vmPFC/mOFC	3.5	21.5	3	3.99
	cOFC	7	19	7.5	5.21
	lOFC	19	8	-4	4.95
	Precentral gyrus	26	1.5	1.5	5.29

Table 3.3: Decision-related activity
Activity that exceeded a whole brain cluster-corrected significance level in four subdivisions of frontal cortex (GLM-2).

3.3.2.2 'Less is more' neural pattern

Next, we investigated the neural origins of the 'less is more' effect apparent in the monkeys' behaviour in both this chapter and Chapter 2, and which was particularly affected by vmPFC/mOFC lesions in Chapter 2. We examined whether there was greater vmPFC/mOFC activation when an animal had to choose between the HV and the CV options as opposed to when they chose between other option pairs (i.e. HV versus LV and CV versus LV). An analysis was therefore performed, within the area that had been lesioned in Chapter 2 comparing activity recorded during decisions between HV and CV options as opposed to decisions between either HV and LV options or decisions between CV and LV options.

Strong vmPFC/mOFC activity was observed when the monkeys had to choose between the CV and the HV options (GLM-1; Fig. 3.6a). This activity was significantly greater than that observed when the monkeys had to choose between the other CS+ pairs (HV versus LV and CV versus LV). This contrast (Critical comparison) was associated with significantly greater activity in vmPFC/mOFC (Fig. 3.6b, c).

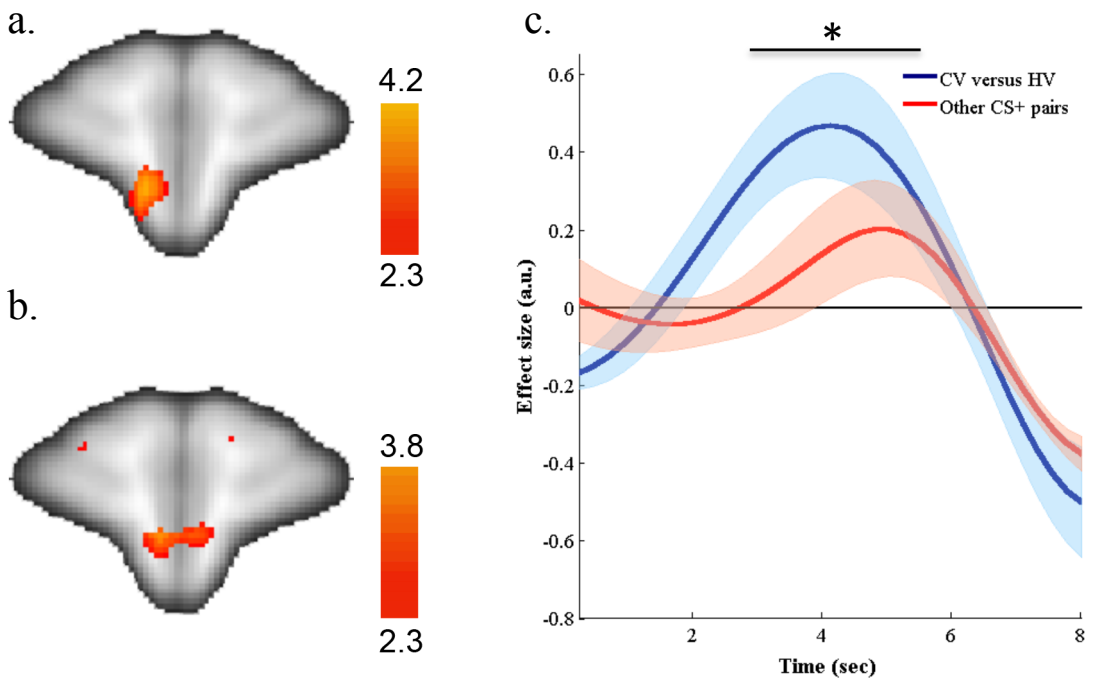


Figure 3.6: Critical comparison
(a) VmPFC/mOFC activity at cue onset time when an animal is presented with a choice between the HV and CV options. **(b)** Activity difference at cue onset time on trials on which macaques chose between HV and CV options as opposed to other CS+ pairs i.e. between the HV and LV options or between the CV and LV options. **(c)** Timecourse of activity associated with HV versus CV decisions (blue) and the other CS+ pairs (red) (GLM-1).

Further, a timecourse analysis from ROI-1 (using appropriate leave-one-out procedures) confirmed a statistically significant difference in vmPFC/mOFC activity associated with the types of decisions (HV versus CV decisions in comparison to decisions between other CS+ pairs: $t_3 = 3.5978$, $p = 0.037$).

3.3.2.3 Difficulty encoding

Because decisions between the LV and either the CS- option or the Blank side of the screen, were less consistent than other decisions they were arguably more difficult than other decisions. To further investigate the possibility that vmPFC/mOFC might reflect the difficulty of the decision being taken, a contrast was used to identify

activity associated with hard decisions (HDs) as opposed to easy decisions (EDs; HDs – EDs; GLM-1; Fig. 3.7).

There was a significant increase in activity in several frontal lobe areas, including in vmPFC/mOFC (x = -3, y = 18, z = 3), when monkeys made HDs as opposed to EDs (Fig. 3.7a). Activity time-courses from this region are also shown for illustration (Fig. 3.7b).

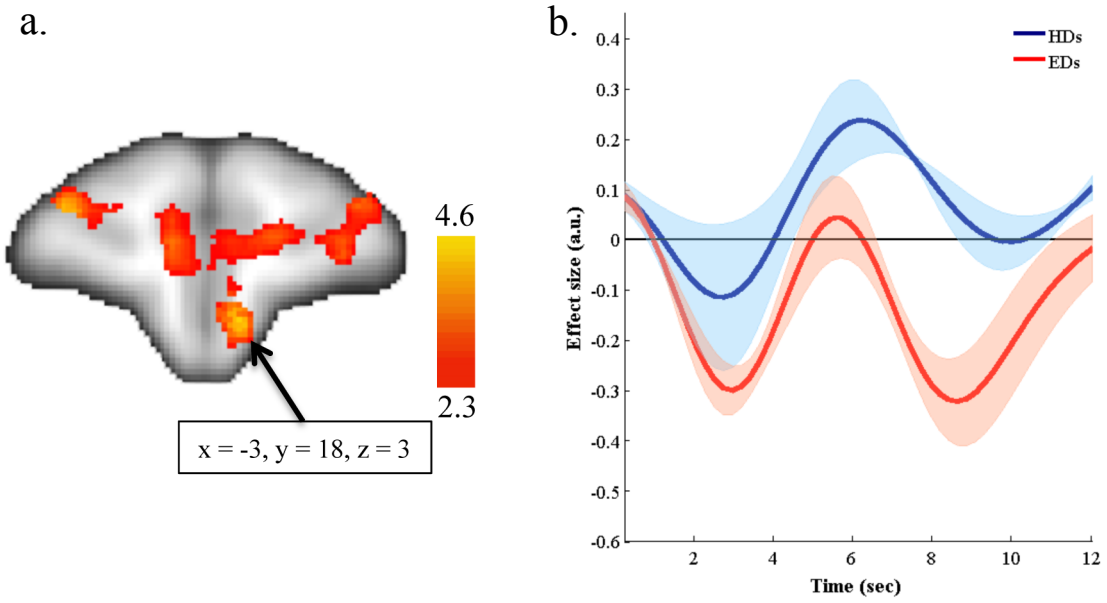


Figure 3.7: Decision difficulty
HDs: Decisions of higher difficulty between the LV option and either the Blank side of the screen or the CS- option. These decisions have a lower parametric value difference (mean parametric value difference = 0.18).
EDs: Decisions of lower difficulty between the HV/CV and either the Blank side of the screen or the CS- option. These decisions have a higher parametric value difference (mean parametric value difference = 0.69).
(a) HDs minus EDs. Activity increases associated with HDs as opposed to EDs were prominent in vmPFC/mOFC. **(b)** Time-courses of activity associated with HDs and EDs in the vmPFC/mOFC are shown for illustration (GLM-1).

3.3.2.4 Value encoding

As already noted, human vmPFC/mOFC activity decreases at the time of decision-making. Likewise, human vmPFC/mOFC activity also decreases with decision-difficulty. By contrast in monkeys, the preceding analyses have shown a

positive relationship between vmPFC /mOFC activity and decision-making that activity increases with decision difficulty. Given the complete reversal of vmPFC/mOFC activity and task relationships is there any evidence of consistent activity patterns across the two species?

Human vmPFC activity has also been related to the process of value representation and comparison that guides decision-making (Boorman et al., 2009, Fitzgerald et al., 2009; Basten et al., 2010; Rushworth et al., 2011; 2012; Hunt et al., 2012; Chau et al., 2014). For example, human vmPFC activity initially increases with the sum of the values of the options that are being considered in a decision and then subsequently with the difference in value between the options as a function of which is being taken (Hunt et al., 2012).

In the next analysis was looked for evidence that macaque vmPFC/mOFC BOLD activity reflected the sum of the values of the choices being considered. We therefore examined the brain activity associated with decisions involving the three different CS+ options and found a suggestion it might reflect value information (GLM-1). For example, decisions between the HV and the CV options were associated with higher activity than decisions between the HV and LV options or the CV and LV options. However, decisions between HV and LV options were only associated with a non-significant trend for higher activity than decisions CV and LV options (Fig. 3.8).



Figure 3.8: VmPFC/mOFC value transitivity
Decision-related activity time-locked at the presentation of the stimuli. **(a)** HV versus CV option minus CV versus LV option. **(b)** HV versus CV option minus HV versus LV option. **(c)** HV versus LV option minus CV versus LV option (GLM-1).

To test this possibility further, we ran a further, separate whole brain analysis that included parametric value regressors for the chosen and the unchosen options respectively (GLM-2). Two regressors corresponding to either the sum of the values of the two options or the difference between the values of the chosen and the unchosen options (GLM-2) during each trial of the task were constructed (Hunt et al., 2012). These regressors included only trials on which a CS+ was chosen from the animals. Trials, in which either a CS- or the Blank side of the screen was chosen, were excluded from further analysis and are discussed more below. Fig. 3.9 illustrates the strong negative relationship between vmPFC/mOFC activity and the differences in value between the chosen and the unchosen options. Thus, as in humans, there is a evidence that macaque vmPFC/mOFC BOLD reflects the value of the choices being considered during a decision but once again it appears as if there is an important difference in the way in which value-related activity manifests in vmPFC/mOFC in the two species; in humans vmPFC/mOFC activity is positively correlated with the difference in value between the choice taken and the choice rejected.

To clarify the nature of macaque vmPFC/mOFC value-related activity (ROI shown in Fig. 3.9b), we examined the time-course of activity recorded when HV choices were taken as a function of the value of the alternative option available during

a given decision (Fig. 3.9c). Now it becomes clear that activity in macaque vmPFC/mOFC activity is, just as in human vmPFC/mOFC, positively related to the value of the choice that is taken. However, as the value comparison becomes more difficult because the other option's value is higher, the vmPFC/mOFC activity is stronger and more protracted. For example, the green time-course in Fig. 3.9c shows that there was strong activity when HV was chosen in preference to the CV option that has the most closely similar value. When vmPFC/mOFC activity was plotted when the same HV choice was made in preference to a low value option (CS- or blank screen), there was less activity (blue time course in Fig. 3.9c). Further analysis, using appropriate leave one out procedures for both localizing effects and determining their peak in time revealed vmPFC activity increased as the value of the chosen option decreased (compare red and blue lines in panels moving left to right; Fig. 3.9c, d; two by two factorial ANOVA; chosen value: $F_{1,3} = 40.980$, $p = 0.008$; unchosen value: $F_{1,3} = 11.562$, $p = 0.042$) and increased as the value of the unchosen option increased (compare red, green, and blue lines; fig.3.9c, d).

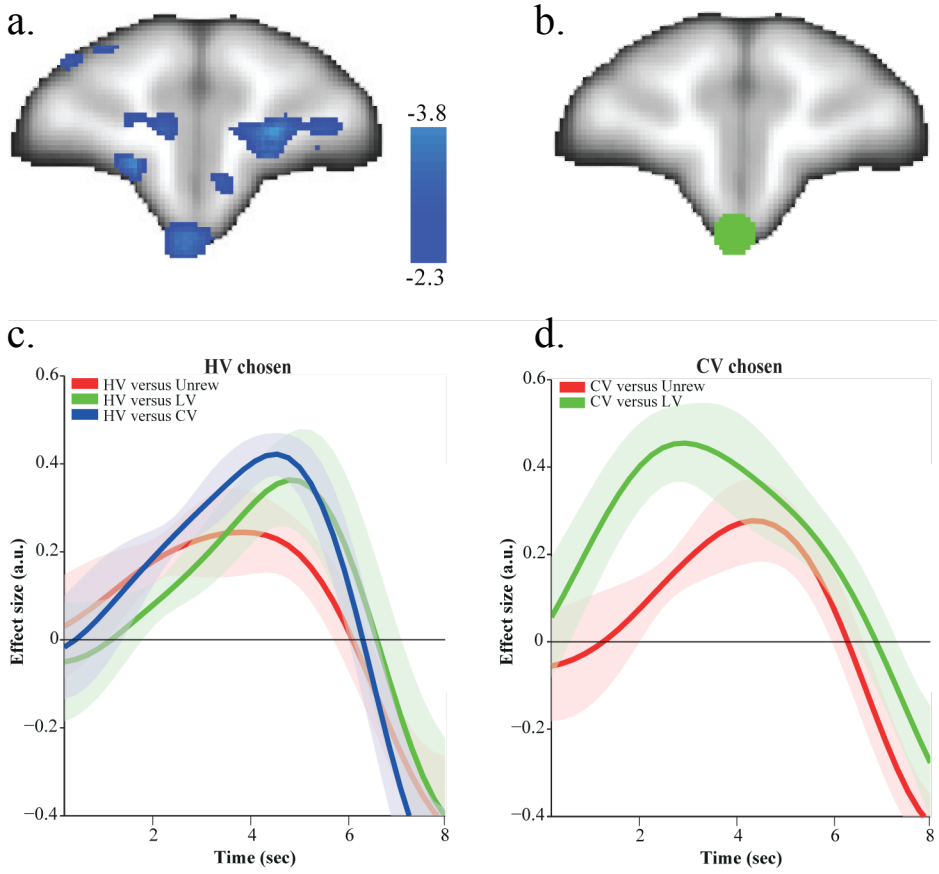


Figure 3.9: Chosen minus Unchosen stimulus value
(a) Activity negatively correlated with the difference the value of the chosen option minus the value of the unchosen option (x = 1, y = 17, z = -2). **(b)** Mask from which the signal was extracted for timecourse analyses. **(c)** HV option versus CV, LV and unvalued options (CS- or blank side of the screen). **(d)** CV option versus LV and Blank side of the screen (GLM-2).

So far our results suggest macaque vmPFC/mOFC reflects aspects of value comparison process, which is consistent with neurophysiological recordings made in adjacent regions (Padoa-Schioppa, 2011; Cai & Padoa-Schioppa, 2014; Strait et al., 2014). However, an alternative interpretation of neural activity in rodent OFC is that it reflects not so much the value of an option but simply the fact that a CS is associated with a very specific outcome. The final analysis tested whether the same was true of macaque vmPFC/mOFC BOLD. The analysis exploited the fact that in GLM-2 we had divided trials into those on which decisions were guided by CS+s, stimuli associated with a specific valued outcome, and decisions on which the choice was not guided by a specific CS+, i.e. trials on which animals chose the CS- or the

blank side of the screen. When we compared these two types of trials there was a significant increase in activity in vmPFC/mOFC (Fig. 3.10a). The activation was just lateral to that related to the difference in value between the two choices close to the medial orbital sulcus (Fig. 3.9a). We examined in the region highlighted in Fig. 3.10a (ROI shown in Fig. 3.10b) when the CS+ that was chosen was the HV option, CV option, or LV option (timecourse analysis in Fig. 3.10c). A comparison of activity time courses revealed that HV ($t_3 = 5.7329$, $p = 0.01$), CV ($t_3 = 6.5149$, $p = 0.007$), and LV ($t_3 = 33.8195$, $p < 0.001$) choices were all associated with statistically significant activity than CS- and blank side ($t_3 = 0.5872$, $p = 0.19$) choices but there was no difference in the activity associated with HV, CV, and LV choices ($p > 0.05$).

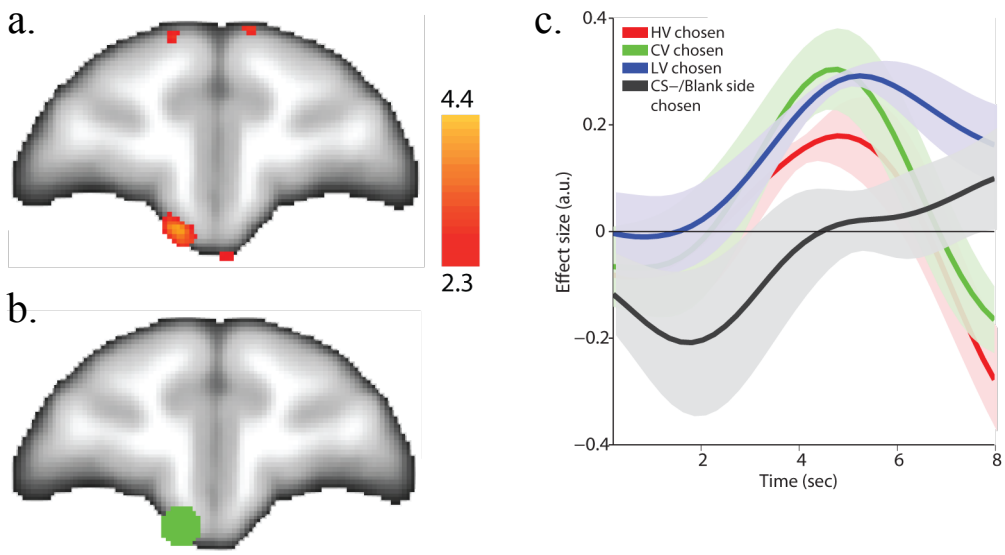


Figure 3.10: VmPFC/mOFC stimulus value information
(a) Brain activity for CS+ stimuli (i.e., HV, CV and LV) chosen minus brain activity for CS- stimuli/Blank side of the screen chosen, time-locked at stimuli presentation. **(b)** Mask from which the signal was extracted for timecourse analyses. **(c)** Stimulus-elicited activity, split per chosen option, as extracted from the peak of the cluster shown in (a) and (b) (x = 4, y = 15, z = -1) (GLM-2).

3.3.2.5 Other findings

3.3.2.5.1 Reward-related vmPFC/mOFC activity

A contrast of activity at the outcome time when reward was delivered (i.e. when an animal chooses a CS+ stimulus and approximately 4s from choice) as opposed to omitted (i.e. when an animal chooses a CS- stimulus or the blank side of the screen and again approximately 4s from choice), revealed a very strong signal, on the medial surface near the rostral sulcus (GLM-1; Fig. 3.11) as well as other areas (Table 3.4).



Figure 3.11: VmPFC/mOFC activity at reward delivery
Reward-related activity time-locked at the onset of reward delivery and for choosing a CS+ stimulus (a reward is delivered approximately 4s after an animal makes a choice), contrasted with brain activity when a CS- or the Blank side of the screen was chosen (again approximately 4s from choice), i.e., CS+ chosen minus CS-/Blank side of the screen chosen (x = -2, y = 21, z = 8) (GLM-2).

Event	Brain Regions	MNI coordinates			z-score
Outcome		**x**	**y**	**z**	
	lOFC	-18	15.5	2.5	5
	Inferior frontal gyrus/ Precentral gyrus	-19.5	7	8.5	7.94
	Precentral gyrus	23.5	7	2	9.12
	Putamen	-13	4.5	2	6.62
	Putamen	14	2	-1.5	8.62
	Thalamus	-1.5	-7	-1.5	6.37
	Superior temporal gyrus	23	7	-8	4.83
	Superior frontal gyrus	2.5	3.5	19.5	6.54
	Superior temporal gyrus	-18	7	-10	4.6
	Superior temporal gyrus	18	-3.5	-9	7.78
	Middle temporal gyrus	-24	-12.5	-7	4.4
	Superior frontal gyrus	-5.5	6.5	23.5	5.76
	Occipital gyrus	-8.5	-47.5	6	6
	Fusiform gyrus	18	-17.5	-12.5	4.79
	Anterior cingulate cortex	-4.5	-6	16	5.26
	Amygdala	12.5	-3.5	-12	8.8
	Hippocampus	13	-10	11.5	5.61
	Caudate nucleus	-2.5	7	7	4.5

Table 3.4: Reward-related activity
Brain regions that the activity was modulated by the occurrence of the reward delivery event across all task trials i.e., CS+ chosen minus CS-/Blank side of the screen chosen. Multiple regions across the whole brain exceeded a cluster-corrected significance level (GLM-2).

3.3.2.5.2 Nucleus Accumbens

Investigation of possible subcortical striatal brain activity at the cue onset time revealed, among else, a strong nucleus accumbens (NAc) activity (GLM-1; Fig. 3.12).

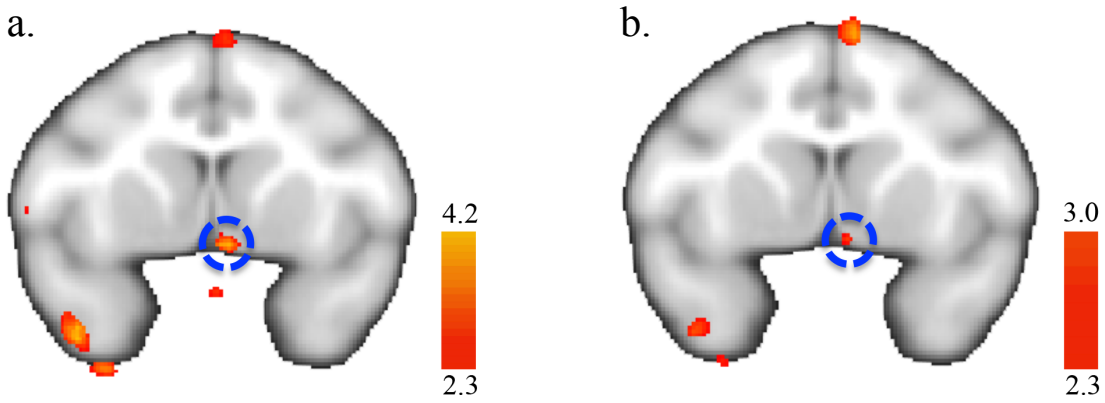


Figure 3.12: NAc activity at cue onset
Decision-related activity time-locked at the presentation of the stimuli. **(a)** Activity during one-option trials, i.e., trials in which a single CS+ is presented against the blank screen. **(b)** Activity during two-option trials, i.e., trials in which two CSs are presented on the screen (blue dot circle, x = -2, y = 4, z = -3) (GLM-1).

Although, no clear statistical significant difference between the results was found there was a trend. In Chapters 4 and 5 of this thesis, I will present a more refined and spatially specific analysis of dopamine related activity in the nucleus accumbens using fast-scan cyclic voltammetry.

