Computational Model-Based Functional Magnetic Resonance Imaging of Reinforcement Learning In Humans

Thesis Submitted for the partial fulfilment of the requirements of the University of Hertfordshire for the Degree of

“Doctor of Philosophy”

by

Burak Erdeniz

The Programme of research was carried out as a joint venture between the School of Computer Science and Psychology at the University of Hertfordshire

September 2012

This work is carried out under the supervision of

Dr. John Done

Dr. Neil Davey

Mr. Ray Frank

Dr. Reinoud Maex

ABSRACT

The aim of this thesis is to determine the changes in BOLD signal of the human brain during various stages of reinforcement learning. In order to accomplish that goal two probabilistic reinforcement-learning tasks were developed and assessed with healthy participants by using functional magnetic resonance imaging (fMRI). For both experiments the brain imaging data of the participants were analysed by using a combination of univariate and model–based techniques.

In Experiment 1 there were three types of stimulus-response pairs where they predict either a reward, a neutral or a monetary loss outcome with a certain probability. The Experiment 1 tested the following research questions: Where does the activity occur in the brain for expecting and receiving a monetary reward and a punishment ? Does avoiding a loss outcome activate similar brain regions as gain outcomes and vice a verse does avoiding a reward outcome activate similar brain regions as loss outcomes? Where in the brain prediction errors, and predictions for rewards and losses are calculated? What are the neural correlates of reward and loss predictions for reward and loss during early and late phases in learning? The results of the Experiment 1 have shown that expectation for reward and losses activate overlapping brain areas mainly in the anterior cingulate cortex and basal ganglia but outcomes of rewards and losses activate separate brain regions, outcomes of losses mainly activate insula and amygdala whereas reward activate bilateral medial frontal gyrus. The model-based analysis also revealed early versus late learning related changes. It was found that predicted-value in early trials is coded in the ventro-medial orbito frontal cortex but later in learning the activation for the predicted value was found in the putamen.

The second experiment was designed to find out the differences in processing novel versus familiar reward-predictive stimuli. The results revealed that dorso-lateral prefrontal cortex and several regions in the parietal cortex showed greater activation for novel stimuli than for familiar stimuli. As an extension to the fourth research question of Experiment 1, reward predicted-values of the conditional stimuli and prediction errors of unconditional stimuli were also assessed in Experiment 2. The results revealed that during learning there is a significant activation of the prediction error mainly in the ventral striatum with extension to various cortical regions but for familiar stimuli no prediction error activity was observed. Moreover, predicted values for novel stimuli activate mainly ventro-medial orbito frontal cortex and precuneus whereas the predicted value of familiar stimuli activates putamen. The results of Experiment 2 for the predicted-values reviewed together with the early versus later predicted values in Experiment 1 suggest that during learning of CS-US pairs activation in the brain shifts from ventro-medial orbito frontal structures to sensori-motor parts of the striatum.

Acknowledgements

Firstly, I would like to thank to my primary supervisor John Done who gave me the opportunity for doing this research. In the last four years being the only person working on a neuroimaging in the department was extremely challenging for me but John supported me financially, motivationally and intellectually in every period of these four years. Without his encouraging questions during our meetings it would be impossible for me to go deeper in the rabbit hole. My special thanks also go to my co-supervisors Neil Davey and Ray Frank whom helped me to understand many computational and bureaucratic problems. I would also like to thank to my fourth co-supervisor Reinoud Maex, for giving me insight about computational neuroscience and helping me interpreting electrophysiological studies in more detail, as well as teaching me how to write articles.

During these three years, I often felt alone, solitary, and sometimes depressed but in those days my friends Giseli de Sousa and Jean Martina always there for me. They always motivated me to find something good in bad. Also I want to thank to Elliot Clayton Brown without him I couldn't be able to come to my third year.

I also want to thank to Volker Steuber, Lucy Annett, Emrah Duzel and Oguz Tanridag for useful discussions and feedback for my research. Moreover, I would like to thank to Vicky Goh, James Sterling, Linda, and Ian from Paul Strickland Scanner Center in the Mount Vernon Hospital, who helped me to overcome many technical issues related to scanner. Special thanks will go to my dearest friend and mentor Semir Zeki. I