3.4 Discussion

In this experiment we found further evidence that macaque monkeys exhibit the 'less is more' effect, both in the pattern of choices made when decisions were taken between the stimuli associated with HV and CV outcomes; as in Experiment 1, there was a strong preference for the HV option over the CV option and this preference remained stable in all sessions. Further evidence of preference for the HV over the CV option was apparent in the RTs when single options were presented in isolation. Related results have been reported with human subjects (e.g. Hsee et al., 1998). Both RT and choice data suggested that the macaques preferred the HV option

to the CV option even though the CV option comprised both juice outcomes and therefore might have been expected to have a value that corresponded to the sum of the values of both the HV and LV options. Instead, however, it appeared that the CV option's value was biased away from the sum of HV and LV options' value towards their mean. Moreover, once again, it was clear that despite this choice pattern, the LV option had a positive value for the monkeys; the monkeys chose to receive the different juice types and chose not to reject any of them by pressing the sensor under the blank side of the screen in the large majority of the one-option trials, indicating that all reward types had positive values.

It is possible that an interaction between the tastes associated with the two juices that comprised the CV option could have been unpleasant for the monkeys. This, however, seems unlikely, for a number of reasons. First, the macaques drank both the HV and LV juice when these were delivered freely in their home cage. Second, they drank both juice types when the CV option was chosen. Third, the CV option was preferred to the LV option and even the LV option was preferred to simply forgoing any juice by pressing the sensor next to the blank side of the screen. Fourth, we found that the effect is generalizable because it happened with both the food rewards used in Chapter 2 and the liquid rewards used here.

On first impression it seems as if the choices of control macaques on HV versus CV decisions are irrational. However, humans sometimes exhibit a similar tendency to average the value of groups of items instead of summing them (Hsee & Kunreuther, 2000; List et al., 2002). For example, collectors were inclined to pay more for set of high value baseball cards than they were for an identical set of high value cards together with some additional low value cards (List, 2000). Moreover, such value estimates may be more rational if decision-making is seen within the

broader context of the foraging decisions macaques have evolved to make (Pearson et al., 2014), when the added value of the CV option over the HV option may be outweighed by the handling costs of its LV component and the opportunity cost incurred by failing to move on to other opportunities that might be present on subsequent decisions. However, in the current experiment, the longer delay imposed after the end of trials on which a CS- or blank screen was chosen, and no reward delivered, meant that the monkeys did not move on to the next trial any more quickly than on trials on which CS+s were chosen and rewards were delivered. Monkeys, however, are not always sensitive to the precise delays imposed by experimenters between trials (Blanchard et al., 2013).

Activity in human vmPFC covaries with the reward value of objects in the environment and of potential choices (Abitbol et al., 2015; Howard et al., 2015; Lebreton et al., 2015; Li et al., 2016; Noonan et al., 2011). Analyses of the time course of such activity and of the relation between inter-subject variability in signal size and variation in decision-making behaviour suggest it may guide decision-making (Boorman et al., 2009; Chau et al., 2014; Hunt et al., 2012; Jocham et al., 2014; Jocham et al., 2012; Kolling et al., 2012). At the same time, however, when lesions are made in vmPFC and adjacent orbitofrontal cortex (OFC) in macaques they do not impair all value-guided decisions (Rudebeck and Murray, 2011) casting doubt on a general role for vmPFC in reward-guided decision-making. For example, even when lesions are restricted to central OFC areas 11 and 13 they impair performance in reward devaluation tasks in which animals have to update their expectations about the value of a specific type of reward that a choice will deliver on the basis of their current levels of satiety (Rudebeck and Murray, 2011) but even large lesions of OFC

do not disrupt decisions in other simpler stimulus-reward association tasks (Rudebeck et al., 2013).

The discrepancy between the results of human neuroimaging studies and macaque lesion studies might be explained in a number of ways. First, there is a surprising degree of uncertainty about the anatomical identity of the vmPFC region in which reward-guided decision-making activity arises in humans and its relation to vmPFC and OFC areas in other primates such as macaques (Wallis, 2012). It is possible that macaque lesion studies have not investigated the homologue of the critical region identified in human neuroimaging studies. Second, the behavioural tasks used in the two species may not assess comparable cognitive processes. Several human neuroimaging studies that have reported value-guided decision-making activity in vmPFC employed multi-faceted choice options. For example, several studies have employed choices varying both in terms of their associated reward probability and reward magnitude but vmPFC activity is only prominent when decision-making depends on consideration of both facets (Hunt et al., 2012). VmPFC may be critical when decision-making depends on the complementary abilities to either integrate information about multiple aspects of a choice outcome into a unified representation (Barron et al., 2013) or to focus on the most relevant aspect (Hunt et al., 2014).

When we used the same technique, fMRI, as had been used in human experiments we were also able to identify activity in macaque vmPFC/mOFC that was time-locked to decision events. There was a large area of increased activity extending from the medial orbital sulcus and across the gyrus rectus within the vmPFC.

The lesion experiment in Chapter 2 had suggested a special role for vmPFC/mOFC in multiple component decisions such as those made between HV and

CV. The lesions we made may, however, have damaged white matter connections subjacent to the cortex (Rudebeck et al., 2013) and so a further analysis also examined whether we could confirm that vmPFC/mOFC was particularly concerned with HV-CV decisions as opposed to decisions made between HV- and LV-associated stimuli and CV- and LV-associated stimuli. There was significantly more activity in the former case. Together, the behavioural and lesion results of Chapter 2 and the neuroimaging results of this chapter suggest that the mOFC/vmPFC plays an important role in representing the values of multi-faceted options.

It is important to note that there are important differences between decision-related frontal cortical activity during value-guided decision-making in humans and macaques. First, such activity in humans is associated with brain areas on the medial wall (Abitbol et al., 2015; Boorman et al., 2009; Chau et al., 2014; Howard et al., 2015; Hunt et al., 2012; Jocham et al., 2014; Jocham et al., 2012; Kolling et al., 2012; Lebreton et al., 2015; Li et al., 2016; Noonan et al., 2011) rather than the more lateral position within vmPFC found in the macaques in this study. However, that similar patterns of activity were found in these human and macaque areas is consistent with Mackey's and Petrides' claim that both regions, in humans and macaques, are parts of area 14, even if they are perhaps distinct subdivisions of area 14. (Mackey & Petrides, 2010, 2014).

Second, while decision-making was associated with increased BOLD activity in macaque vmPFC/mOFC in the current experiment, it has been associated with BOLD decrements in humans (upon which positive value effects are superimposed). There has been considerable interest in the precise interpretation of both decision event-related and difficulty-driven decrements and increments in BOLD in vmPFC in comparison to other brain areas such as anterior cingulate cortex (ACC) in human

neuroimaging studies (Abitbol et al., 2015; Hare et al., 2011; Rushworth et al., 2012). For example, the fact that human ACC activity is positively correlated with decision difficulty has been interpreted as indicating that ACC's key function is related to difficulty (see review by Kolling et al., 2016). Therefore, in addition to examining the main effect of decision-making on vmPFC/mOFC activity we also examined whether macaque vmPFC/mOFC activity changed positively or negatively as a function of decision difficulty. Decisions were divided into two types, Hard Decisions and Easy Decisions on the basis of the consistency of macaques' choices (consistent choices on a given decision suggest that there is a clear value difference between the options). Unlike in human subjects, macaque vmPFC/mOFC activity increased with decision difficulty. One possibility is that regardless of whether BOLD activity is positively or negatively correlated with decision difficulty it may still reflect the operation of a comparator processes instantiated by similar neural mechanisms. Whether activity is positively or negatively related to difficulty might simply depend on whether high firing attractor states in the networks are maintained only until the decision is made or if they are still maintained afterwards; in the latter case difficulty-related activity that is integrated across time (by the sluggish BOLD signal) may be negatively related to difficulty (Kolling et al., 2016).

If macaque vmPFC/mOFC has a similar role in value-guided decision-making to human vmPFC medial wall areas then, despite the fact that activity in macaques and humans is positively and negatively correlated with decision difficulty respectively, it should be possible to identify more direct evidence of a value comparison processes (Boorman et al., 2009; Chau et al., 2014; Hunt et al., 2012; Jocham et al., 2014; Jocham et al., 2012; Kolling et al., 2012). We examined whether activity in the same macaque vmPFC area covaried with the difference between the

value of the choice taken as opposed to the value of the choice forgone. Interestingly a strong deactivation was observed demonstrating the opposite pattern from what happens to the human vmPFC (Fig. 3.10a, b). A closer look at the time-courses, extracted from a relatively medial part of the vmPFC/mOFC, however, revealed a region in which activity, at the time of the decision, reflected the overall value of the components of each choice (Fig. 3.9c, d). Moreover, activity was higher when the monkey had to make a choice between two CSs+ than between a CS+ and non-valued stimulus (i.e. a CS- or the Blank side of the screen). In a more lateral part of vmPFC/mOFC activity was closely related to the values of the two choices available and simply whether the monkeys' behaviour was guided by a CS+ with a specific reward association as opposed to choosing to forgo the trial to move on to the next by pressing the blank screen or choosing the CS- (Fig. 3.10c, d). Activity in this part of vmPFC/mOFC is consistent with recent research suggesting that some neurons in the rodent OFC have activity that reflects specific learned associations between CSs and particular reward outcome identities rather than parametrically scaling with the value of the rewards (McDannald et al., 2014, Lopatina et al., 2015).

Value-related activity has been recorded in many brain regions but the combination of fMRI and lesion results presented in this chapter and in Chapter 2 suggest that, amongst other functions, area 14 in vmPFC/mOFC may have a particular role in construction of value estimates to guide decision-making when choice outcomes have multiple component parts and it is necessary to integrate information about these parts (Barron et al., 2013) or focus selectively on one particularly part (Fellows, 2006; Hunt et al., 2014). This may be true even when it is not necessary for values to be estimated "on the fly" from "knowledge of the causal structure of the environment" (Jones et al., 2012) (although this may be the case for adjacent central

OFC regions such as 11 and 13 in macaque (Rudebeck and Murray, 2011). However, the results suggest that the value estimates constructed by vmPFC/mOFC do not always result in optimal value maximizing behaviour (Camille et al., 2011).

Another interesting finding of this experiment was a NAc signal recorded at the time that the stimuli were presented on the screen. NAc is an important dopaminergic brain structure with a significant role in decision-making, motivation and learning (Flagel et al., 2011; Parkinson et al., 2002; Kelley, 2002) and the activity of the dopamine neurons within this structure reflects the value of expected rewards (Gan et al., 2010; Hollon et al., 2014; Syed et al., 2015). The activity observed in NAc could therefore reflect the value of the different reinforcers available. Chapters 4 and 5 present a finer grained analysis of dopaminergic activity in the same NAc brain region during an appetitive decision-making task. However, a different technique (fast-scan cyclic voltammetry) and species (rats) will be used for this purpose.

4 The influence of reward identity on mesolimbic dopamine release

Summary

The phasic activity of midbrain dopamine neurons and dopamine release in regions such as the nucleus accumbens (NAc) signals predictions of future reward and discrepancies between such predictions and received rewards (Gan et al., 2010; Hart et al., 2014; Montague et al., 1996; Schultz et al., 1997; Syed et al., 2016). These reward prediction errors (RPEs) rely on stored values learned through experience, meaning that theoretically any changes in reward *identity* should not influence dopamine so long as changes in value are controlled for. Here, we investigated this issue by recording dopamine release while rats performed a two-reward choice task. We focused on a subset of trials where the delivered reward was either greater-than-expected or was switched to the alternative reinforcer of equivalent value. We observed that both types of surprise caused phasic increases in dopamine levels, though changes in reward identity only had this effect when expecting sucrose solution and obtaining food. Cue-elicited dopamine release was also greater for food- than sucrose-associated cues. Therefore, these putative identity prediction errors are likely reward prediction errors in a different frame of reference to subjective preference.

4.1 Introduction

The ability to successfully adapt to new situations relies heavily on our ability to learn from our previous experiences and effectively use that knowledge to guide our future behaviour and decision-making. Such prior knowledge enables us, among other things, to generate predictions about potential outcomes and adjust our actions accordingly, if needed. Of course, our predictions do not always lead to the expected outcomes. Such discrepancies in value are termed reward prediction errors (RPEs).

The activity of midbrain dopamine neurons has been shown to correlate with RPEs (Bayer & Glimcher, 2005; Cohen & Uchida, 2012; Montague et al., 1996; Schultz et al., 1997). These signals have been shown to be encoded on a common value scale, integrated across different reward attributes, which reflects individuals' subjective preference for particular outcomes rather than the objective properties of rewards (Lak et al., 2014; Stauffer et al., 2014). For instance, in the study by Lak et al., (2014) the activity of dopamine cells in the substantia nigra and the ventral tegmental area were recorded while monkeys were presented with different stimuli associated with rewards that varied over multiple dimensions, such as identity, value and probability. The monkeys were found to make their choices as if they were integrating across all the different attributes of reward onto a common value scale. Presentation of stimuli associated with rewards of equal value, but of different attributes, led to indistinguishable differences in dopamine firing rates. Similarly, analyses at the time of the reward delivery showed that dopamine responses to rewards corresponded either to the amount of the reward being delivered or to the integrated value of the amount and the attributes of the reward.

One brain region that is tightly linked with dopaminergic activity and RPEs is the nucleus accumbens (NAc), which is part of the ventral striatum. The NAc core is a

key structure in evaluating, modulating and updating incentive motivation for appetitive reinforcers (Flagel et al., 2011; Parkinson et al., 2002; Kelley, 2002) and transient dopamine release here follows a reward prediction error term in temporal difference learning models (Day et al., 2007; Hart et al., 2014). Cue-evoked dopamine in the NAc core was found to be positively correlated with cue-evoked food seeking behaviour (Ostlund et al., 2014) and has been shown to scale with predicted future rewards. Importantly, dopamine release also scales with discrepancies from expectation (Hart et al., 2014; Nasrallah et al., 2011; Klanker et al., 2015). Blocking dopamine transmission here disrupts the acquisition of learning about the incentive properties of cues and also the ability of cues to motivate responding (Flagel et al., 2011; Nicola, 2010).

According to TDRL, RPEs should depend on learned numeric values – good or bad - that are stored independent of the identity of the reward (Niv & Schoenbaum, 2008). Any change in reward identity separate to reward value, theoretically at least, should not therefore influence dopamine release. To investigate whether this is correct, we used fast-scan cyclic voltammetry to record dopamine release in the NAc core as animals made choices for two different reward types – food pellets or sucrose solution. The volume and concentration of the sucrose solution was adjusted such that it was on average of equivalent value to a food pellet. We focused on a subset of trials where animals received either received a larger quantity of the expected reward type or the standard amount of the alternative reward type. In a second analysis we investigated the patterns of cue-elicited dopamine release for these options.

4.2 Methods

17 male Sprague-Dawley rats weighing 350-400g were used in this experiment (Harlan Olac, Bicester, UK). This included 11 naïve rats aged 3-6 months at the start of testing as well as an additional 6 rats, aged 6 months, which had previously participated in a separate 2-option appetitive decision-making experiment. A total of 6 rats were excluded from the electrochemistry analysis; 3 for misplaced electrodes outside the nucleus accumbens core and 3 for broken/noisy electrodes, leaving a total of 11 rats that are included in this experiment. Animals were maintained on a 12-hr light/dark cycle, were initially group housed in cages of 3 but then were individually housed after surgery and during the testing period. All testing was carried out during the light phase. During the training and testing periods, access to food was restricted such that rats' weights were kept between 85-90% of their free-feeding body weight. Water was continuously available in the home cages. All procedures were in compliance with the United Kingdom Animals Scientific Procedures Act (1986).

4.2.1 Surgical Procedures

Animals were anaesthetised using isoflurane (4% induction and 1.5% for maintenance) and given buprenorphine (Vetergesic, 0.1 ml/kg) at the start of the surgical procedure. Body temperature was maintained at 37±0.5 °C with the use of a homeothermic heating blanket. Corneal dehydration was prevented with application of ophthalmic ointment (Lacri-Lube, Allergan, UK). After induction, the rat's head was shaved and secured in a stereotaxic frame. The head was cleaned with dilute hibiscrub, 70% alcohol and a local anaesthetic, bupivacaine, was applied to the area. The scull was then exposed and holes were drilled for an Ag/AgCl reference electrode

(AP: -3.7, ML: -.4), 4 anchoring screws (Precision Technology Supplied Ltd, UK) and a recording electrode inserted, custom-made carbon fibre microelectrodes were then lowered into the NAc core (AP: +1.4, ML: ±1.3, DV: 7.0). The carbon fibre microelectrodes and reference electrode were attached to a headstage connector, which was secured in place along with an anchoring headpost using dental cement (Kemdent, Swindon, UK). Following surgery, animals were administered additional buprenorphine (0.1 ml/kg) and meloxicam (Metacam, 0.2 ml/kg). Meloxicam was also administered for at least 3 days following surgery. Animals had on average three weeks of post-surgery recovery with food and water ad libitum, prior to food restriction and further behavioural training/recording.

4.2.2 Behavioural Paradigm

4.2.2.1 Apparatus

Testing was carried out in custom-designed operant chambers (30.5 x 24.1 x 29.2cm; Med Associates, VT, USA). Each chamber was housed within a sound-attenuating cabinet ventilated with a fan, which provided background noise of ~64dB. Each chamber contained two retractable levers, situated 9.5cm on either side of a reward magazine that contained a receptacle into which both 45mg standard grain-based pellets (Test Diet, distributed through Sandown Scientific, UK) and sucrose solution could be delivered. Above each there was a cue light. The magazine was fitted with an infrared beam that signaled when animals entered the receptacle. Each chamber was also fitted with a house-light.

4.2.2.2 Training

Rats were first given experience of each type of reward in the operant chamber by placing a number of pellets or amount of sucrose solution in the reward receptacle during two brief sessions separated by ~1 hour (reward order counterbalanced across animals). On the next day, the rats received a magazine training session with both reinforcers. In this session (duration ~30mins), pellets or sucrose solution were delivered on VI60 schedule. Subsequently, rats were taught to press levers to gain reward. Half of the rats were trained that the left lever would lead to sucrose solution and the right lever to pellets, whereas the other half were trained on the opposite configuration. These associations remained fixed throughout the data acquisition. Once the animals were making ~100 responses on both levers, they moved on to the main paradigm.

The main paradigm used during both behavioural and subsequent voltammetric recording sessions, consisted of individual trials where animals could press a lever to gain one of the two reward types (Fig. 4.1).

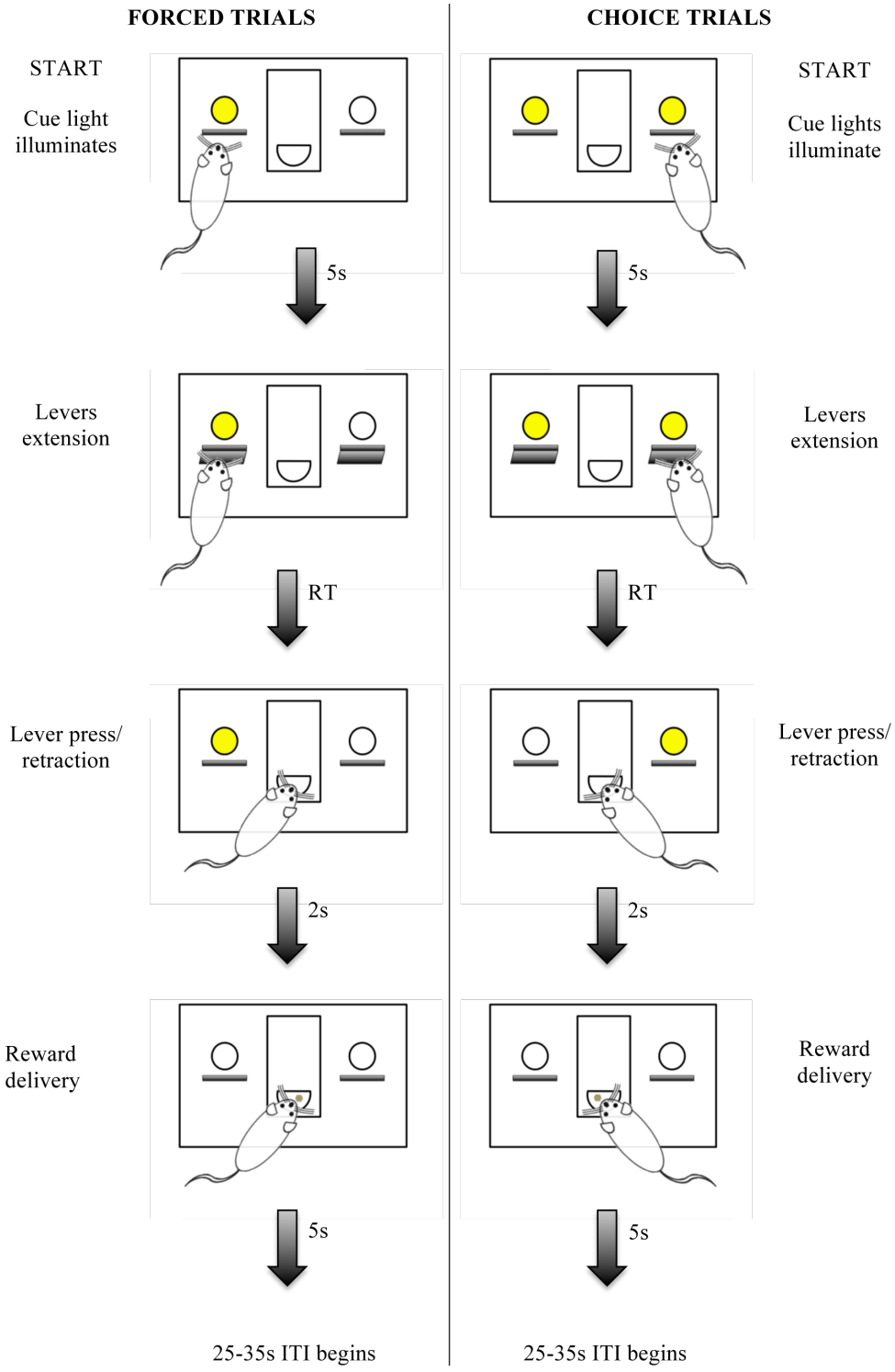


Figure 4.1: Task design
(a) Schematic of a typical forced trial (forced "left"). At trial onset, the left cue light illuminates and 5s later both left and right levers extend. A response on the left lever will lead to reward being delivered 2s after in the central food magazine; a response on the right lever will be an error and will not lead to reward. 80% of the trials were standard trials (the expected size and type of reward was delivered), 15% were SWITCH trials (the standard size of the *other* reward was delivered) and 5% were MORE trials (an increased size of the expected reward was delivered. **(b) Schematic of a typical choice trial.** At trial onset, both cue lights illuminate and 5s later both levers extend. A response to either lever will lead to reward being delivered 2s after in the central food magazine.

Sessions contained 'forced' trials where only one reward type was available, and 'free choice' trials where both reward types were available. At the start of a forced trial, the house light would turn on and one of the two cue lights would illuminate. After a delay of 5sec, both levers would extend. A single lever press on the option under the illuminated cue would cause both levers to retract and reward to be delivered to the receptacle 2s later. The cue light over the selected option remained illuminated during this delay to reduce any working memory load between choice and reward delivery. There was then an inter-trial interval (ITI) of 25 – 35s. A response on the lever under the non-illuminated cue light was counted as an error and did not lead to any reward delivery. If the rat failed to make a response within 10sec, this trial was counted as an omission. After either an error or omission, the levers would retract, cue and house lights turn off, and the ITI would immediately start. Choice trials were identical to forced trials except that both cue lights would illuminate and selection of either option within 10s would result in reward delivery 2s after this choice. The houselight was on throughout the session.

4.2.2.3 Reward value titration

To establish the association between each response option and a specific reward during titration sessions, the rats were given a session with only forced trials (40 to each lever in a pseudorandom order). They then had a session where forced trials were interspersed with equal numbers of choice trials, which could be used to determine each animal's preference for the reward types. Each session consisted of 80 trials made up of blocks of 4 forced trials (2 to the left, 2 to the right lever, pseudorandom order) followed by 4 choice trials. Based on pilot data collected using a different reward delivery system, rats were initially tested with 190µl of 30%

sucrose solution (Test A) and then subsequently with 95µl of 20% sucrose solution (Test B) across separate sessions. Our criteria for determining whether this was an appropriate volume and concentration of sucrose for the group of rats were (a) stable performance across two Pre-Test titration sessions (no significant change across the sessions) and (b) average number of food choices not being significantly different from 50%.

4.2.2.4 Reward identity task

We used fast-scan cyclic voltammetry to record dopamine release as animals performed a modified version of the 2-choice/2-reward decision-making task described above. All sessions consisted of 120 trials, consisting of 10 blocks of 12 trials (8 forced trials, 4 to each lever in a pseudorandom order, followed by 4 free choice trials). On 80% of trials, animals received the reinforcer associated with the selected option ("expected reward"). However, on 10% of the forced trials and 5% of the choice trials, the animals received the reward associated with the other lever ("SWITCH" trials). On another 5% of the forced trials the animals received four times more reward than expected though of the expected identity ("MORE" trials).

8 of the 11 animals that underwent FCV completed 4 separate recording sessions, consisting of 2 baseline sessions and 2 selective satiation sessions (3 animals completed only Baseline 1). The session order was always Baseline A - Selective Satiation A - Baseline B - Selective Satiation B, with the order of satiation (food or sucrose solution) counterbalanced across animals. In this chapter, only data from the two Baseline sessions will be presented; the selective satiation data will be presented in Chapter 5.

4.2.3 Fast-scan cyclic voltammetry

Fast-scan cyclic voltammetry (FSCV) recordings from chronically implanted carbon fibre electrodes were performed as described previously (Clark et al., 2010; Gan et al., 2010). Voltammetric scans were performed at a frequency of 10Hz throughout the session. Prior to a scan, the carbon fibre was held at a potential of -0.4V (vs AG/AgCl) and then, during the scan, ramped up to +1.3V and back to -0.4V at 400 V/s. The application of this waveform causes redox reactions in electrochemically active species, such as dopamine, at the surface of the carbon fibre that can be recorded as changes in current over time. Based on previously established criteria the recorded current in response to uncued pellet and sucrose delivery, obtained at the start and end of each recording session, was used to determine the chemical sensitivity of the recording electrode to dopamine on that particular session. An extracted cyclic voltammogram was linearly regressed against a dopamine standard, with $r^2 \geq .75$ set as the criterion based on the discriminability of dopamine from other common neurochemicals in a flow cell (Gan et al., 2010).

4.2.4 Data analysis

Each animal's preference for food over sucrose solution was calculated as the number of food choices / (number of food choices + number of sucrose choices). Subjective preference in each session was tested as a 2-tailed t-test against indifference (50%) and consistency across sessions was tested as a repeated measures ANOVA with session as a within-subjects factor. Response latencies were calculated as the time from lever extension to a response.

Voltammetric analysis was initially carried out using software written in LabVIEW (National Instruments). Data were low-pass filtered at 2kHz. To isolate

changes in dopamine concentration from other electrochemical signals, a principal component analysis was performed using a standard training set of stimulated dopamine release detected by chronically implanted electrodes, with dopamine treated as the first principal component among other unrelated electrochemical fluctuations such as changes in pH (Heien et al., 2004; Heien et al., 2005). The data were smoothed using a 0.5s moving window. Trials where the PCA failed to successfully extract dopamine current on >50% of data points in a trial were excluded. Once dopamine-related current changes were extracted all further analysis was undertaken using Matlab® (Mathworks, MA, USA).

To quantify which factors affected dopamine levels, regression coefficients were estimated for each animal at each time point in either a 7s window spanning from 2s before reward delivery to 5s after reward or a 9s window from 2s before cue onset to 7s after (cue and lever extension periods). A linear model was used with a constant term, representing an ordinary least-squares fit of the given regressors to the data over trials.

For analysis of the reward-evoked dopamine on all correct trials, the regressors were: (1) reward type (food trials were assigned 1, sucrose solution were assigned –1), (2) trial type (forced 1, choice –1), (3) SWITCH trial, (4) MORE trial, (5) cumulative number of food pellets consumed, and (6) preference for food in the choice trials at the end of each block (transformed to be 1 for 100% preference for food and –1 for 100% preference for sucrose solution, as well as interaction terms for (7) reward type x SWITCH, (8) reward type x MORE, and (9) reward type x preference. For the analysis of cue-evoked dopamine on all correct trials, the regressors were again: (1) reward type, (2) trial type, (3) amount of food consumed, and (4) block-by-block food preference, and 3 terms for the interaction between

reward type and the other regressors. For the analysis of the influence of reward-evoked dopamine on trial t-1 on cue-evoked dopamine on trial t, we extracted the peak dopamine level in a 5s window post-reward for each trial and examined its influence on cue-evoked dopamine in the next trial in which that reinforcer was selected (either a correct forced trial of the same type or a choice trial where that reinforcer was chosen). We also included regressors for reward type and trial type in this analysis.

Each trial in each regressor was modeled with a single value. All regressors except for choice performance, whether continuous or categorical, were mean-centered. Regression coefficients in each animal were averaged. The significance of the regression coefficients in the 5s post-event period was tested against a population of 1000 coefficients obtained by randomly permuting the pairings between the regressors and the data. Permutation tests were considered significant at any time point when the regression coefficient from the real data exceeded the maximum or minimum of the permuted population of coefficients ($p < 0.05$, corrected for multiple comparisons over the 5s after event onset; i.e., $p < 0.001$ uncorrected).

The discriminability of dopamine signals between pairs of trial types (e.g., expected sucrose versus surprise food) was analysed in each individual animal at each time point in a 5s period after either reward delivery or cue onset (between cue onset and lever extension) using the area under the receiver operating characteristic curve (auROC). The auROC from each animal was then averaged and significant discriminability at each time point was determined using 1000 random permutations of the trial types and re-computing the auROC to generate a null distribution. Permutation tests were considered significant at any time point when $p < 0.05$, corrected for multiple comparisons (i.e., $p < 0.001$ across 50 timepoints). To calculate

a dopamine discriminability index (Fig. 4.6, 4.7, 4.9), we calculated the auROC using the average dopamine in the 5s window after reward delivery for a particular trial type.

For the analysis of the reward period, we compared (a) value surprise: (expected food versus MORE food / expected sucrose versus MORE sucrose); and (b) identity surprise (expected food versus SWITCH sucrose / expected sucrose versus SWITCH food). For the analysis of the cue period, we compared (a) all forced trials (food versus sucrose solution for all trials, irrespective of whether they would be MORE or SWITCH trials at reward delivery), (b) all choice trials (food chosen or sucrose solution chosen), or (c) trial types (forced food versus choose food or forced sucrose solution versus choose sucrose solution).

4.2.5 Histology

At the end of the experiment, the rats were deeply anaesthetised with sodium pentobarbitone (200mg/kg) and electrolytic microlesions were made at the electrode locations before they transcardially perfused with saline followed by a 10% formol saline solution. Brains were extracted and placed into a formol saline solution. Subsequently, the brains were placed in a sucrose/formalin solution for 24 hours, before being frozen and sectioned in 50 μm coronal slices using a cryostat. Sections were stained with cresyl violet and electrode locations were verified by two experimenters to confirm the electrode locations. Five coronal slices demonstrating the electrodes recording locations are presented in Fig. 4.2.

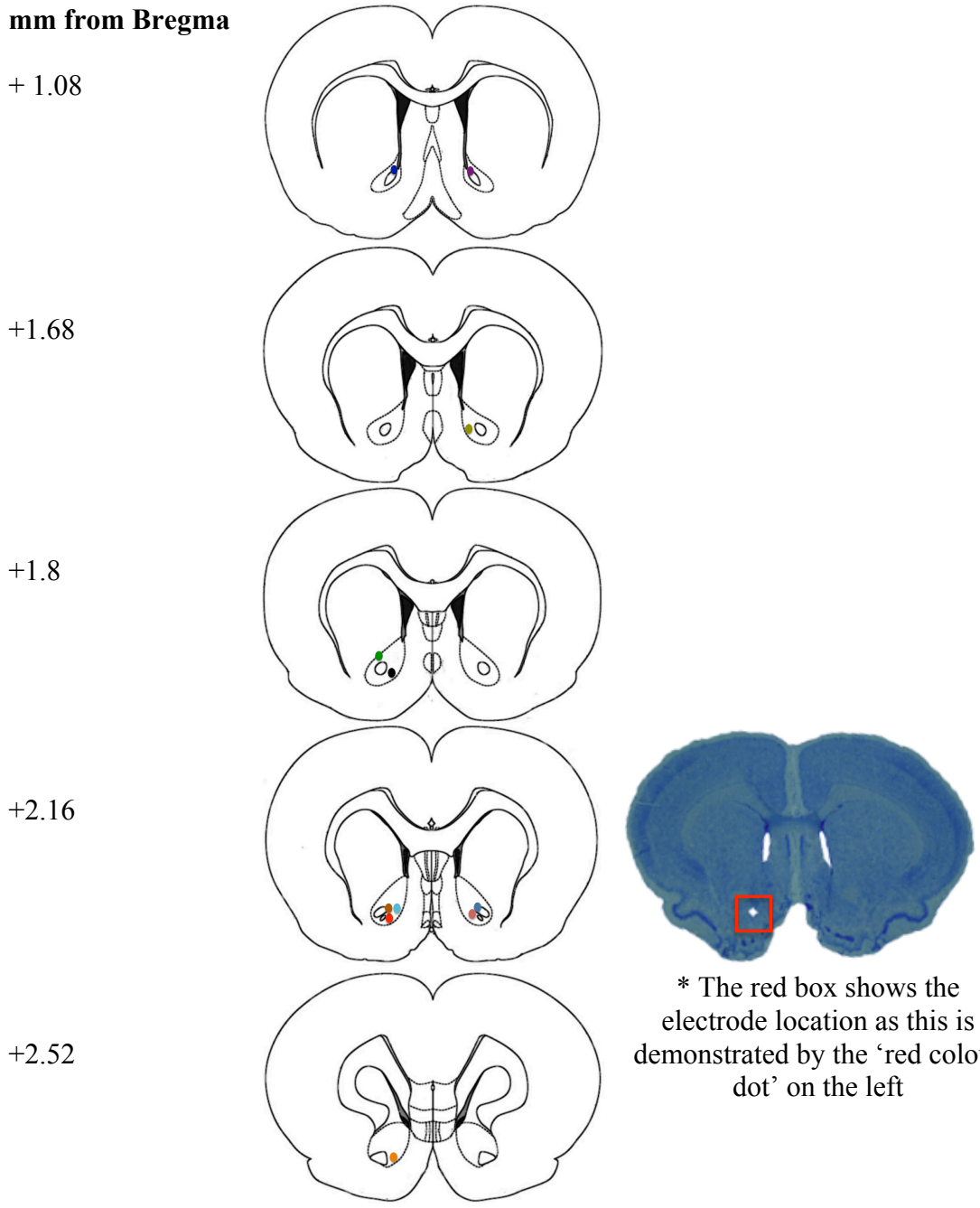


Figure 4.2: Representation of recording sites
Schematic, along with an example photomicrograph, of the recording locations in the nucleus accumbens core. The numbers next to each section indicate distance in mm anterior to bregma. (Adapted from the atlas of Paxinos and Watson, 2005).

4.3 Results

4.3.1 Behavioural performance during Baseline food-restricted sessions

Food-restricted rats were trained to perform a 2-option operant decision-making task where selection of each option was associated with a particular type of reward (food pellet or 95µl of 20% sucrose solution or 190µl of 30% sucrose solution) (Fig. 4.1). They were initially given a sampling session with 80 "forced" trials, where only one type of reward was available on each trial, followed by two behavioural reward titration sessions ('Pre-Test A and B') where they both sampled each reward type in forced trials and then also made choices between them on "choice" trials. The choice trials allowed us to determine if the rats had a preference for one of the rewards. If an animal chose the option associated with a particular reward on a majority of free choice trials in any session, it is described as having a subjective preference for that option and therefore that its subjective value of that reward was higher than the other reward type in that session.

The group as a whole had a very strong preference for sucrose over pellets when tested with 190µl of 30% sucrose solution ($t_{10} = 5.497$, $p < 0.001$). However, when tested with 95µl of 20% sucrose solution it was subjectively indifferent to the reward types in both Pre-Test sessions (t-test against 50% for food choices: session 1: $t_{10} = -0.473$, $p = 0.65$ and session 2: $t_{10} = 1.089$, $p = 0.30$) (Fig. 4.3).

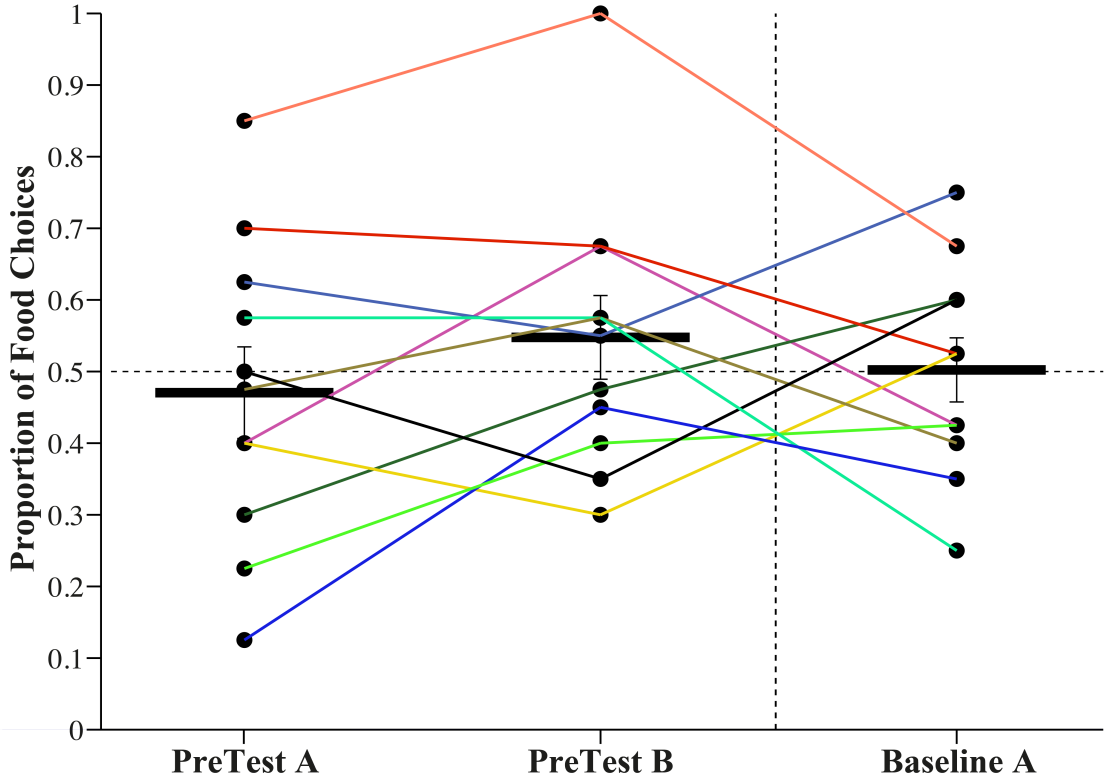


Figure 4.3: Proportion of food choices
Behavioural preference during choice trials during two Pre-Test sessions and the first recording session (Baseline A). Separate coloured lines, showing the variation of food choices for each animal across the three sessions.

The rats also responded at comparable latencies to press the lever associated with each reward type when only one of the options was available (session 1: $t_{10} = 0.327$, $p = 0.751$ and session 2: $t_{10} = 0.894$, $p = 0.392$).

In the first recording session ('Baseline A'), the group of rats overall displayed the same lack of preference for either reward type on choice trials ($t_{10} = 0.041$, $p = 0.97$), no difference in response latencies to the two options on forced trials ($t_9 = 1.13$, $p = 0.29$; the data from one animal was lost owing to a computer error) and no differences in the numbers of wrong lever choices or missed trials (both < 2% of trials, $t_{10} < 1.42$, $p > 0.18$). The stability of the pattern of choices was confirmed by a repeated measures ANOVA comparing choices in the 2 titration sessions and the baseline session, which showed no effect of session ($F_{2, 20} = 1.25$, $p = 0.309$) (Fig.

4.3). Therefore, while there was some variability in each rat's subjective valuation of the two options, the preference of the group as a whole was on average equal and consistent across sessions. This allowed us to investigate how mesolimbic dopamine encodes surprising changes in reward *identity* while controlling for reward value.

4.3.2 Dopamine release in response to outcome surprise in a food-restricted state

We monitored dopamine release in the nucleus accumbens (NAc) core (Fig. 4.2) using fast-scan cyclic voltammetry while rats performed a modified version of this reward identity decision paradigm. To determine how different aspects of outcome surprise are encoded by dopamine release, we focused our initial analyses on dopamine levels on forced trials during the reward delivery period (Fig. 4.4).

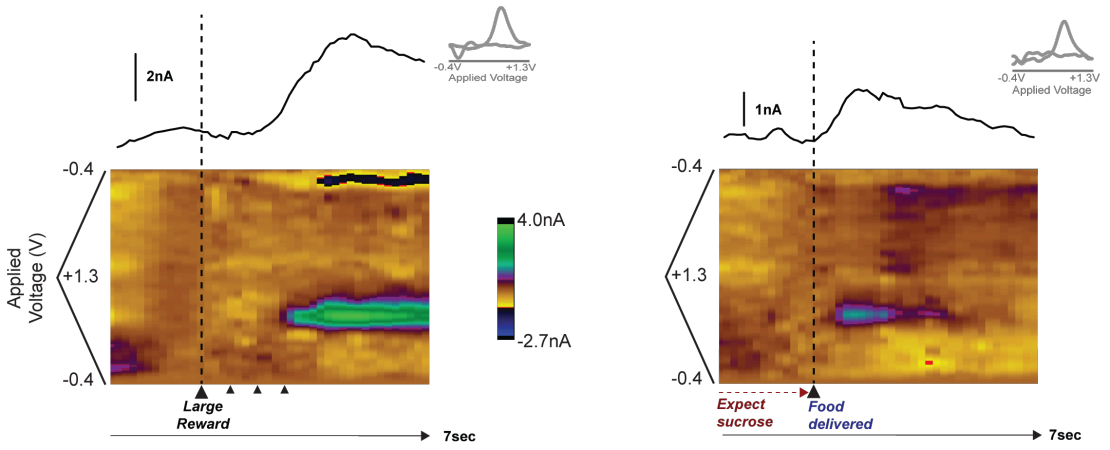


Figure 4.4: Individual example MORE (left) and SWITCH (right) trials
Each panel depicts the recorded current x applied voltage in a pseudocolour plot from 2s before and 5s after reward delivery. The upper trace depicts the extract dopamine signal along with an example cyclic voltammogram identifying the detected current as dopamine.

Note that until the reward is dispensed, surprise trials are otherwise identical to standard trials (Fig. 4.5).

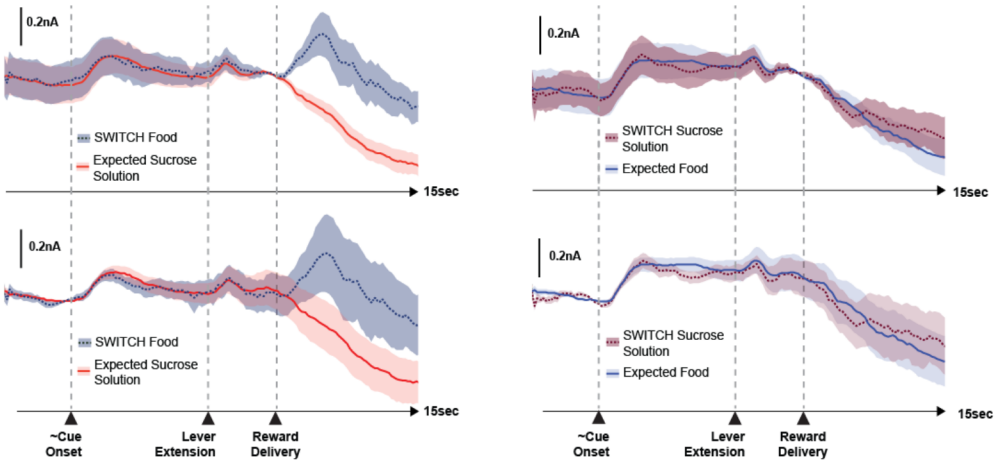


Figure 4.5: Reward delivery during Baseline sessions
Baseline A dopamine signals on expected and SWITCH trials for the whole trial length (from before cue onset to 5s after reward delivery) with the background subtraction set to 0.5s before reward delivery (upper panels) or 8s before reward delivery (~0.5s before cue onset). Note that as the SWITCH of reward type is only signalled at reward delivery, dopamine signals prior to this point are equivalent.

A linear regression showed that there was a significant positive influence of MORE trials on reward-elicited dopamine release, as well as an interaction between reward type and MORE trial (Fig. 4.6a). The unexpected delivery of a larger-than-expected magnitude of either food or sucrose solution caused a significant increase in dopamine compared to the standard trials ($p < 0.05$, corrected for multiple comparisons) (Fig. 4.6b).

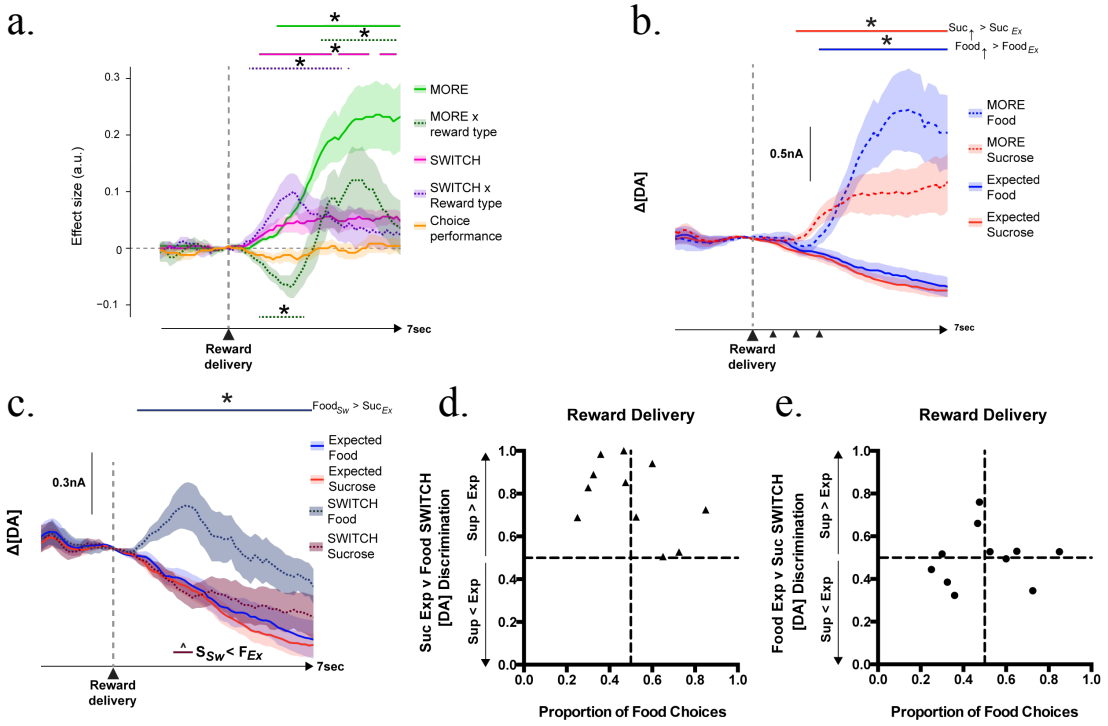


Figure 4.6: Dopamine signals in Baseline A
(a) Average effect sizes from a general linear model of post-reward dopamine signals with regressors for reward type (food / sucrose), trial types (forced / choice, MORE, SWITCH), choice performance, and interactions between reward type and these factors. **(b, c)** Average dopamine signals in the 2s before and 5s after reward delivery for the Expected food and sucrose trials (filled blue and red line, respectively) with the MORE food and sucrose trials (dotted blue and red line, respectively) (**b**) and SWITCH food and sucrose trials (dotted dark blue and dark red line, respectively) **(c)** Arrows show the time of reward delivery (4 arrows for 4 food pellets / boluses of sucrose solution in MORE trials). **(d, e)** Dopamine discrimination index for each animal (see Experimental Procedures) in the 5s post reward for SWITCH Food versus Expected Sucrose **(d)** or SWITCH Sucrose versus Expected Food (**e**) plotted against each animal's food choices. (Lines: *, $p < 0.05$ permutation tests, corrected for multiple comparisons; ^, $p < 0.05$ permutation tests, *uncorrected for multiple comparisons*. All averages, mean ± SEM).

Importantly, there was also a significant positive influence of SWITCH trials and a significant interaction between SWITCH and reward type on dopamine release (Fig. 4.6a, c-e). Note that if dopamine only reflects deviations from predicted value, then the delivery of a reward of different identity but equivalent value should not cause any change in dopamine release. Specifically, when the rats unexpectedly received food reward after responding on the sucrose solution-associated lever, there was a significant, time-locked increase in NAc core dopamine, even though the animals were displaying no overall preference for this option on the choice trials

(standard *sucrose solution* trials versus SWITCH *food* trials - where prior to reward delivery sucrose solution had been expected -: $p < 0.05$, corrected for multiple comparisons) (Fig. 4.6c, d). This was consistent irrespective of training history (Fig. 4.7a) and was unrelated to choice performance (no influence of choices on patterns of dopamine release, Fig. 4.6a; or correlation between choices and dopamine discriminability: $r^2 = 0.183$, $p = 0.167$, Fig. 4.6d). When each animal was tested individually, it was found that there was a significant increase on SWITCH food trials in dopamine in 7 of the 11 rats (Mann-Whitney U: $p < 0.05$).

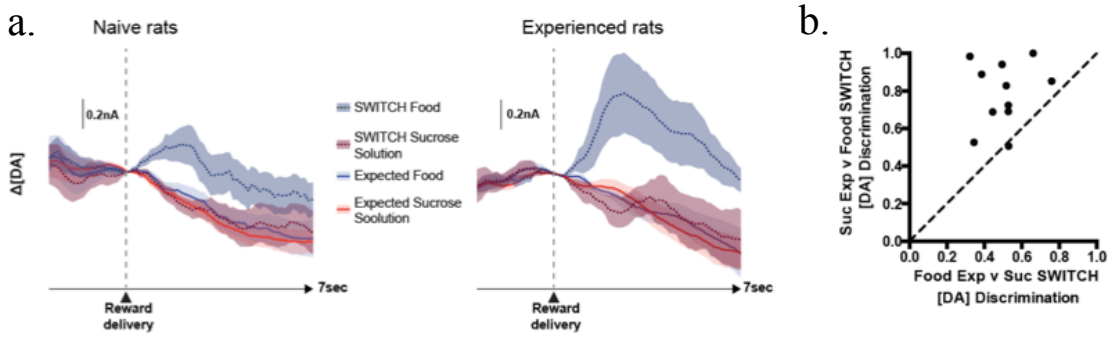


Figure 4.7: Comparison of naive versus experienced rats
(a) Baseline A dopamine signals on expected and surprise reward trials, separated by whether the rats were naïve (left) or had previously been through a separate 2-option decision-making task (right). **(b)** Comparison for dopamine discrimination indices for Food SWITCH or Sucrose Solution SWITCH trials for the 11 individual animals in Baseline A. 11/12 animals fall to the left of the dotted line, showing higher discrimination on Food than Sucrose SWITCH trials.

However, this effect was not observed on all SWITCH trials. When the rats had made a response on the food-associated lever but received sucrose liquid reward there was no consistent detectable change in dopamine release (comparison of expected food versus surprising sucrose trials: $p > 0.05$) (Fig. 4.6c, e). In fact, the closest observable phasic trend was for an initial *reduction* in dopamine on surprise sucrose solution trials compared to standard food trials, though this only reached significance in this session at an uncorrected threshold (Fig. 4.6c). None of the

individual rats showed a significant increase of dopamine on SWITCH sucrose solution trials and 2 of the 11 had a statistically reliable decrease (Mann-Whitney U: $p < 0.05$). Moreover, 10/11 rats had lower dopamine discrimination on SWITCH sucrose trials than SWITCH food ones (Fig. 4.7b).

The selective increase in dopamine on food SWITCH trials, as well as a small transient reduction in dopamine on sucrose solution SWITCH trials, was replicated in a second Baseline session ('Baseline B') run after the first selective satiety manipulation (see Chapter 5), demonstrating that the effect was consistent and could not be explained by the novelty of either reinforcer (Fig. 4.8a, b). Moreover, a similar pattern of results was also observed on the free choice trials in both Baseline sessions (Fig. 4.8c).

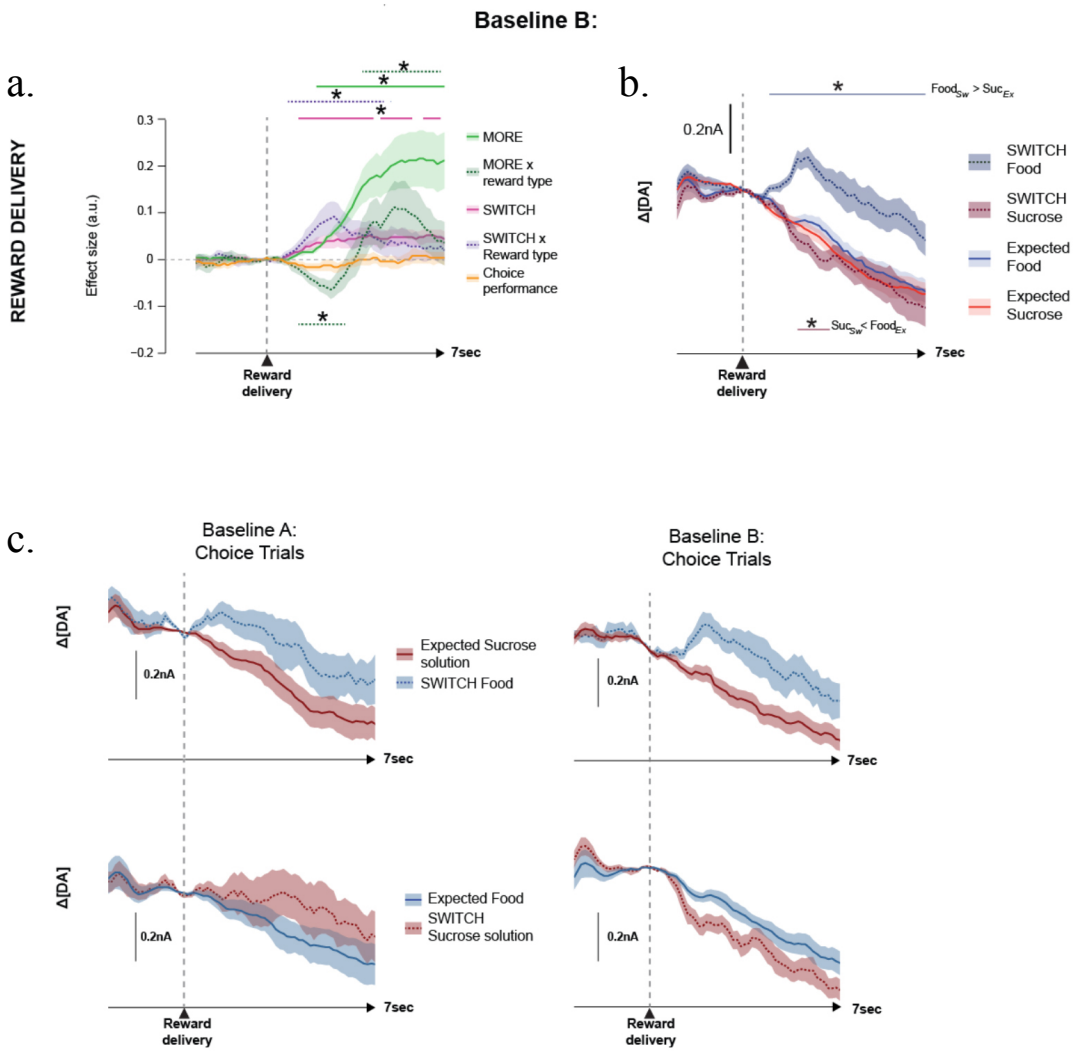


Figure 4.8: Reward delivery dopamine data
Average regression coefficients **(a)** and dopamine release **(b)** in Baseline B, which occurred in between the two selective satiation sessions. **(c)** Dopamine on Choice SWITCH trials in Baseline A and B. Note that as there were only 6 choice SWITCH trials in a session, it was not possible to do any meaningful statistical analysis on these data.

4.3.3 Dopamine release in response to predictive cues in a food-restricted state

The first set of results implies that NAc core dopamine release is selectively encoding a surprising change in reward identity from sweet liquid to food even after controlling for differences in value derived from the observed in the pattern of the rats' choices. This could mean that dopamine is encoding some aspect of the sensory surprise or change in motor programmes when unexpectedly required to consume food. In this case, changes in dopamine release should only be observed at reward

delivery and not at cue-onset. However, another possibility is that this difference in dopamine release is in fact also a reward prediction error signal, with the dopamine-associated value for food in a hungry state being higher than for sucrose solution, even if the rats' subjective preference did not reflect this. If so, cue-elicited dopamine should be greater on food than sucrose solution trials.

As can be observed, the latter was what we observed in both Baseline sessions: significantly larger cue-elicited dopamine release on forced food trials compared to forced sucrose trials (Fig. 4.9).

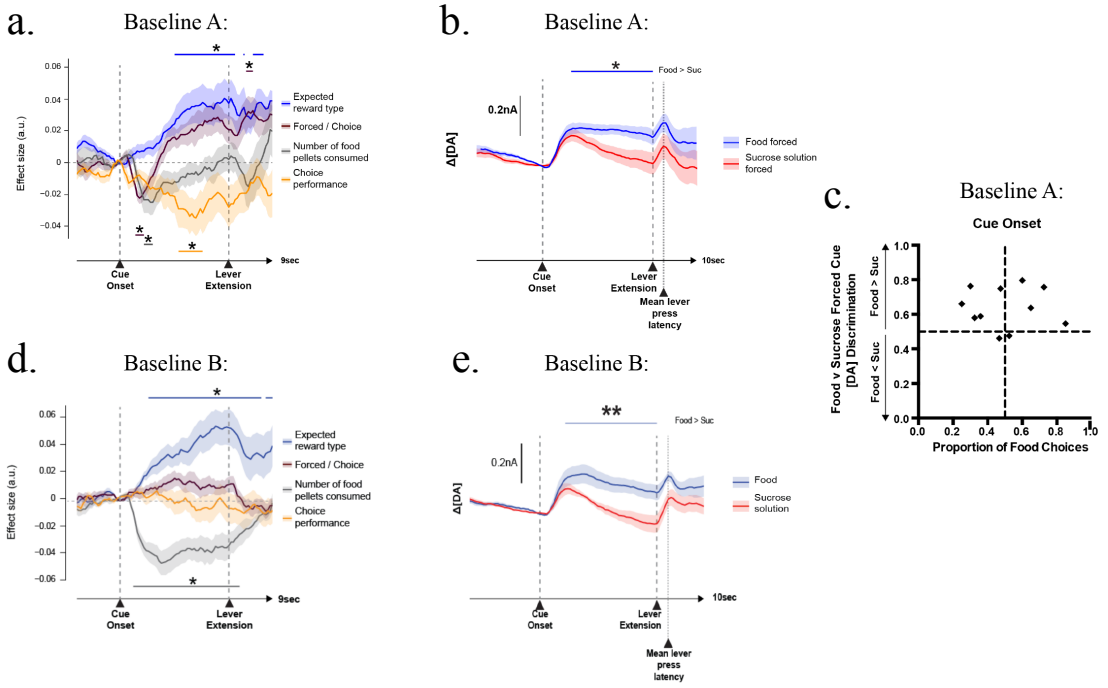


Figure 4.9: Forced versus Choice trials dopamine release
(a, d) Average effect sizes from a general linear model of post-cue dopamine signals with regressors for reward type (food / sucrose), trial type (forced / choice), amount of food consumed, choice performance, and interactions between reward type and these other factors. **(b, e)** Average dopamine signals in the 3s before and 7s after cue presentation. **(c)** Dopamine discrimination index for each animal in the 5s post cue period for Forced Food versus Forced Sucrose. (Lines: *, $p < 0.05$ permutation tests, corrected for multiple comparisons; ^, $p < 0.05$ permutation tests, *uncorrected for multiple comparisons*. All averages, mean ± SEM).

Again, the size of this difference did not positively reflect the strength of rats' subjective preference for food (Fig. 4.9a, c). However, the size of the dopamine

release was influenced positively by the magnitude of reward-elicited dopamine on the previous trial of that type (Fig. 4.10a) and also negatively by the amount of food consumed in the session (Fig. 4.9a, d, Fig. 4.10b).

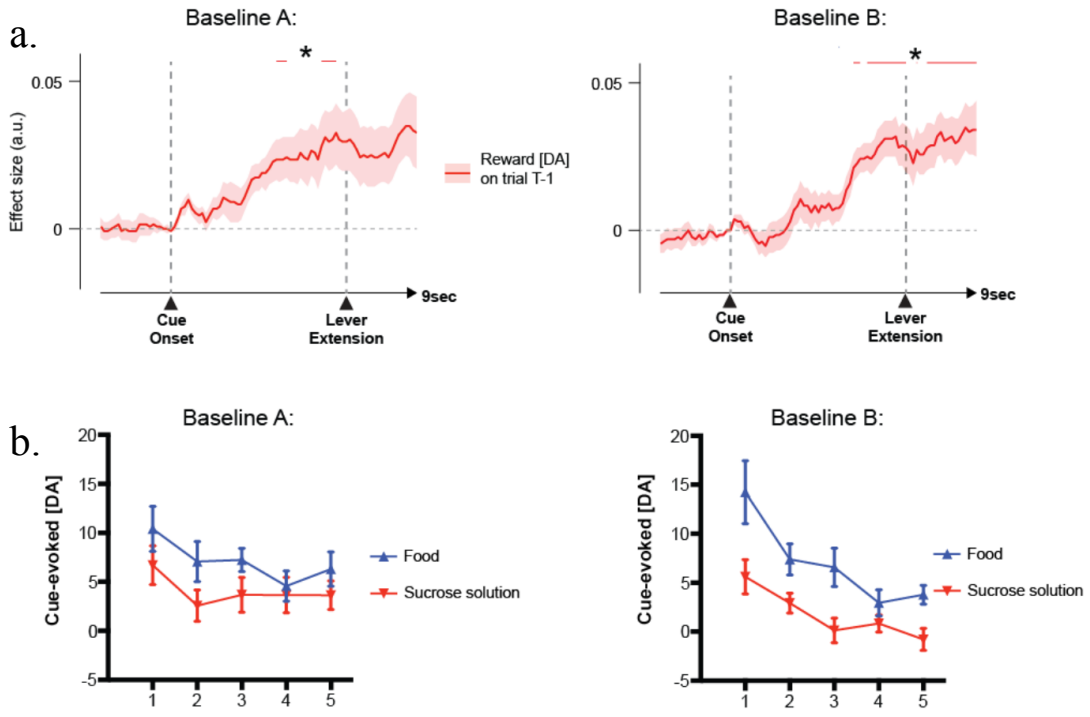


Figure 4.10: Cue-elicited dopamine data
(a) Average regression coefficient in Baseline A and B for the influence of the previous reward-elicited dopamine on subsequent cue-elicited dopamine. **(b)** Average dopamine levels in the 5s post-cue period divided into 5 equally sized bins each of 2 blocks of trials in Baseline A and B.

To investigate further this consumption-related change in cue-elicited release, dopamine levels either on forced food or forced sucrose solution trials were analysed across 5 equally sized blocks in the session. This revealed a significant linear main effect of block on food trials in both sessions ($F > 5.8$, $p < 0.037$), driven by a significant reduction in food cue-evoked dopamine as increasing amounts of food were consumed ($p < 0.05$). Although a similar trend was present on sucrose solution trials, this only reached significance in Baseline B (Baseline A: $p = 0.16$; Baseline B: $p = 0.02$).

This general pattern of cue-elicited dopamine was also observed when the choice trials were analysed. Interestingly, however, dopamine levels were significantly attenuated on choice trials when food was subsequently selected compared to forced food trials (Fig. 4.11). This does not simply reflect the sensory variation between the trial types (two versus one cue lights being presented on choice or forced trials, respectively), as there were no differences when comparing dopamine on sucrose solution choice and forced trials. Instead, dopamine-based representation of the value of food appears altered by the presence of sucrose solution as a viable alternative.

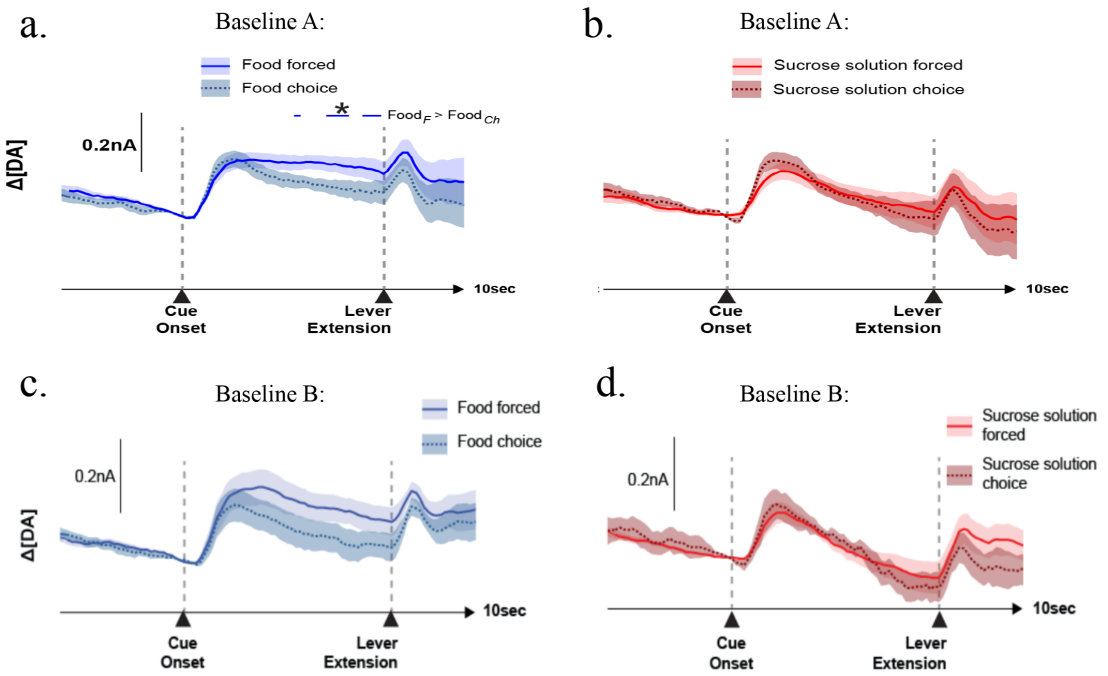


Figure 4.11: Forced versus Choice trials
(a, b) Comparison of average cue-evoked dopamine on Forced Food or Sucrose trials with Choice Food or Sucrose trials in Baseline A. **(c, d)** Comparison of average dopamine release on forced or free choice trials in Baseline B (Lines: *, $p < 0.05$ permutation tests, corrected for multiple comparisons; ^, $p < 0.05$ permutation tests, *uncorrected for multiple comparisons*. All averages, mean ± SEM).

4.4 Discussion

In this experiment, we showed that there is increased dopamine release in NAc core after (a) surprising changes in reward value and (b) surprising changes in reward identity when controlling for subjective value. However, the latter change was selective for situations when going from sucrose solution to food but not from food to sucrose solution. Food cues also elicited greater dopamine release than sucrose solution cues.

We focused our analyses in two time windows: at reward delivery and at cue onset. During the reward delivery time, surprising increases in the magnitude of reward elicited large increases in dopamine release. This finding is consistent with what has already been demonstrated in previous research in both dopamine cells and cue-evoked release (Hart et al, 2014; Tobler et al., 2005; Roesch et al., 2007). Interestingly, when the animals were expecting to receive sucrose but surprisingly a pellet was delivered a significant increase in dopamine release was observed. However, there was no significant in increase in dopamine release when animals were expecting pellets and they were instead receiving a bolus of sucrose.

In order to control for differences in the sensory properties of the rewards or in the delivery methods a series of steps were followed. First, both the sucrose solution and the food pellet were delivered to the same food cup with as similar latencies as possible. Second, the delivery of sucrose solution was made as salient as the pellet delivery, by adding a brief click sound when the solenoid opened to mimic the sound that a pellet makes when it drops in the food cup. Thus, any difference in perception of the latency of either reward is minimised. If the difference in dopamine release reflected the sensory properties alone, then one might expect dopamine release to be higher on Expected Food than Expected Sucrose trials, and unrelated to patterns of

dopamine elicited by food and sucrose cues. It would therefore seem that differential dopamine release is not being driven primarily by differences in the sensory properties or delivery methods.

This finding at first glance might appear to contradict the predictions of TDRL models that would suggest that RPEs, and consequently dopamine, should respond to discrepancies in the overall value of the different options but should not be influenced by changes in identity. However, a pure identity surprise signal that was independent of value would be expected to be seen whenever there was a surprising change in identity, which was not observed. Moreover, these differences between food and sucrose solution trials were also present at cue onset. Analyses at the time that the cues were presented showed larger dopamine release was observed when the rats were presented with cues signalling a forced food than a forced sucrose solution trial. Therefore, a more parsimonious explanation might be that dopamine is consistently signalling a reward prediction error, where value is encoded in a different frame of reference than subjective preference.

If so, this would contradict recent work in non-human primates (Lak et al., 2014) in which midbrain dopamine neurons have been shown to encode reward prediction errors that reflect the animals' subjective preference for different reward type. This may reflect differences in the amount of training (the monkeys had extensive experience with different reinforcers), differences between midbrain dopamine cell activity and patterns of dopamine release, or differences in the behavioural state of the animals.

One hypothesis that merits further investigation is to determine whether the differences we observe are directly related with the fact that the animals enter the task in a food-, but not fluid-, restricted state. One possibility is that dopamine encodes this

need for nutritional food separate to the behavioural preference of the animal, as food is nutritionally more valuable to a hungry rat. A single food pellet contains nearly two times the calories compared to the chosen volume and concentration of sucrose solution, as well as elements such as protein and minerals not available in the liquid. Other research focused on the nutritional aspects of different rewards has already showed that the caloric value of different rewards have greater influence over NAc activity and mesolimbic dopamine release than sweet taste, even though the latter has strong influences over animals' decision policies (Beeler et al., 2012; McCutcheon et al., 2012; Ren et al., 2010; de Araujo et al., 2013).

To determine the influence of motivational state on dopamine responses, a devaluation experiment was conducted and presented in detail on Chapter 5.

5 Mesolimbic dopamine reflects state-dependent prediction errors

Summary

Mesolimbic dopamine encodes the benefits of a course of action. Dopamine levels are also strongly shaped by current physiological needs and by the nutritional properties of reinforcers (Ahn & Phillips, 1999; de Araujo et al., 2013; McCutcheon et al., 2012). To date, the relationship between phasic dopamine, reward prediction errors, motivational state and reward-guided choices remains poorly understood. In the previous chapter we described the patterns of dopamine release as food-deprived rats performed a two-choice decision-making instrumental task. The current chapter will focus on how those patterns of dopamine release on the same task were modulated by selective satiation with one of the two reinforcers. Both choice behaviour and dopamine release was found to be state-dependent. Following food devaluation, rats' preference shifted towards the sucrose solution and on trials where the reinforcers were unexpectedly swapped, there were now increases in dopamine release following surprise delivery of sucrose solution, but not food, This supports the conclusion from the previous chapter that mesolimbic dopamine does not encode identity prediction errors, but instead signals reward prediction errors in a frame of reference that more closely reflects current nutritional needs than subjective preference. These changes following selective satiation were rapid and selective, with dopamine release returning to pre-satiation patterns when the animals were re-tested in a standard food-restricted state. Such rapid and selective adaptation of dopamine-

associated value signals could provide an important signal to promote efficient foraging for a varied diet.

5.1 Introduction

When foraging for food, animals need to continuously take into account their current physiological state and needs. In a state of hunger, it may be more beneficial to choose a food with higher caloric value than one with a lower value, assuming that the cost for gaining each one of these is similar. A more nutritious food will provide more energy and at the same time will increase the chances of surviving. Nevertheless, there are frequently situations when animals behave in a suboptimal way and prefer to choose a food based solely on specific attributes and ignoring an overall more valuable option. A case of such suboptimality in decision-making has already been described in Chapters 2 & 3 of the current thesis, in which macaque monkeys demonstrated a strong preference for an option which was objectively less valuable i.e. a highly valued reward, when the alternative option was the same highly valued reward presented in compound with a lower valued reward. The ventromedial prefrontal cortex/medial orbitofrontal cortex (vmPFC/mOFC) in those experiments exhibited strong activation at the time that cues associated with the two reward types were presented. However, in that experiment, the monkeys' state was stable from session to session. Therefore, it is not clear what role motivational state plays in shaping trial-by-trial reward value responses and reward prediction errors and how these influence behaviour.

The phasic activity of midbrain dopamine neurons and dopamine release in regions such as the nucleus accumbens (NAc) signals predictions of future reward and discrepancies between such predictions and received rewards (Gan et al., 2010; Montague et al., 1996; Schultz et al., 1997; Syed et al., 2016). However, the value of such future rewards are not fixed but instead depend on an organism's current nutritional needs, particularly in comparison with recent consumption. Several studies

have shown that dopamine levels in the presence of rewards are influenced by current physiological state and also the nutritional content of reinforcers (Ahn & Phillips, 1999; Bassareo & Di Chiara, 1999l Beeler et al., 2012; de Araujo et al., 2013; McCutheon et al., 2012).

For instance, an in vivo brain microdialysis study by Wilson et al., (1995) investigated the anticipatory and consummatory effects of food on tonic dopamine levels in the NAc. Dopamine levels increased more in food-restricted rats than non-food-restricted rats both during the anticipation of an upcoming sweet reward and consumption of the reward. This was not simply related to the amount of reward consumed as a separate experiment where intake was equalised also found greater increases in dopamine in food-restricted rats. This finding of increased dopamine levels in a food-restricted state was also replicated more recently in a study where rats had to make lever presses to gain reward (Ostlund et al., 2011). Phasic NAc core dopamine release is similarly increased in food-restricted animals (Cone et al., 2014). Together, this shows that dopamine levels and phasic dopamine release are influenced by the amount of food consumed and the motivational state of the animal at the time of the task during goal-directed actions.

In Chapter 4, we found that food restricted rats exhibited greater dopamine release in the presence of options associated with food delivery than sucrose solution delivery. The current experiment aimed to investigate the dynamic influence of motivational state on these signals by examining patterns of dopamine release after animals had free access to one of the two reinforcers ("selective satiety": Rolls et al., 1983; Ahn & Phillips, 1999). Specifically, we aimed to investigate whether the food-specific identity surprise signals that were observed in the previous experiment (Chapter 4) were state dependent. Additionally, by monitoring how patterns of

dopamine release updated following selective satiety, we could investigate how dopamine prediction errors are influenced by selective changes in subjective value and how value predictions and behavioural preferences updated with experience of the reinforcers in a new state.

5.2 Methods

All experimental procedures were the same as the experiment in Chapter 4. However, only 13 rats were used for this experiment. This included 8 rats out of the 11 naive rats that were used in the previous experiment as well as the 5 out of the 6 more experienced rats. A total of 5 rats were excluded from the electrochemistry analysis for the reasons explained in Chapter 4, leaving electrochemistry data from a total of 8 rats that are reported.

All animals that underwent FCV completed 2 separate selective satiation-recording sessions, separated by a single session run in baseline food-restricted state. Selective satiation sessions had exactly the same structure as the sessions described on Chapter 4 but with the difference that prior to those sessions animals were given 1 hour's free access to either sucrose solution or food pellets in the testing chamber before the task commenced. During food devaluation, the animals did not have access to any fluids. The session order was Baseline A (presented in Chapter 4), Selective Satiation A, Baseline B (presented in Chapter 4), Selective Satiation B, with the order of satiation (food or sucrose solution) counterbalanced across animals. Each of the 4 sessions took place on separate days.

5.2.1 Data analysis

For the analyses, we combined the data across the counterbalanced food / sucrose devaluation sessions as well as the baseline data from Chapter 4 to look at the effects in the whole group for each manipulation. We focused here on (a) SWITCH signals again, but also (b) contrasted dopamine signals on the expected reward trials in sessions depending on whether the reward type was devalued or not (non-devalued is termed "valued" throughout) (e.g., expected valued sucrose versus expected devalued food / expected valued food versus expected devalued sucrose). As there were only a few MORE trials (increased amounts of the expected reward) in each session, it was not possible to directly compare these trial types using the auROC approach. Therefore, we found the average signal in a 3s window from 0.5-3.5s after reward delivery, and compared these using a repeated measures ANOVA within subjects factors devaluation session and reward type and devaluation order as a between subjects factor (NB. the results were unchanged if a peak measure was taken instead). To examine changes in dopamine release after a state change on standard trials, we combined the data across reward types in the Selective Satiation sessions into valued and devalued trials and then, using an auROC, compared dopamine on expected reward dopamine on these two trial types against signals recorded in the baseline session presented in Chapter 4. We also made comparable analyses on the data from each reinforcer in isolation. To look at how the valued signals changed during the session, we first broke these sessions down into 5 bins, each containing 2 blocks of trials (i.e., 8 forced food and 8 forced sucrose solution) and subtracted the average dopamine signal in a 0.5-3.5s window after reward delivery on expected reward trials in the Baseline session and the valued expected reward trials during Selective Satiation session. To account for variance in the data, we log transformed

the values before running a repeated measures ANOVA, with trial bin as a within-subjects factor, and *post-hoc* 1-sample t-tests against zero to determine in which bins there was significantly greater dopamine during the valued trials than baseline trials.

We also analysed the cue data from the selective satiation sessions in a similar manner. First, we again compared all forced trials (food versus sucrose solution) either in food devaluation or sucrose solution devaluation sessions. We then examined dopamine release following presentation of food or sucrose solution cues in a valued or devalued state and contrasted it with dopamine release to that cue in the immediately preceding Baseline session (i.e., if food was devalued in Selective Satiation session B for a particular rat, the Baseline B data was used for comparison whereas if it was devalued in Selective Satiation session A, Baseline A data was used). Finally, the average dopamine levels in the 5s post-cue period on forced trials were taken for each of 5 equally sized bins and analysed. In particular, we focused on changes observed in the first 2-block bin (1st 8 forced trials for each reinforcer).

5.3 Results

5.3.1 Behavioural results

Dopamine release in a food-restricted state as presented in Chapter 4, is consistently greater for surprising delivery of food over sucrose solution, and for food cues over sucrose solution cues, even while the rats were displaying no overall subjective preference for the food. One possibility is that phasic dopamine release is encoding a state-dependent value signal. In order to test this, we recorded dopamine release in a subset of these same animals (n=8) during additional sessions of the same reward identity decision paradigm using a selective satiety procedure (Ahn & Phillips,

1999; Berridge, 1991; Rolls et al., 1986). This allowed us to manipulate the animals' state, and consequently both the subjective value of one of the options and the rats' immediate nutritional needs. Having free access to one of the two reinforcers resulted in the rats consuming, on average, either 11.5g (±1.24g SEM) of pellets or 24.75ml (±1.81ml SEM) of sucrose solution in the hour prior to the test session. This manipulation reliably altered the animals' preference for the reward types (t-test against 50% for food choices: devalue food pellets: $t_7 = -2.84$, $p = 0.01$; devalue sucrose solution: $t_7 = 4.13$, $p < 0.01$) (Fig. 5.1).

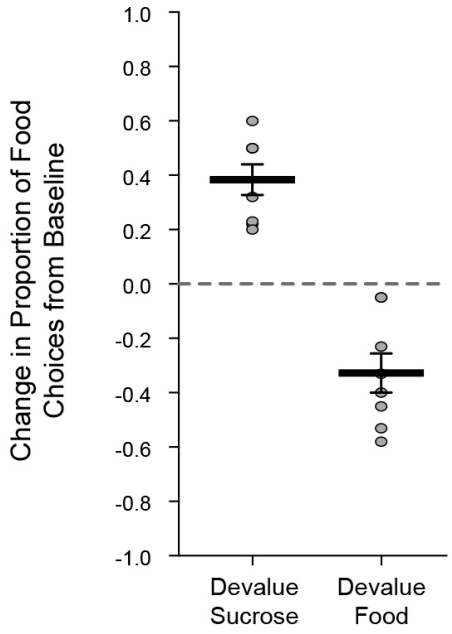


Figure 5.1: Preference change
Change in preference for food after an hour's access to sucrose solution (Devalue Sucrose) or to food (Devalue Food).

There was no difference in the magnitude of this change following satiation with either the food or sucrose solution ($t_7 = 0.29$, $p = 0.78$). There was also a significant increase in the number of missed trials and wrong lever choices on forced trials in the devaluation sessions compared to baseline sessions (main effect of devaluation: both $F_{1,7} > 7.02$, $p < 0.034$), an effect driven by a selective increase on

the devalued option (interaction between reward type and devaluation session: wrong choices $F_{1,7} = 9.75$, $p = 0.02$; missed trials: $F_{1,7} = 5.53$, $p = 0.051$). These changes in subjective valuation were temporary and specific to the devaluation session. As already mentioned in the previous chapter, preference returned to indifference in the baseline B session run in between the counterbalanced devaluation sessions ($t_7 = 1.08$, $p = 0.32$) and there was no change in choices from the pre-devaluation Baseline A session ($t_7 = 1.18$, $p = 0.28$).

5.3.2 Dopamine release at reward delivery following sensory-specific satiation procedures

To examine the effect of selective satiety on value-related dopamine signals at the time of reward delivery, we again ran a linear regression on the two devaluation sessions (Fig. 5.2a, b).

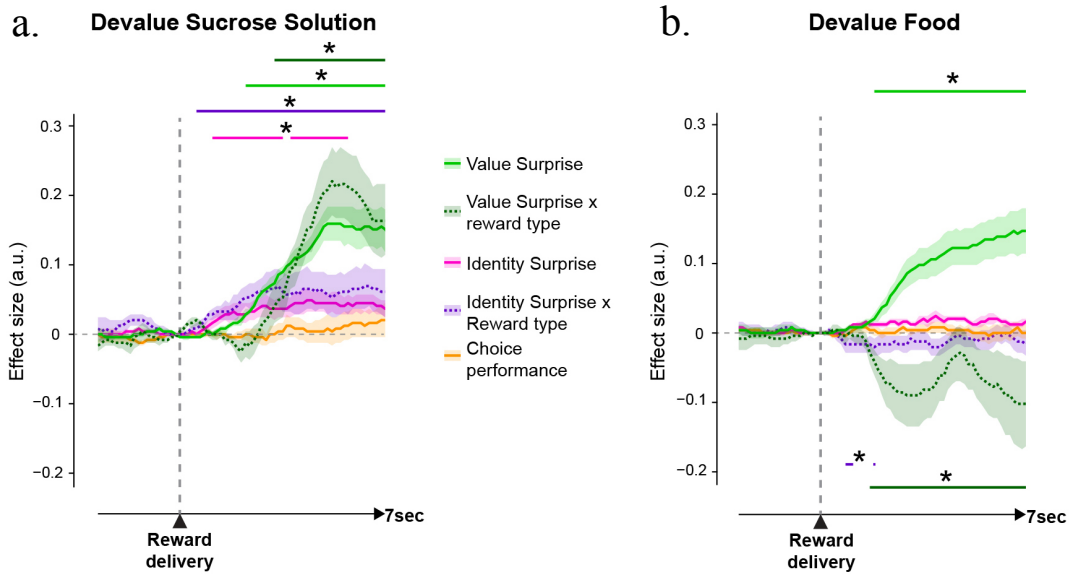


Figure 5.2: Effect size during devaluation
Average effect sizes (mean ± SEM) from a general linear model of post-reward dopamine signals after sucrose **(a)** or food **(b)** was selectively devalued. (Lines: *, $p < 0.05$ permutation tests, corrected for multiple comparisons; All averages, mean ± SEM).

While there was again a strong influence of MORE trials on dopamine release, the sign of the interaction term switched depending on whether food or sucrose solution was devalued (Fig. 5.2). As can be seen in Fig. 5.3, this was caused by a selected attenuation in dopamine release when receiving an increased quantity of the devalued reward. This demonstrates reinforcer specific satiety effects on value surprise trials.

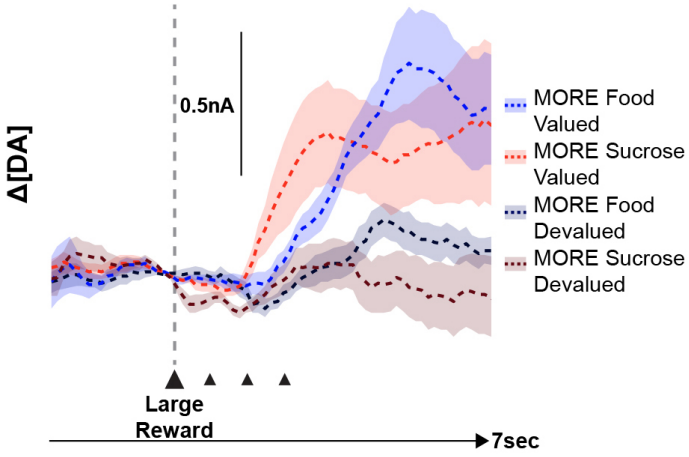


Figure 5.3: Dopamine release in MORE trials
Average dopamine release on the MORE food or sucrose trials divided up based on which reward type was devalued by selective satiation prior to the session.

Importantly, the same influence of selective satiety was also observed on SWITCH trials. In sucrose solution devaluation sessions, as expected, there was once again a transient increase in dopamine following the surprise delivery of a valued food pellet following a response on the sucrose solution lever (Fig. 5.2a, 5.4a, c). These signals were significantly more discriminable than during the baseline session (paired t-test on the dopamine discrimination index ($t_7 = 2.88$, $p = 0.028$). Crucially, the opposite pattern was observed in the food devaluation sessions: now, it was the surprising delivery of the valued sucrose solution that caused a selective increase in

dopamine release, whereas there was no observable increase following surprising pellet delivery (Fig. 5.2b, 5.4b, d).

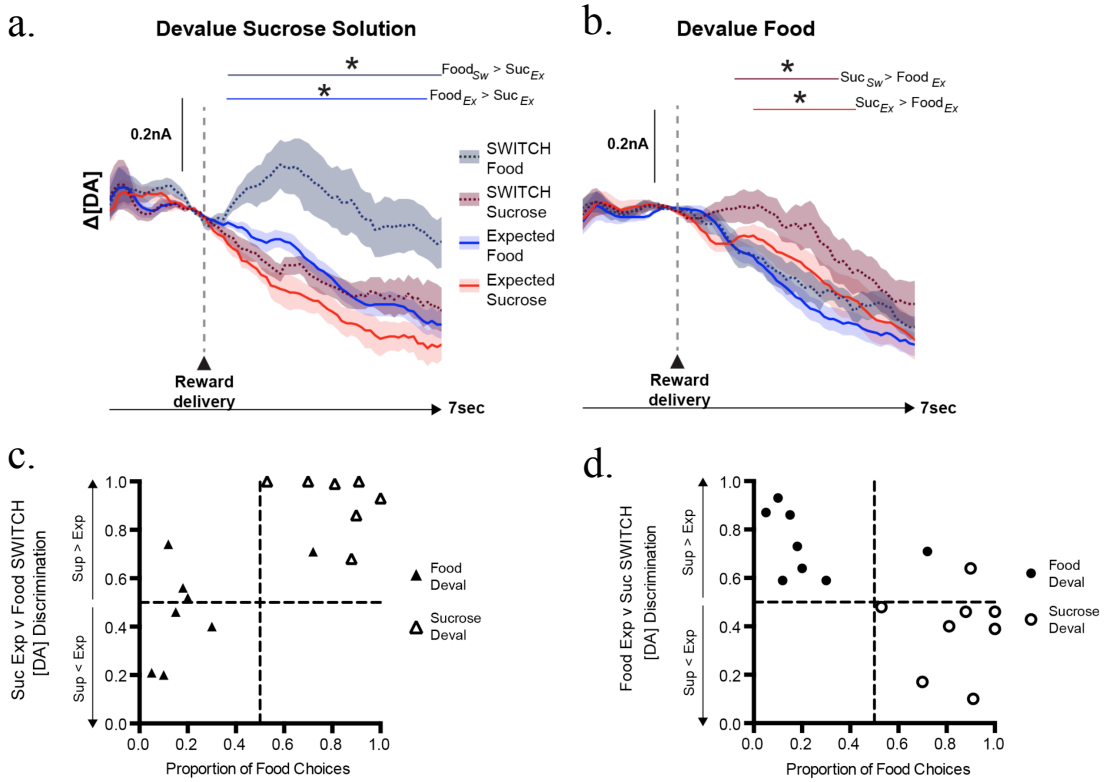


Figure 5.4: Dopamine release on Expected and SWITCH trials
Dopamine signals on Expected and SWITCH trials after sucrose **(a)** or food **(b)** was selectively devalued. Dopamine discrimination index for each animal in the 5s post reward for SWITCH Food versus Expected Sucrose **(c)** or SWITCH Sucrose versus Expected Food **(d)** plotted against each animal's food choices, divided up into food (filled symbols) or sucrose (open symbols) devaluation sessions. (Lines: *, $p < 0.05$ permutation tests, corrected for multiple comparisons; All averages, mean ± SEM).

Surprise-evoked dopamine release can thus be modulated by the state-based value of the reinforcers, further demonstrating that the pattern of dopamine observed in the Baseline sessions in Chapter 4 is distinct from the physical properties of the rewards.

It was also evident that dopamine release differed even on the standard trials for the valued and devalued options in a new state, even though the anticipated type of reward was always delivered (Fig. 5.4, 5.5). To investigate what could have been influencing this, we directly contrasted dopamine time-locked to reward delivery in

the devaluation sessions against an equivalent period in the baseline session (Fig. 5.5a). There was no consistent difference in the change in average dopamine levels after receipt of the devalued reward when compared to receiving that same reward in the baseline session ($p > 0.05$). Instead, there was a small but significant increase in dopamine when receiving the valued option for both reinforcers when compared to same situation during baseline testing. Moreover, when the session was divided into 5 blocks, this difference was found to be on average largest at the beginning of the session and diminished linearly as the session progressed (linear main effect of block: $F_{1,7} = 5.88$, $p = 0.046$) (Fig. 5.5b).

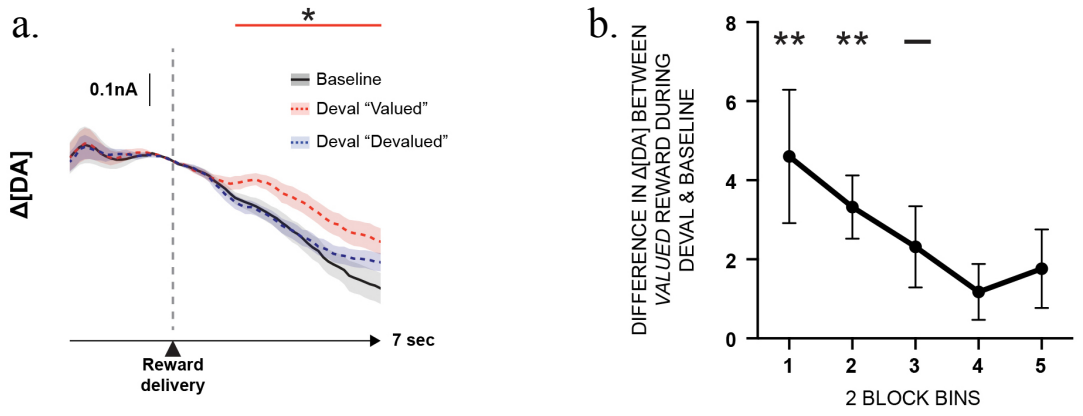


Figure 5.5: Dopamine release during Devaluation and Baseline sessions
(a) Comparison of dopamine signals when receiving expected rewards in Baseline and Devaluation sessions as a function of which reward type was devalued (non-devalued reward = "Valued"). **(b)** Difference between average dopamine release after reward delivery when receiving the valued option in Devaluation sessions or this same reward type in Baseline A (collapsed over reinforcers), divided into 5 bins each of 2 blocks of trials. (Lines: *, $p < 0.05$ permutation tests, corrected for multiple comparisons; **, $p < 0.05$; –, $p = 0.058$, 2-tailed t-test against 0. All averages, mean ± SEM).

5.3.3 Rapid updating of cue-elicited dopamine release after changes in state

The pattern of dopamine release at reward delivery suggests that value predictions are shaped by current motivational state in a reinforcer-specific manner.

If so, state-based modulations of value predictions should also be observable at cue onset.

As can be observed in Fig. 5.6 and 5.7, this is exactly what we found. While food cues continued to elicit significantly greater dopamine release than sucrose solution cues in sucrose solution devaluation session (Fig. 5.6a), this reversed after food devaluation, with dopamine release for food cues now lower than after sucrose solution cues (Fig. 5.6b). This was borne out by a linear regression that showed a significant effect of reward type on cue-evoked dopamine in both sessions, but with the sign modulated by the identity of the pre-fed reinforcer (Fig. 5.6c, d).

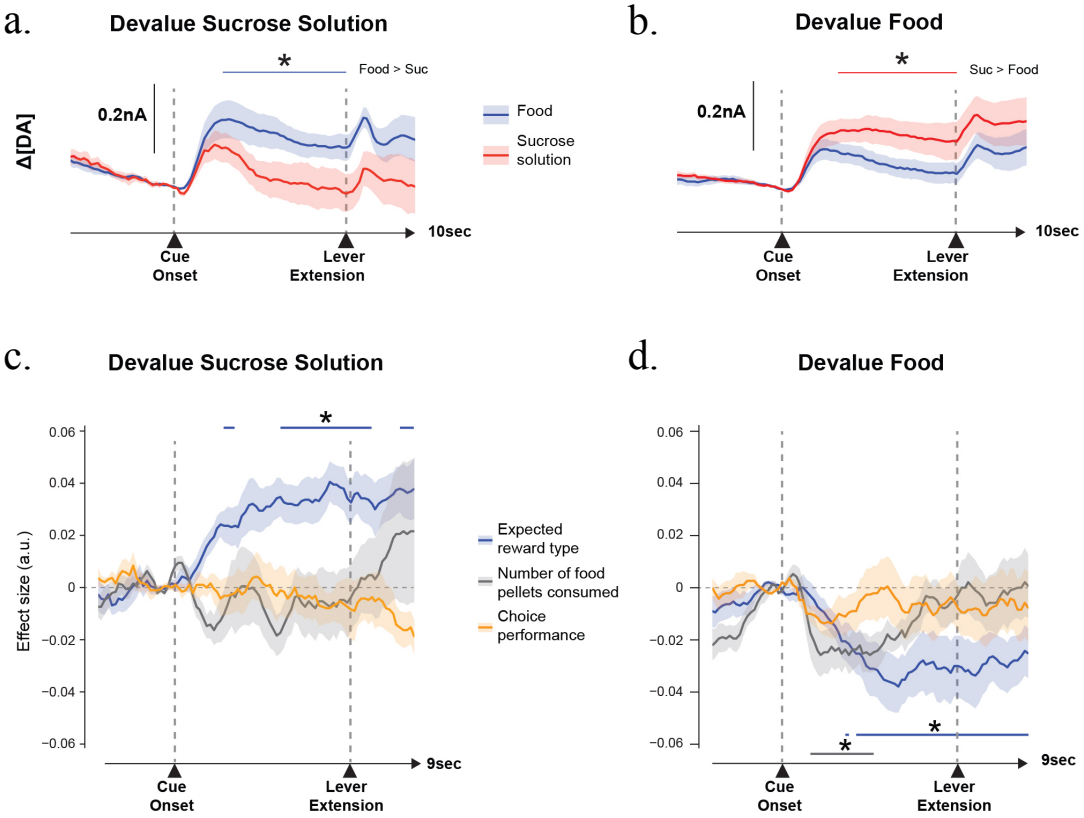


Figure 5.6: Dopamine release during Devaluation sessions
Average dopamine signals (**a**, **b**) or effect sizes (**c**, **d**) from a general linear model of cue-evoked dopamine after selective satiation on sucrose solution (**a, c**) or food (**b, d**). (Lines: *, $p < 0.05$ permutation tests, corrected for multiple comparisons. All averages, mean ± SEM).

Moreover, these effects occurred rapidly, being evident within the first block of the session (analysis of valued or devalued cue dopamine: $F_{1,6} = 11.93$, $p = 0.014$; no main effect or interaction with reward type: both $F < 1.39$, $p > 0.28$; $n = 7$ as one animal was excluded for having $\leq 5\%$ responses on the devalued option) (Fig. 5.7a). Interestingly, while selective satiation uniformly decreased dopamine after devalued reward cues, there was an asymmetric effect valued cue-elicited dopamine (Fig. 5.7b). Specifically, after food devaluation, dopamine levels were on average significantly greater after sucrose solution cues compared to baseline sessions ($p < 0.05$). By contrast, there was no statistically reliable change in either direction in response to valued food cues after sucrose solution devaluation. In other words, after eating to satiety, the predicted value of the sweet liquid option increased. Nonetheless, while the selective satiety procedures reliably biased choice behaviour and modulated dopamine release, there was again no consistent relationship between the size of cue-elicited signals in a particular session for a particular animal and its preference for one reinforcer over the other on choice trials (Fig. 5.6c-d, 5.7a, c).

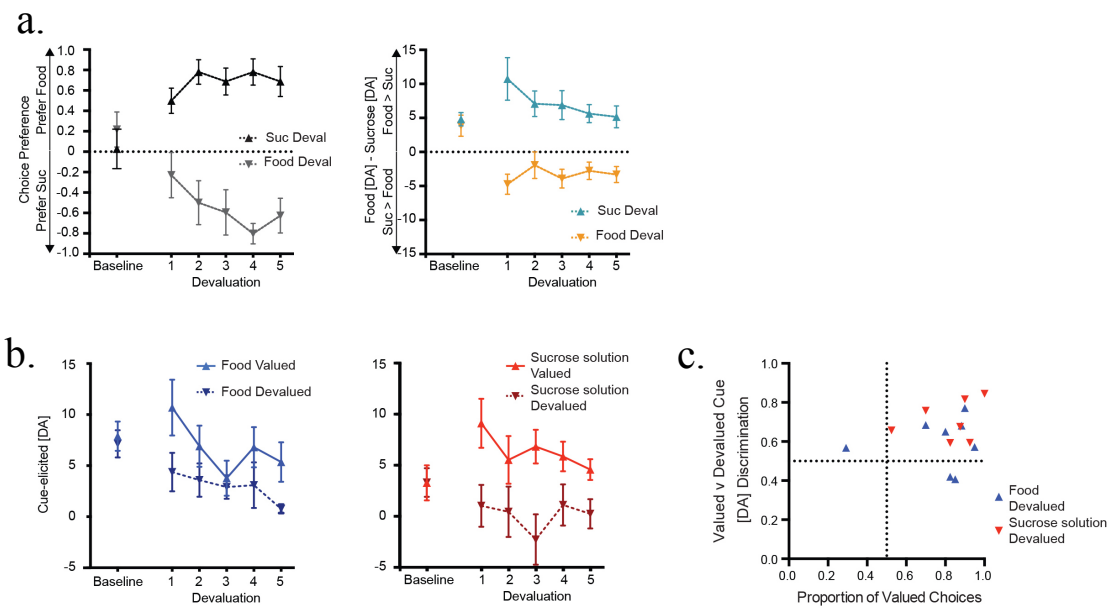


Figure 5.7: Dynamic changes in cue-evoked dopamine after selective satiation
(a) Change in preference and relative cue-evoked dopamine plotted over 5 bins each of 2 blocks of trials following sucrose solution or food devaluation. Choice in expressed as a change from 50% [(2 x Proportion of food choices) – 1]. Relative cue-evoked dopamine is the difference between average food cue and sucrose solution cue dopamine levels during the 5s post-cue period (between cue onset and lever extension). The average difference across the whole of the immediately preceding Baseline session is presented for comparison. **(b)** Average dopamine in the 5s post-cue period after food or sucrose solution cues, divided up into 5 bins, based on whether that reinforcer or the other was devalued in that session. The average difference across the whole of the immediately preceding Baseline session is presented for comparison. Difference in average post-cue dopamine release for each reinforcer in a valued or devalued state plotted over 5 bins each of 2 blocks of trials. Average difference in dopamine in the immediately preceding Baseline session is presented for comparison. **(c)** Dopamine discrimination index for each animal in the 5s post cue period for both Valued Food versus Devalued Sucrose Solution (red triangles) and for Value Sucrose Solution versus Devalued Food (blue triangles), plotted against each animal's choices of the valued reinforcer in that session. There was no reliable relationship between these measures ($r = 0.183$, $p = 0.51$).

Importantly, although selective satiation strongly modulated dopamine levels in the devaluation session, this did not have a lasting influence over patterns of dopamine release. The first presentation of the previously devalued option in Baseline B immediately elicited comparable levels of dopamine as when that same cue had been presented during the first Baseline A session (comparison between 1st trial dopamine in Baseline A and Baseline B, separated by which option was devalued during Devaluation A: all $F < 0.96$, $p > 0.36$) (Fig. 5.8a).

While these analyses show a rapid influence of selective satiety on dopamine release, it is not clear if this is purely an experience-dependent effect based on learning the value of the options in the new state or whether dopamine cue signals can update even before consuming the devalued reward during the session. To examine this, we analysed dopamine release elicited by the first presentation in the session of both the valued and devalued options (Fig. 5.8b).

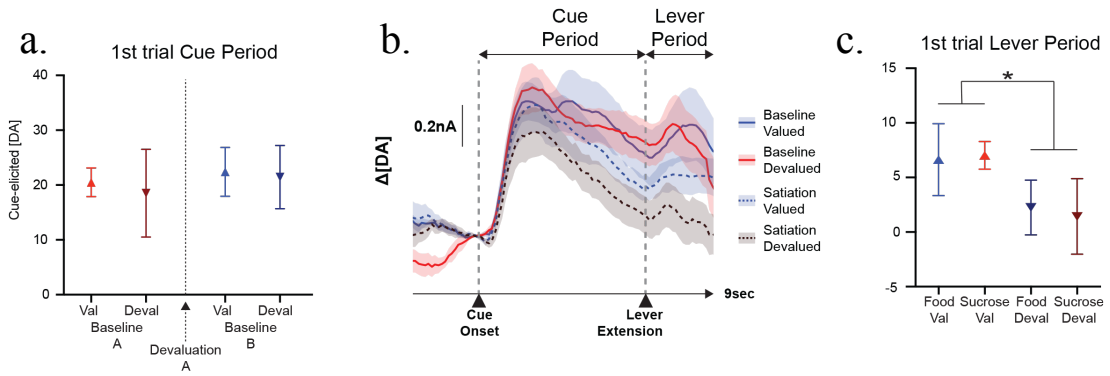


Figure 5.8: Cue-elicited dopamine release on 1st trials of the session
(a) Average dopamine on the first food / sucrose solution trial in a 5s post-cue window in Baseline A and B. Data are divided up by which reinforcer was devalued in the first Devaluation session. **(b)** Average dopamine signals after cue onset on the first food / sucrose solution trial in the Baseline and Devaluation sessions. Baseline data are here divided up based on the reinforcer that would be devalued in the *subsequent* devaluation session. **(c)** Average dopamine on the first food or sucrose solution trial in a 2s post-lever extension window prior to reward delivery, averaged across the Devaluation sessions. Levels were significantly reduced on the 1st devalued trial compared to the 1st valued trial (*, $p < 0.05$, ANOVA).

This revealed an overall attenuation of dopamine release on the initial trial of the Devaluation sessions compared to the preceding Baseline session (main effect of Session type: $F_{1,6} = 12.44$, $p = 0.012$, including trial order as a between-subject factors). However, this reduction was not significantly greater after devalued cues than valued cues (interactions including Session type x Devaluation: all $F < 0.27$, $p > 0.62$). This implies that selective satiety induces an immediate general, rather than stimulus-specific, reduction in cue-elicited dopamine signals, but that rats need

experience of the reinforcer in the new state to fully update learned cue associations. Nonetheless, inspection of Fig. 5.4b suggests that dopamine levels were on average lower on the first *devalued* trial compared to cue presentations in valued states, particularly in the later period between lever extension and reward delivery. We therefore also directly compared average dopamine levels during the lever extension/response period, prior to reward delivery, on the first valued and devalued trial. This confirmed that dopamine levels were significantly attenuated after presentation of the devalued lever compared to after the valued lever even though the reinforcers had yet to be directly experienced in the new state (main effect of Devaluation: $F_{1,6} = 12.63$, $p = 0.012$) (Fig. 5.8c).

5.4 Discussion

These results demonstrate that mesolimbic dopamine flexibly encodes reward prediction error signals shaped by the specific properties of a reward to satisfy a current need. Midbrain dopamine neurons in primates tested for multiple days in a similar state have been shown to encode reward prediction errors that reflect the animals' subjective preference for different reward type (Lak et al., 2014). Here we observed a rapid, experience-driven updating of NAc core dopamine signals, both to predictive cues and reward delivery, to reflect the subjective value of stimuli following selective satiation.

In combination with the results of Chapter 4, these results appear consistent with key predictions of model-free temporal difference learning models that dopamine firing rates or release will not encode surprising changes in reward identity independent of reward value, if differences in state-based value are accounted for

(Niv & Schoenbaum, 2008). This does not rule out that coding of reward identity prediction errors may exist in other contexts where the value difference between the options are less prominent or when a change in identity is more relevant for behaviour. For instance, the SWITCH trials here occurred as rare fluctuations in an otherwise stable task, but in other paradigms, such as unblocking or reversal learning, a change in reward identity can be more long lasting and of more significance for behaviour (Huetteroth et al., 2015; McDannald et al., 2014). Equally, it is possible that distinct dopamine pathways may contain additional information about reward identity or other aspects of reward. The current data were collected from the NAc core as dopamine release in this structure has been shown to signal discrepancies from expectation (Day et al., 2007; Hart et al., 2014). However, in rodents, the NAc shell rather than the core, and specifically the D1-receptor expressing medium spiny neurons in this region, have been associated with the ability of specific reinforcers to motivate and invigorate responding and promote feeding (Corbit et al., 2014; Laurent et al., 2014; O'Connor et al., 2015).

There was a strong modulation of both food- and sucrose solution-elicited dopamine by the amount of reinforcer consumed within and prior to the session. Such selective modulation of stored value signals by specific satiety may be important to promote efficient and varied foraging behaviours. While cue-elicited dopamine release updated within the first block of trials following selective satiation, this appeared to be predominantly shaped by direct incentive learning as there was no selective change in cue-evoked dopamine – at least prior to lever presentation – when the devalued option was first presented. Nonetheless, there was an overall reduction in cue-elicited dopamine after selective satiation that was evident on the first trials on the session. This is in line with previous studies showing a general activating role for

the NAc core, and dopamine transmission in this region, in the presence of reward-associated cues to motivate and invigorate available actions (Corbit et al., 2011; du Hoffmann & Nicola, 2014).

Experiments in rats with NAc lesions during devaluation tasks provide interesting insights about the role of this region in using outcome information to guide behaviour. Excitotoxic cell body lesions of the NAc core cause disruptions of the general ability of a cue to motivate responding, but leave intact the specific memory of an association between a cue and a particular reward to bias behaviour towards actions that gave that reward (Corbit et al., 2011). Similarly, after a dopamine-specific lesion of the NAc core (which leaves cells bodies intact), rats were found to be just as sensitive as sham-lesioned animals to outcome devaluation in an instrumental setting. They did, however, show an overall decrease in their responding and also showed non-selective reductions in responding in a Pavlovian devaluation paradigm (Lex & Hauber, 2010). However, during an earlier experiment by Corbit et al., (2001) the opposite result was found while rats had to execute an instrumental task that demanded increased workload (RR-20) after selective devaluation. On this occasion, rats did not show any selectively satiation effect although they overall reduced their responding towards both available options. This evidence may suggest that dopamine function plays a substantial role during tasks with high but not low workload demands.

However, it is notable that there were already selective differences in dopamine levels in the period after lever extension while the rat was making a response and waiting for either the valued or devalued reward. Some aspects of dopamine signalling can thus partially update without direct experience of the outcome (Bromberg-Martin et al., 2010). Moreover, cue-elicited dopamine returned to

pre-Devaluation patterns by the start of the subsequent Baseline session run in a food-restricted state, in spite of the large difference between release elicited by the valued and devalued cues at the end of the devaluation session. This implies that mesolimbic dopamine systems have access to stored memories of learned incentive values when returning to a familiar state.

These data add to evidence indicating a close link between mesolimbic dopamine, physiological state and metabolic processing (de Araujo et al., 2012; McCutcheon, 2015; Sclafani et al., 2011). In contrast to the lack of consistent relationship between dopamine release and choice behaviour, we observed a strong modulation of both food- and sucrose solution-elicited dopamine by the amount of reinforcer consumed within and prior to the session. Ventral tegmental area dopamine neurons receive excitatory inputs from the lateral hypothalamus (Watabe-Uchida et al., 2012) and dopamine cell activity and NAc core dopamine release is influenced by physiological state (Bassareo et al., 1999; Branch et al., 2013) and by peptides involved in appetite (Cone et al., 2014). To determine the influence of motivational state, and peptides known to influence motivational state, on phasic dopamine release, Cone et al., (2014) recorded dopamine release from the NAc core while central receptors for the hormone ghrelin were pharmacologically manipulated. Ghrelin is a peptide hormone known for stimulating appetite and inducing a positive energy balance leading to weight gain (Inui et al., 2004). Ghrelin receptors can be found throughout the brain, including the hypothalamus and the VTA (Zigman et al., 2006). Rats with either *ad libitum* or restricted food access received unsignalled delivery of food pellets while their NAc dopamine release was recorded. Infusion of ghrelin to the lateral ventricles increased dopamine release in ad libitum rats while a ghrelin antagonist suppressed dopamine release in food-restricted animals.

In contrast with the passive feeding tasks mentioned in the Introduction, and taking into account the fact that tonic dopamine plays a key role in effort-based decision-making, a study by Ostlund et al., (2011) investigated the role of dopamine during instrumental conditioning. During 3 testing sessions rats had to press a lever for earning pellets (RR-10). All three sessions were identical with the only difference that during the first and third session rats were hungry whereas in the second session rats were sated. During the food-deprived sessions an increase in NAc core and shell dopamine release was observed, which was attenuated in the sated session. Rats with NAc core placements also showed a decreased dopamine release as the workload requirements increased. Moreover, the specific *type* of reinforcer, and not just its general properties (i.e., food to satisfy hunger), influence dopamine levels. A study from Ahn and Phillips (1999) demonstrated increased NAc dopamine release during the anticipation and presentation of a second meal consisting of a different food to one consumed earlier. By contrast, presentation of the same food type again had no reliable influence on dopamine levels.

In the current experiment, decisions will be made based not only on the respective taste and palatability of a food pellet versus a bolus of sucrose solution but also on their nutritional value in a given state. The rapid adaptation of mesolimbic dopamine signals following a change in state would potentially allow it to play an important role in prioritising behaviours based on the available opportunities.

6 General Discussion

In this thesis, I presented a series of reward-guided decision-making experiments in non-human primates (macaque monkeys) and rodents (rats). For this purpose, I have used two different brain-recording techniques: functional magnetic resonance imaging (fMRI) and fast-scan cyclic voltammetry (FSCV).

In my experiments with macaque monkeys (Chapters 2 and 3) I investigated a well-known behavioural effect, that I have referred to as the 'less is more' effect, in which monkeys had to choose between single and compound options. Interestingly control monkeys exhibited the same pattern of behaviour that humans do when performing similar tasks (Hsee, et al., 2001; Kahneman, 2011; Kahneman & Tversky, 1979; List, 2002). More precisely, instead of choosing the more valuable option that consisted of two rewards, a highly valued and a lower, but still positively valued reward, they repeatedly chose only the single highly valued option. It appears that the representation of the two rewards associated with the compound option is biased away from the sum of the values of the two constituent parts and towards the options' mean values. As a result the animals fail to maximise the number of rewards that they might receive in a session. On first inspection the pattern of behaviour appears irrational although it was noted that it can be adaptive for a foraging animal to disregard food items that entail a handling cost, in terms of time and effort, that disrupts their ability to pursue other much better opportunities (Pearson et al., 2014; Stephens & Krebs, 1986).

Even more strikingly, when the same task was given to monkeys with vmPFC/mOFC lesions, as described in detail in Chapter 2, they exhibited a more rational choice pattern by choosing the compound option over the single highly valued option more often than the control group. Devaluation of the low valued

reward that led to a larger value discrepancy between the values of the component rewards that constituted the compound option resulted in the vmPFC/mOFC animals becoming even more rational, or at least indifferent, in their choices. In contrast, the controls became heavily suboptimal exhibiting even stronger preference for the single highly valued option over the compound option. However, devaluation of the highly valued reward resulted in the controls becoming optimal in their decisions while the vmPFC/mOFC animals remained relatively unchanged in their choices.

Based on these intriguing results, we devised an fMRI version of the same paradigm with the control animals, in order to identify the neural origins of this irrational decision-making pattern. We paid particular attention to activity in the vmPFC/mOFC when an animal engaged in the types of decisions that were most affected by the lesion. We focused our analyses at the time that the cues were presented to the monkeys since at that time value estimations associated with the two potential choices must be compared (Bouret & Richmond, 2010; Litt et al., 2011; Tom et al., 2007). Whole brain analyses revealed that at the time of the cue presentation, strong vmPFC/mOFC activity is present. In particular, a strong cluster of brain activity was observed within the approximate region in which the lesion had been placed in Chapter 2. The peak of that signal was chosen as the region of interest (ROI) for all further analyses. Throughout the analyses the critical regions appeared to be close to the medial orbital sulcus area highlighted by Padoa-Schioppa and colleagues (Cai & Padoa-Schioppa, 2014; Padoa-Schioppa & Assad, 2006) but extending into the adjacent gyrus rectus.

Whole brain comparisons between different choice trials, in which two reward-associated CS+ stimuli were offered to the animals, revealed that the strongest activity in vmPFC/mOFC was observed when the animals had to choose between the

highly valued single element stimulus and the compound stimulus. This activity was significantly higher than when the animal had to choose between the other two CS+ pairs. Again, the activity difference was observed within the lesion area. We also found evidence for value signals within the vmPFC/mOFC. The value difference between the chosen and the unchosen option was associated with strong deactivation in the vmPFC/mOFC. A timecourse signal extracted from the peak of the largest vmPFC/mOFC cluster, revealed an activity pattern that incorporates the overall value of each choice, regardless of whether an option was chosen or unchosen. Another value-related signal was found to account for the stimulus value. A pattern of activity consistent with value transitivity pattern was also observed; vmPFC/mOFC activity approximately covaried with the overall value sum associated with each different CS+ pairs that the monkeys chose between. Additional analyses demonstrated other attributes of the vmPFC/mOFC role in decision-making such as the fact that, unlike in human subjects, vmPFC/mOFC was positively rather than negatively correlated with decision difficulty (Hare et al., 2011; Lebreton et al., 2015). This suggests that interpretations of brain function that have focused on whether or not an area's activity is positively or negatively correlated with decision-difficulty may need revision (Kolling et al., 2016; Rushworth et al., 2012).

In Chapters 4 and 5, I presented data on reward-guided learning, decision-making, and dopamine function in rats. As already noted, I used FSCV for recording dopamine activity from the nucleus accumbens (NAc) core. The phasic activity of dopamine neurons signal predictions of expected rewards and discrepancies between the predicted and received rewards that can then be used to guide learning and other aspects of behaviour (Corbit et al., 2011; Gan et al., 2010; Hart et al., 2014; Montague et a., 1996; Schultz et al., 1997; Syed et al., 2016). The NAc and the vmPFC/mOFC

are interconnected in the primate brain (Haber et al., 2006) and have related roles in reward-guided learning and decision-making (Strait et al., 2014; Strait et al., 2015). However, the granular cortex present in macaque and human vmPFC/mOFC is only present in primates and research on dopamine modulation of neurons in the PFC has led to contradictions and controversies (Seamans & Young, 2004). As in the experiments in Chapters 2 and 3, I used different types of rewards during my investigations of rats. An important difference, however, was that my different reinforcers were titrated so as the animals could show, on average, equal preference towards two rewards used in the same testing session while my aim in the monkey experiments had been to investigate reward preferences that are related to reward type. Having two reinforcers that were of equal preference, and consequently of equal subjective value, allowed me to investigate whether dopamine encodes the identity of the rewards. Reward prediction errors at the time of the reward delivery under certain circumstances could demonstrate such coding.

For this purpose, I have used a simple two-choice two-reward instrumental decision-making task in which rats had to press levers associated with the delivery of two distinct reinforcers: sucrose solution and standard grain-based pellets. On a subset of trials each of the levers could lead either to a larger amount of reward or to the reward contingent with the opposite lever - a reward with a different identity but of equal value. I was able to show that dopamine neurons encoded the difference in value between prior expectation and actual reward. Most interestingly, however, I was able to show that when a reward of different identity to the expected reward was delivered, dopamine neurons exhibited a significant phasic increase. The effect was observed only when the rats were pressing the lever associated with sucrose delivery but instead a grain-based pellet was delivered. Analyses at the time of the cue onset

revealed that the difference in dopamine release between the two reinforcers was evident even from that time point. Surprisingly, however, at the same time, the rats were choosing the two reinforcers equally even though we already know that dopamine signal scales with expected reward (Gan et al., 2010; Hollon et al., 2014; Syed et al., 2016). We assumed that the difference in the observed dopamine signals could be related to the different nutritional value of the two reinforcers as well as to the fact that rats entered the task in a food deprived state but without any prior water restriction.

To investigate this possibility, we ran an additional experiment, described in Chapter 5, in which a subset from the same group of rats executed the same task but right after devaluing each of the reinforcers in separate sessions. We found that the unidirectional identity surprise signals that had previously occurred in response to unanticipated grain pellets were no longer evident and instead only value reward prediction errors were observed. These results demonstrate that mesolimbic dopamine flexibly encodes reward prediction error signals in a state-dependent frame of reference. They further add to evidence indicating a close link between mesolimbic dopamine, physiological state and metabolic processing (de Araujo et al., 2012; McCutcheon, 2015; Sclafani et al., 2011).

In all of my experiments the question of what is the best way to make a decision was at stake. It seems that the nature of vmPFC/mOFC value signal, under the conditions investigated in the compound option decision-making task, lead animals to make decisions that appeared suboptimal from many perspectives. The monkeys failed to sum the value of the two component rewards of the Compound option and instead it appeared that their reward expectation for the compound cue was biased towards the average value of the two rewards. They made the correct

estimations only when the value difference between the two rewards of the Compound option were closer in terms of value, for example, after devaluation of the highly valued reward. The dopamine neurons of the NAc however, demonstrated strong signalling towards the 'correct direction' and the more nutritious option but the rats failed to follow what actually the dopamine neurons were pointing to demonstrating a strong discrepancy between dopamine neural activity and behaviour. Similarly, rats managed to overcome their 'suboptimality' after selective devaluation sessions. For both species there was strong NAc activity when the choice (or two-option) cues were presented. In rats, there was a stronger dopamine signal in NAc on one-option trials on which only the food option was available (so no decision had to be made) and the two option decision-making trials on which food was chosen. Similarly in Chapter 3, I noted that in monkeys a stronger BOLD signal was observed on one-option trials compared to two-option decision trials. Therefore it is possible that the presence of an alternative option suppresses reward related activity.

In future experiments it would be very interesting to run additional voltammetry experiments with rats in a water deprived state to see if a reward type prediction error signal might be seen in response to unexpected fluid rewards if these rewards had a higher nutritional value to the animal given its physiological state. I would anticipate that the results would look analogous to those reported in Chapter 4 but nevertheless the interpretation of the effect that I have given would become particularly compelling if bidirectional evidence for reward type prediction errors were found. Also, it would be quite exciting to run the same macaque fMRI experiment that I reported in Chapter 3 but after one of the control monkeys' rewards had been selectively devalued. I hypothesise that the brain signals during the HV versus CV choice would scale relative to the devaluation type. Moreover, other

interesting experiments could be based on running the same experiments presented on this thesis but for both species. Future research could further aim to bridge the gap between animal and human neuroscience by using modified versions of the tasks in human subjects.

Bibliography

Abitbol, R., Lebreton, M., Hollard, G., Richmond, B. J., Bouret, S., & Pessiglione, M. (2015). Neural mechanisms underlying contextual dependency of subjective values: Converging evidence from macaques and humans. *J Neurosci 35*, 2308-2320.

Ahn, S., & Phillips, A.G. (1999). Dopaminergic correlates of sensory-specific satiety in the medial prefrontal cortex and nucleus accumbens of the rat. *J Neurosci 19*, 21-26.

Anand, P. (1993). *Foundations of Rational Choice Under Risk*. Oxford, Oxford University Press.

Asp, E., Manzel, K., Koestner, B., Cole, C.A., Denburg, N. L., & Tranel, D. (2012). A neuropsychological test of belief and doubt: damage to ventromedial prefrontal cortex increases credulity for misleading advertising. *Front Neurosci 6*, 1-9.

Baliki, M.N., Mansour, A., Baria, A.T., Huang, L., Berger, S.E., Fields, H.L., & Apkarian, A.V. (2013). Parceling human accumbens into putative core and shell dissociates encoding of values for reward and pain. *J Neurosci 33,* 16383-16393.

Balleine, B., Daw, N., & O'Doherty, J. (2008). Multiple forms of value learning and the function of dopamine. In: *Neuroeconomics*. United Kingdom: Academic Press, 367-387.

Balleine, B.W., & Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology 37*, 407-419.

Balleine, B.W., Leung, B.K., & Ostlund, S.B. (2010). The orbitofrontal cortex, predicted value, and choice. *Ann N Y Acad Sci 1239*, 43-50.

Barron, H.C., Dolan, R.J., & Behrens, T.E. (2013). Online evaluation of novel

choices by simultaneous representation of multiple memories. *Nat Neurosci 16*, 1492-1498.

Barron, H.C., Garvert, M.M., & Behrens, T.E.J. (2015). Reassessing VMPFC: full of confidence? *Nat Neurosci 18*, 1064-1066.

Bassareo, V., & Di Chiara, G. (1999). Modulation of feeding-induced activation of mesolimbic dopamine transmission by appetitive stimuli and its relation to motivational state. *Eur J Neurosci 11*, 4389-4397.

Basten, U., Biele, G., Heekeren, H.R., & Fiebach, C.J. (2010). How the brain integrates costs and benefits during decision making. *Proc Natl Acad Sci U S A 107*, 21767-21772.

Baxter, M.G., Parker, A., Lindner, C.C.C., Izquierdo, A.D., & Murray, E.A. (2000). Control of response selection by reinforcer value requires interaction of amygdala and orbital frontal cortex. *J Neurosci 20*, 4311-4319.

Bechara, A., Tranel, D., & Damasio, H. (2000). Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain 123*, 2189-2202.

Beeler, J.A., McCutcheon, J.E., Cao, Z.F., Murakami, M., Alexander, E., Roitman, M.F., & Zhuang, X. (2012). Taste uncoupled from nutrition fails to sustain the reinforcing properties of food. *Eur J Neurosci 36*, 2533-2546.

Beran, M.J., Ratliff, C.L., & Evans, T.A. (2009) Natural choice in chimpanzees (Pan troglodytes): Perceptual and temporal effects on selective value. *Learn Motiv 40*, 186-196.

Berridge, K.C. (1991). Modulation of taste affect by hunger, caloric satiety, and sensory-specific satiety in the rat. *Appetite 16*, 103-120.

Bilder, R.M., Volavkan J., Lachman, H.M., & Grace, A. (2004). The catechol-O-

methyl-transferase polymorphism: Relations to the tonic-phasic dopamine hypothesis and neuropsychiatric phenotypes. *Neuropsychopharmacology 29*, 1943-1961.

Bayer, H.M., & Glimcher, P.W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron 47*, 129-141.

Blanchard, T.C., Pearson, J.M., & Hayden, B.Y. (2013). Postreward delays and systematic biases in measures of animal temporal discounting. *Proc Natl Acad Sci U S A 110*, 15491-15496.

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

Bossaerts, P., Preuschoff, K., & Hsu, M. (2008). The Neurobiological Foundations of Valuation in Human Decision Making under Uncertainty. In *Neuroeconomics: Decision Making and the Brain*, edited by Glimcher, P.W., Camerer, C.F., Fehr, E., & Poldrack, R.A. New York: Academic Press, 353-365.

Botvinick, M.M. & An, J. (2008) Goal-directed decision making in prefrontal cortex: A computational framework. *Adv Neural Inf Process Syst 21*,169-176.

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

Branch, S.Y., Goertz, R.B., Sharpe, A.L., Pierce, J., Roy, S., Ko, D., Paladini, C.A., & Beckstead, M.J. (2013). Food restriction increases glutamate receptor-mediated burst firing of dopamine neurons. *J Neurosci 33*, 13861-13872.

Bromberg-Martin, E.S., Matsumoto, M., Hong, S., & Hikosaka, O. (2010). A pallidus-habenula-dopamine pathway signals inferred stimulus values. *J*

Neurophysiol 104, 1068-1076.

Burke, K.A., Franz, T.M., Miller, D.N., & Schoenbaum, G. (2008). The role of orbitofrontal cortex in the pursuit of happiness and more specific rewards. *Nature 454*, 340-344.

Butter C.M. (1969). Perseveration in extinction and in discrimination reversal tasks following selective frontal ablations in Macaca mulatta. *Physiol Behav 4*, 163-171.

Cai, X., & Padoa-Schioppa, C. (2014). Contributions of orbitofrontal and lateral prefrontal cortices to economic choice and the good-to-action transformation. *Neuron 81*, 1140-1151.

Camille, N., Griffiths, C.A., Vo, K., Fellows, L.K., & Kable, J.W. (2011). Ventromedial frontal lobe damage disrupts value maximization in humans. *J Neurosci 31,* 7527-7532.

Caplin, A., & Dean, M. (2007). Dopamine, reward prediction error, and economics. *Quart J Econ 123*, 663-701.

Cardinal, R.N., Pennicott, C., Sugathapala, C.L., Robbins, T.W., & Everitt, B.J. (2001). Impulsive choice induced in rats by lesions of the nucleus accumbens core. *Science 292*, 2499-2501.

Carmichael, S.T., & Price, J.L. (1994). Architectonic subdivision of the orbital and medial prefrontal cortex in the macaque monkey. *J Comp Neurol. 346*, 366-402.

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. *Nat Neurosci 17*, 463-470.

Clarke, H.F., Robbins, T.W. & Roberts, A.C. (2008). Lesions of the medial striatum in monkeys produce perseverative impairments during reversal learning similar to those produced by lesions of the orbitofrontal cortex. *J Neurosci 28*, 10972-10982.

Cohen, J., Haesler, S., Vong, L., Lowell, B.B., & Ushida, N. (2012). Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature 482*, 85-88.

Collins, A. & Koechlin, E. (2012). Reasoning, learning, and creativity: frontal lobe function and human decision making. *PLoS Biol 10*, e1001293.

Cone, J.J., McCutcheon, J.E., & Roitman, M.F. (2014). Ghrelin acts as an interface between physiological state and phasic dopamine signaling. *J Neurosci 34*, 4905-4913.

Corbit, L.H., & Balleine, B.W. (2011). The general and outcome-specific forms of Pavlovian-instrumental transfer are differentially mediated by the nucleus accumbens core and shell. *J Neurosci 31*, 11786-11794.

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

Dahlstrom, A. & Fuxe, K. (1965). Evidence for the existence of an outflow of noradrenaline nerve fibres in the ventral roots of the rat spinal cord. *Experientia 21***,** 409-10.

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*, 1102-1105.

Daw, N.D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioural control. *Nat Neurosci 8*, 1704-1711.

Day, J.J., Jones, J.L., & Carelli, R.M. (2011). Nucleus accumbens neurons encode predicted and ongoing reward costs in rats. *Eur J Neurosci 33*, 308-321.

Day, J.J., Roitman, M.F., Wightman, R.M., & Carelli, R.M. (2007). Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nat Neurosci 10*, 1020-1028.

de Araujo, I.E., Ferreira, J.G., Tellez, L.A., Ren, X., & Yeckel, C.W. (2012). The gut-brain dopamine axis: a regulatory system for caloric intake. *Physiol Behav 106*, 394-399.

de Araujo, I.E., Lin, T., Veldhuizen, M.G., & Small, D.M. (2013). Metabolic regulation of brain response to food cues. *Curr Biol 23*, 878-883.

Deci, E.L. (1971). Effects of externally mediated rewards on intrinsic motivation. *J Pers Soc Psychol 18*, 105-115.

Delamater, A.R. (1995). Outcome-selective effects of intertrial reinforcement in a Pavlovian appetitive conditioning paradigm with rats. *Anim Learn Behav 23*, 31-39.

Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgment. *J Pers Soc Psychol 66*, 819-829.

Di Chiara, G. (2002). Nucleus accumbens shell and core dopamine: differential role in behaviour and addiction. *Behav Brain Res 137*, 75-114.

du Hoffmann, J., & Nicola, S.M. (2014). Dopamine invigorates reward seeking by promoting cue-evoked excitation in the nucleus accumbens. *J Neurosci 34*, 14349-14364.

Everitt, B.J., Morris, K.A., O'Brien, A., & Robbins, T.W. (1991). The basolateral amygdala-ventral striatal system and conditioned place preference: further evidence of limbic-striatal interactions underlying reward-related processes. *Neuroscience 42*, 1-18.

Fellows, L.K. (2006). Deciding how to decide: ventromedial frontal lobe damage

affects information acquisition in multi-attribute decision making. *Brain 129*, 944-952.

Fellows, L.K. (2007). The role of ventromedial prefrontal cortex in decision making: Judgement under uncertainty per se? *Cereb Cortex 17*, 2669-2674.

Fellows L. K. (2011). Orbitofrontal contributions to value-based decision making: evidence from humans with frontal lobe damage. *Ann N Y Acad Sci 1239*, 51-58.

Fellows, L.K. (2011). The Neurology of Value. In: Neurobiology of Sensation and *Reward*. Chapter 16. Boca Raton (FL): CRC Press.

Ferry, A.T., Ongur D., An, X., & Price, J.L. (2000). Prefrontal cortical projections to the striatum in macaque monkeys: evidence for an organization related to prefrontal networks. *J Comp Neurol 425*, 447-470.

Fiorillo, C.D., Tobler, P.N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science 299*, 1898-1902.

Fiorillo, C.D., Yun, S.R., Song, M.R. (2013). Diversity and homogeneity in responses of midbrain dopamine neurons. *J Neurosci 33*, 4693-4709.

Flagel S.B., Clark J.J., Robinson T.E., Mayo L., Czuj A., Willuhn I., Akers C.A., Clinton S.M., Phillips P.E.M., & Akil, H. (2011). A selective role for dopamine in stimulus-reward learning. *Nature 469*, 53-57.

Flagel, S.B., Robinson, T.E., Clark, J.J., Clinton, S.M., Watson, S.J., Seeman, P., Phillips, P.E., & Akil, H. (2010). An animal model of genetic vulnerability to behavioral disinhibition and responsiveness to reward-related cues: implications for addiction. *Neuropsychopharmacology 35*, 388-400.

Gallagher, H., & Frith, C.D. (2003). Functional imaging in theory of mind. *Trends Cogn Sci 7*, 77-83.

Gallagher, M., McMahan, R.W., & Schoenbaum, G. (1999). Orbitofrontal cortex and

representation of incentive value in associative learning. *J Neurosci 19*, 6610-6614.

Gan, J.O., Walton, M.E., & Phillips, P.E. (2010). Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. *Nat Neurosci 13*, 25-27.

Ganzach, Y. (2001). Judging risk and return of financial assets. *Organ Behav Hum Dec 83*, 353-370.

Garcia-Cabezas, M.A., Martinez-Sanchez, P., Sanchez-Gonzalez, M.A., Garzon, M. & Cavada, C. (2008). Dopamine innervation in the thalamus: monkey versus rat. *Cereb Cortex 19*, 424-434.

Garris, P.A., Christensen, J.R.C., Rebec, G.V., & Wightman, R.M. (1997). Real-time measurement of electrically evoked extracellular dopamine in the striatum of freely moving rats. *J Neurochem 68*, 152-161.

Ghods-Sharifi, S., & Floresco, S.B. (2010). Differential effects on effort discounting induced by inactivations of the nucleus accumbens core or shell. *Behav Neurosci 124*, 179-191.

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

Grace, A.A. (1991). Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: A hypothesis for the etiology of schizophrenia. *Neuroscience 41*, 1-24.

Haber, S.N. (2008). Basal Ganglia. In: *Encyclopedia of Neuroscience*, edited by Binder, M.D., Hirokawa, N. & Windhorst, U., Berlin Heidelberg: Springer-Verlag.

Haber, S.N., Kim, K.S., Mailly, P., & Calzavara, R. (2006). Reward-related cortical inputs define a large striatal region in primates that interface with associative cortical connections, providing a substrate for incentive-based learning. *The J Neurosci 26*, 8368-8376.

Haber, S.N., & Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology 35*, 4-26.

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

Hart, A.S., Rutledge, R.B., Glimcher, P.W., & Phillips, P.E.M. (2014). Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. *J Neurosci 34*, 698-704.

Hare, T.A., Schultz, W., Camerer, C.F., O'Doherty, J.P., & Rangel, A. (2011). Transformation of stimulus value signals into motor commands during simple choice. *Proc Natl Acad Sci U S A 108*, 18120-18125.

Heekeren, H.R., Wartenburger, I., Schmidt, H., Schwintowski, H.P., & Villringer, A. (2003). An fMRI study of simple ethical decision-making. *Neuroreport 14*, 1215-1219.

Heien, M.L., Johnson, M.A., & Wightman, R.M. (2004). Resolving neurotransmitters detected by fast-scan cyclic voltammetry. *Anal Chem 76*, 5697-5704.

Heien, M.L., Khan, A.S., Ariansen, J.L., Cheer, J.F., Phillips, P.E., Wassum, K.M., & Wightman, R.M. (2005). Real-time measurement of dopamine fluctuations after cocaine in the brain of behaving rats. *Proc Natl Acad Sci U S A 102*, 10023-10028.

Heien, M.L., Phillips, P.E., Stuber, G.D., Seipel, A.T., & Wightman RM. (2003). Overoxidation of carbon-fiber microelectrodes enhances dopamine adsorption and increases sensitivity. *Analyst 128*, 1413-1419.

Hernandez, G., Rajabi, H., Steward, J., Arvanitogiannis, A. & Shizgal, P. (2008). Dopamine tone increases similarly during predictable and unpredictable administration of rewarding brain stimulation at short inter-train intervals. *Behav Brain Res 188*, 227-232.

Heshmati, M., & Russo, S.J. (2015). Anhedonia and the brain reward circuitry in depression. *Curr Behav Neurosci Rep 2*, 146-153.

Hollon, N.G., Arnold, M.M., Gan, J.O., Walton, M.E., & Phillips, P.E. (2014). Dopamine-associated cached values are not sufficient as the basis for action selection. *Proc Natl Acad Sci U S A 111*, 18357-18362.

Howard, J.D., Gottfried, J.A., Tobler, P.N., & Kahnt, T. (2015). Identity-specific coding of future rewards in the human orbitofrontal cortex. *Proc Natl Acad Sci U S A 112*, 5195-5200.

Hsee, C.K. (1998). Less is better: when low-value options are valued more highly than high-value options. *J Behav Decis Mak 11*, 107-121.

Hsee, C.K. & Kunreuther, H. (2000). The affection effect in insurance decisions. *Journal of Risk and Uncertainty 20*, 141-159.

Huettel, S.A., Song, A.W., & McCarthy, G. (2009). *Functional Magnetic Resonance Imaging* (2 ed.), Massachusetts: Sinauer, ISBN 978-0-87893-286-3.

Huetteroth, W., Perisse, E., Lin, S., Klappenbach, M., Burke, C., & Waddell, S. (2015). Sweet taste and nutrient value subdivide rewarding dopaminergic neurons in Drosophila. *Curr Biol 25*, 751-758.

Hunt, L.T., Dolan, R.J., & Behrens, T.E. (2014). Hierarchical competitions subserving multi-attribute choice. *Nat Neurosci 17*, 1613-1622.

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. *Nat Neurosci 15*, 470-476.

Inui, A., Asakawa, A., Bowers, C.Y., Mantovani, G., Laviano, A., Meguid, M.M. & Fujimiya, M. (2004). Ghrelin, appetite, and gastric motility: the emerging role of the stomach as an endocrine organ. *FASEB J 18*, 439-456.

Izquierdo, A., & Murray, E.A. (2010). Functional interaction of medial mediodorsal thalamic nucleus but not nucleus accumbens with amygdala and orbital prefrontal cortex is essential for adaptive response selection after reinforcer devaluation. *J Neurosci 30*, 661-669.

Izquierdo, A., Suda, R.K., & Murray, E.A. (2004). Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *J Neurosci 24*, 7540-7548.

Izquierdo A., Suda R.K., & Murray E.A. (2005). Comparison of the effects of bilateral orbital prefrontal cortex lesions and amygdala lesions on emotional responses in rhesus monkeys. *J Neurosci 25*, 8534-8542.

Jocham, G., Furlong, P.M., Kroger, I.L., Kahn, M.C., Hunt, L.T., & Behrens, T.E. (2014). Dissociable contributions of ventromedial prefrontal and posterior parietal cortex to value-guided choice. *Neuroimage 100*, 498-506.

Jocham, G., Hunt, L.T., Near, J., & Behrens, T.E. (2012). A mechanism for value-guided choice based on the excitation-inhibition balance in prefrontal cortex. *Nat Neurosci 15*, 960-961.

Jones, J.L., Esber, G.R., McDannald, M.A., Gruber, A.J., Hernandez, A., Mirenzi, A., & Schoenbaum, G. (2012). Orbitofrontal cortex supports behavior and learning using inferred but not cached values. *Science 338*, 953-956.

Kable, W.J. & Glimcher, P.W. (2007). The neural correlates of subjective value during intertemporal choice. *Nat Neurosci 10*, 1625-1633.

Kable, J.W., & Glimcher, P.W. (2009). The neurobiology of decision: consensus and controversy. *Neuron 63*, 733-745.

Kahneman, D., (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux, New York.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under

risk. *Econometrica 47*, 263-291.

Kelley, A.E. (2004). Ventral striatal control of appetitive motivation: role in ingestive behavior and reward-related learning. *Neurosci Biobehav Rev 27*, 765-776.

Kennerley, S.W., Dahmubed, A.F., Lara, A.H., & Wallis, J.D. (2009). Neurons in the frontal lobe encode the value of multiple decision variables. *J Cogn Neurosci. 21*, 1162-1178.

Kennerley, S.W., Walton, M.E., Behrens, T.E., Buckley, M.J., & Rushworth, M.F. (2006). Optimal decision making and the anterior cingulate cortex. *Nat Neurosci 9*, 940-947.

Knutson B., Adams C. M., Fong G. W., & Hommer D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J Neurosci 21*, RC159.

Kolling, N., Behrens, T.E., Mars, R.B., & Rushworth, M.F. (2012). Neural mechanisms of foraging. *Science 336*, 95-98.

Kolling, N., Behrens, T., Wittmann, M., & Rushworth, M. (2016). Multiple signals in anterior cingulate cortex. *Curr Opin Neurobiol 37*, 36-43.

Lak, A., Stauffer, W.R., & Schultz, W. (2014). Dopamine prediction error responses integrate subjective value from different reward dimensions. *Proc Natl Acad Sci U S A 111*, 2343-2348.

Laurent, V., Bertran-Gonzalez, J., Chieng, B.C., & Balleine, B.W. (2014). Delta-opioid and dopaminergic processes in accumbens shell modulate the cholinergic control of predictive learning and choice. *J Neurosci 34*, 1358-1369.

Lebreton, M., Abitbol, R., Daunizeau, J., & Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation system. *Nat Neurosci 18*, 1159-1167.

Lebreton, M., Jorge, S., Michel, V., Thirion, B., & Pessiglione, M. (2009). An automatic valuation system in the human brain: evidence from functional Neuroimaging. *Neuron 64*, 431-439.

Levita, L., Dalley, J.W., & Robbins, T.W. (2002). Disruption of Pavlovian contextual conditioning by excitotoxic lesions of the nucleus accumbens core. *Behav Neurosci 116*, 539-52,

Lex, B., & Hauber, W. (2010). The role of nucleus accumbens dopamine in outcome encoding in instrumental and Pavlovian conditioning. *Neurobiol Learn Mem 93*, 283-290.

Li, Y., Vanni-Mercier, G., Isnard, J., Mauguière, F., & Dreher, J.C. (2016). The neural dynamics of reward value and risk coding in the human orbitofrontal cortex. *Brain 139*, 1295-1309.

List, J.A. (2002). Preference reversals of a different kind: The "more is less" phenomenon. *Am Econ Rev 92*, 1636-1643.

Litt, A., Plassmann, H., Shiv, B., & Rangel, A. (2011). Dissociating valuation and saliency signals during decision-making. *Cereb. Cortex 21*, 95-102.

Lopatina, N., McDannald, M.A., Styer, C.V., Sadacca, B.F., Cheer, J.F., & Schoenbaum, G. (2015). Lateral orbitofrontal neurons acquire responses to upshifted, downshifted, or blocked cues during unblocking. *Elife 4*, pii: e11299.

McLaren, D.G., Kosmatka, K.J., Oakes, T.R., Kroenke, C.D., Kohama, S.G., Matochik, J.A., Ingram, D.K. & Johnson, S.C. (2008). A population-average MRI-based atlas collection of the rhesus macaque. *Neuroimage 45*, 52-59.

Mackey, S., & Petrides, M. (2010). Quantitative demonstration of comparable architectonic areas within the ventromedial and lateral orbital frontal cortex in the human and the macaque monkey brains. *Eur J Neurosci 32*, 1940-1950.

Mackey, S., & Petrides, M. (2014). Architecture and morphology of the human ventromedial prefrontal cortex. *Eur J Neurosci 40*, 2777-2796.

Málková, L.E., Gaffan, D., & Murray, E.A. (1997). Excitotoxic lesions of the amygdala fail to produce impairment in visual learning for auditory secondary reinforcement but interfere with reinforcer devaluation effects in rhesus monkeys. *J Neurosci 17*, 6011-6020.

Matsumoto, M., & Hikosaka, O. (2009). Two types of dopamine neuron distinctively convey positive and negative motivational signals. *Nature 459*, 837-841.

McCutcheon, J.E. (2015). The role of dopamine in the pursuit of nutritional value. *Physiol Behav 152*, 408-415.

McDannald, M.A., Esber, G.R., Wegener, M.A., Wied, H.M., Liu, T.L., Stalnaker, T.A., Jones, J.L., Trageser, J., & Schoenbaum, G. (2014). Orbitofrontal neurons acquire responses to 'valueless' Pavlovian cues during unblocking, *ELife 3*, e02653.

McDannald, M., Lucantonio, F., Burke, K., Niv, Y., & Schoenbaum, G. (2011). Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. *J Neurosci 31*, 2700-2705.

Meredith, G.E., Agolia, R., Arts, M.P., Groenewegen, H.J., & Zahm, D.S. (1992). Morphological differences between projection neurons of the core and shell in the nucleus accumbens of the rat. *Neuroscience 50*, 149-162.

Monosov, I. E., & Hikosaka, O. (2012). Regionally distinct processing of rewards and punishments by the ventromedial prefrontal cortex. *J Neurosci 32*, 10318-10330.

Montague P. R., Dayan P., Sejnowski T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci 16*, 1936-1947.

Mogenson G.J., Jones D.L., & Yim C.Y. (1980). From motivation to action: functional interface between the limbic system and the motor system. *Prog Neurobiol 14*, 69-97.

Murphy, R.A. & Baker, A.G. (2004). A role for CS-US contingency in Pavlovian conditioning. *J Exp Psychol Anim Behav Process 30*, 229-239.

Nestler, E.J., Hyman, S.E., & Malenka, R.C (2009). *Molecular Neuropharmacology: A Foundation for Clinical Neuroscience*, 2nd Ed., McGraw-Hill, New York.

Nicola, S.M. (2010). The flexible approach hypothesis: unification of effort and cue-responding hypotheses for the role of nucleus accumbens dopamine in the activation of reward-seeking behavior. *J Neursoci 30*, 16585-16600.

Niv, Y., & Montague P.R., (2008). Theoretical and Empirical Studies of Learning. In: *Neuroeconomics,* 331-351. United Kingdom: Academic Press.

Niv, Y. & Schoenbaum, G. (2008). Dialogues on prediction errors. *Trends Cogn Sci 12*, 265-272.

Noonan, M.P., Mars, R.B., & Rushworth, M.F. (2011). Distinct Roles of Three Frontal Cortical Areas in Reward-Guided Behavior. *J Neurosci 31*, 14399-14412.

Noonan, M.P., Walton, M.E., Behrens, T.E., Sallet, J., Buckley, M.J., & Rushworth, M.F. (2010). Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proc Natl Acad Sci USA 107*, 20547-20552.

Nyberg, E.M., Tanabe, J., Honce, J.M., Krmpotich, T., Shelton, E., Hederman, J., & Berman, B.D. (2015). Morphologic changes in the mesolimbic pathway in Parkinson's disease motor subtypes. *Parkinsonism Relat Disord 21*, 536-540.

O'Doherty, J., Kringelbach, M.L., Rolls, E.T., Hornak, J., & Andrews, C. (2001).

Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat Neurosci 4*, 95-102.

O'Doherty, J.P. (2011). Contributions of the ventromedial prefrontal cortex to goal-directed action selection. *Ann N Y Acad Sci 1239*, 118-129.

Olds, J. (1958). Pleasure centers in the brain. *Sci Am 195*,105-116.

Olds, J. & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of the rat brain. *J Comp Physiol Psychol 47*, 419-427.

Ostlund, S.B. & Balleine, B.W. (2007). Orbitofrontal cortex mediates outcome encoding in Pavlovian but not instrumental conditioning. *J Neurosci 27*, 4819-4825.

Ostlund, S.B., LeBlanc, K.H., Kosheleff, A.R., Wassum, K.M., & Maidment, N.T. (2014). Phasic mesolimbic dopamine signaling encodes the facilitation of incentive motivation produced by repeated cocaine exposure. *Neuropsychopharmacology 39*, 2441-249.

O'Connor, E.C., Kremer, Y., Lefort, S., Harada, M., Pascoli, V., Rohner, C., & Luscher, C. (2015). Accumbal D1R Neurons Projecting to Lateral Hypothalamus Authorize Feeding. *Neuron 88*, 553-564.

Padoa-Schioppa, C. (2011). Neurobiology of economic choice: a good-based model. *Annu Rev Neurosci 34*, 333-359.

Padoa-Schioppa, C. & Assad, J.A. (2006). Neurons in the orbitofrontal cortex encode economic value, *Nature 441*, 223-226.

Pandya, D.N., & Yeterian, E.H. (2000). Comparison of prefrontal architecture and connections. In: *The PFC: Executive and cognitive functions*, edited by Roberts, A.C., Robins, T.W., & Weiskrantz. L., New York: Oxford University Press, 51-68.

Parkinson, J.A., Dalley, J.W., Cardinal, R.N., Bamford, A., Fehnert, B., Lachenal, G., Rudarakanchana, N., Halkerston, K.M., Robbins, T.W., & Everitt, B.J. (2002). Nucleus accumbens dopamine depletion impairs both acquisition and performance of appetitive Pavlovian approach behaviour: implications for mesoaccumbens dopamine function. *Behav Brain Res 137*, 149-163.

Paton J.J., Belova M.A., Morrison S.E., & Salzman C.D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature 439*, 865-870.

Pavlov, I.P. (1928). *Lectures on conditioned reflexes*. (Translated by Gantt, W.H.) London: Allen and Unwin.

Pearson, J.M., Watson, K.K., & Platt, M.L. (2014). Decision making: the neuroethological turn. *Neuron 82*, 950-965.

Philiastides, M.G., Biele, G., & Heekeren, H.R. (2010). A mechanistic account of value computation in the human brain. *Proc Natl Acad Sci U S A 107*, 9430-9435.

Phillips, P.E.M., Stuber, G.D., Heien, M.L.A.V., Wightman, R.M. & Carelli, R.M. (2003). Subsecond dopamine release promotes cocaine seeking. *Nature 422*, 614-618.

Pickens, C.L., Saddoris, M.P., Gallagher, M., & Holland, P.C. (2005). Orbitofrontal lesions impair use of cue-outcome associations in a devaluation task. *Behav Neurosci 119*, 317-322.

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. *J Neurosci 30*, 10799-10808.

Plassmann, H., O'Doherty, J.P. & Rangel, A. (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J Neurosci 27*, 9984-9988.

Pothuizen, H.H., Jongen-Rêlo, A.L., Feldon, J., & Yee, B.K. (2005). Double dissociation of the effects of selective nucleus accumbens core and shell lesions on impulsive-choice behaviour and salience learning in rats. *Eur J Neurosci. 22*, 2605-2616.

Ratiu P., I.F. Talos, S. Haker, S. Lieberman, P. Everett (2004). The tale of Phineas Gage, digitally remastered. *J Neurotrauma 21*, 637-643.

Ren, X., Ferreira, J.G., Zhou, L., Shammah-Lagnado, S.J., Yeckel, C.W., & de Araujo, I.E. (2010). Nutrient selection in the absence of taste receptor signaling. *J Neurosci 30*, 8012-8023.

Rescorla, R. A., & Solomon, R. L. (1967). Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning. *Psychol Rev 74*, 151-182.

Roberts A.C. (2006). Primate orbitofrontal cortex and adaptive behaviour. *Trends Cogn Sci 10*, 83-90.

Roesch, M.R., Calu, D.J., & Schoenbaum, G. (2007). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat Neurosci 10*, 1615-1624.

Robinson, D.L., & Wightman, R.M. (2007). Rapid Dopamine Release in Freely Moving Rats. In: *Electrochemical Methods for Neuroscience,* edited by Michael, A.C., & Borland, L.M., CRC Press.

Roitman, M.F., Stuber, G.D., Phillips, P.E., Wightman, R.M., & Carelli, R.M. (2004). Dopamine operates as a subsecond modulator of food seeking. *J Neurosci 24*, 1265-1271.

Roitman, M.F., Wheeler, R.A., Wightman, R.M., & Carelli, R.M. (2008). Real-time chemical responses in the nucleus accumbens differentiate rewarding and aversive

stimuli. *Nat Neurosci 11*, 1376-1377.

Rolls, E.T., Murzi, E., Yaxley, S., Thorpe, S.J., & Simpson, S.J. (1986). Sensory-specific satiety: food-specific reduction in responsiveness of ventral forebrallin neurons after feeding in the monkey. *Brain Res 368*, 79-86.

Rudebeck, P.H., Behrens, T.E., Kennerley, S.W., Baxter, M.G., Buckley, M.J., Walton, M.E., & Rushworth, M.F. (2008). Frontal cortex subregions play distinct roles in choices between actions and stimuli. *J Neurosci 28*, 13775-13785.

Rudebeck P.H., & Murray E.A. (2008). Amygdala and orbitofrontal cortex lesions differentially influence choices during object reversal learning. *J Neurosci 28*, 8338-8343.

Rudebeck, P.H., & Murray, E.A. (2014). The orbitofrontal oracle: cortical mechanisms for the prediction and evaluation of specific behavioural outcomes. *Neuron 84*, 1143-1156.

Rudebeck, P.H., & Murray, E.A. (2011). Dissociable effects of subtotal lesions within the macaque orbital prefrontal cortex on reward-guided behavior. *J Neurosci 31*, 10569-10578.

Rudebeck, P.H., Saunders, R.C., Prescott, A.T., Chau, L.S., & Murray, E.A. (2013). Prefrontal mechanisms of emotion, value and behavioural flexibility. *Nat Neurosci 16*, 1140-1145.

Rudebeck, P.H., Walton, M.E., Smyth, A.N., Bannerman, D.M., & Rushworth, M.F. (2006). Separate neural pathways process different costs. *Nat Neurosci 9*, 1161-1168.

Rudorf, S., & Hare, T.E. (2014). Interactions between dorsolateral and ventromedial prefrontal cortex underlie context-dependent stimulus valuation in goal-directed choice. *J Neurosci 34*, 15988-15996.

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? *Curr Opin Neurobiol 22*, 946-955.

Rushworth M.F.S., Noonan M.P., Boorman, E.D., Walton, M.E., & Behrens T.E. (2011). Frontal cortex and reward-guided learning and decision-making. *Neuron 70*, 1054-1069.

Saddoris, M.P., Cacciapaglia, F., Wightman, R.M., & Carelli, R.M. (2015). Differential dopamine release dynamics in the nucleus accumbens core and shell reveal complementary signals for error prediction and incentive motivation. *J Neurosci 35*, 11572-11582.

Saleem, K.S., & Logothetis, N.K. (2006). *A Combined MRI and Histology Atlas of the Rhesus Monkey Brain*. Amsterdam: Academic Press.

Samuelson, P.A. (1937). A note on measurement of utility. *Rev Econ Stud 4*, 155-161.

Schoenbaum, G., Chiba, A. A. & Gallagher, M. (1998). Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nat Neurosci 1*, 155-159.

Schoenbuam, G., & Roesch, M. (2005). Orbitofrontal cortex, associative learning, and expectancies. *Neuron 47*, 633-636.

Schoenbaum, G., Roesch, M. R., Stalnaker, T.A., Takahashi, Y. K. (2009). A new perspective on the role of the orbitofrontal cortex in adaptive behavior. *Nat Rev Neurosci 10*, 885-892.

Schoenbaum, G., Saddoris, M.P., & Stalnaker, T.A. (2007). Reconciling the roles of orbitofrontal cortex in reversal learning and the encoding of outcome expectancies. *Ann. N Y Acad. Sci. 1121*, 320-335.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *J Neurophysiol 80*, 1-27.

Schultz W. (2006). Behavioral theories and the neurophysiology of reward. *Annu Rev Psychol 57*, 87-11.

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

Sclafani, A., Touzani, K., & Bodnar, R.J. (2011). Dopamine and learned food preferences. *Physiol Behav 104*, 64-68.

Seamans, J.K., & Yang, C.R. (2004). The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Prog Neurobiol 74*, 1-58.

Shirayama ,Y., & Chaki, S. (2006). Neurochemistry of the nucleus accumbens and its relevance to depression and antidepressant action in rodents. *Curr Neuropharmacol 4*, 277-291.

Singh, T., McDannald, M.A., Haney, R.Z., Cerri, D.H., & Schoenbaum, G. (2010). Nucleus accumbens core and shell are necessary for reinforcer devaluation effects on pavlovian conditioned responding. *Front Integr Neurosci 4*, 126.

Smith, D.V., Clithero, J.A., Boltuck, S.E., & Huettel, S.A. (2014a). Functional connectivity with ventromedial prefrontal cortex reflects subjective value for social rewards. *Soc Cogn Affect Neurosci 9*, 2017-2025.

Stalnaker, T.A., Calhoon, G.G., Ogawa, M., Roesch, M.R., & Schoenbaum, G. (2012). Reward prediction error signaling in dorsomedial striatum is action specific. *J Neurosci 32*, 10296-10305.

Stalnaker, T.A., Cooch, N.K., & Schoenbaum, G. (2015). What the orbitofrontal cortex does not do. *Nat Neurosci 18*, 620-627.

Stanovich, K.E., & West, R.F. (2000). Individual differences in reasoning:

Implications for the rationality debate? *Behav Brain Sci 23*, 645-726.

Steinberg, E.E., Keiflin, R., Boivin, J.R., Witten, I.B., Deisseroth, K., & Janak, P.H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nat Neurosci 16*, 966-973.

Stephens, D.W., & Krebs, J.R. (1986). *Foraging Theory* (Princeton, NJ: 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*, 1357-1366.

Strait C.E., Sleezer B.J., & Hayden B.Y. (2015). Signatures of value comparison in ventral striatum neurons. *PLoS Biol 13*, e1002173.

Sugam, J.A., Day, J.J., Wightman, R.M., & Carelli, R.M. (2012). Phasic Nucleus Accumbens Dopamine Encodes Risk-Based Decision-Making Behavior. *Biol Psychiatry 71*, 199-205.

Sutton R.S., & Barto A.G. (1981). Toward a modern theory of adaptive networks: expectation and prediction. *Psychol Rev 88*, 135-170.

Sutton, R.S., & Barto, A.G. (1987). A temporal-difference model of classical conditioning. *Proceedings of the Ninth Conference of the Cognitive Science Society*, 355-378.

Sutton, R.S., & Barto, A.G. (1998). *Reinforcement Learning: An Introduction*, 2002 edition, Bradford Books, MIT Press, Cambridge, MA.

Sharot, T. (2011). The optimism bias. *Curr Biol 21*, 941-945.

Shirayama Y, & Chaki, S. (2006). Neurochemistry of the nucleus accumbens and its relevance to depression and antidepressant action in rodents. *Curr Neuropharmacol 4*, 277-291.

Syed, E.C., Grima, L.L., Magill, P.J., Bogacz, R., Brown, P., & Walton, M.E.

(2015). Action initiation shapes mesolimbic dopamine encoding of future rewards. *Nat Neurosci 19*, 34-36.

Takahashi Y.K., Roesch M.R., Wilson R.C., Toreson K., O'Donnell P., Niv Y., & Schoenbaum G. (2011). Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. *Nat Neurosci 14*, 1590-1597.

Tobler, P.N., Fiorillo, C.D. & Schultz, W. (2005). Adaptive coding of reward value by dopamine neurons. *Science 307*, 1642-1645.

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*, 515-518.

Tremblay, L. & Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. *Nature 398*, 704-708.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: heuristics and biases. *Science 185*, 1124-1131.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychol Rev 90*, 293-315.

Ungless, M.A., Magill, P.J., & Bolam, J.P. (2004). Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. *Science 303*, 2040-2042.

Van Horn, J.D., Irimia, A. Torgerson, C.M., Chambers, M.C., Kikinis, R., & Toga, A.W. (2012). Mapping connectivity damage in the case of Phineas Gage. *PLoS ONE 7*, e37454.

Vaidya, A.R., & Fellows, L.K. (2015). Ventromedial frontal cortex is critical for guiding attention to reward-predictive visual features in humans. *J Neurosci 35*, 12813-12823.

Wallis, J.D. (2012). Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nat Neurosci 15*, 13-19.

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

Watabe-Uchida, M., Zhu, L., Ogawa, S.K., Vamanrao, A., & Uchida, N. (2012). Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron 74*, 858-873.

Wenzel, J.M., Rauscher, N.A., Cheer, J.F., & Oleson, E.B. (2015). A role for phasic dopamine release within the nucleus accumbens in encoding aversion: A review of the neurochemical literature. *ACS Chem Neurosci 6*, 16-26.

White, N.M. (2011). Reward: What is it? How can it be inferred from behavior? In: *Neurobiology of sensation and reward*, edited by Gottfried, J.A., Boca Raton (FL): CRC Press/Taylor & Francis.

Winecoff, A., Clithero, J.A., Carter, R.M., Bergman, S.R., Wang, L., & Huettel, S.A. (2013). Ventromedial prefrontal cortex encodes emotional value. *J Neurosci 33*, 11032-11039.

Wise, S. (2008). Forward frontal fields: phylogeny and fundamental function. *Trends Neurosci 31*, 599-608.

Young, L., Bechara, A., Tranel, D., Damasio, H., Hauser, M., & Damasio, A. (2010). Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. *Neuron 65*, 845-851.

Zigman, J.M., Jones, J.E., Lee, C.E., Saper, C.B., & Elmquist, J.K. (2006). Expression of ghrelin receptor mRNA in the rat and the mouse brain. *J Comp Neurol 494*, 528-548.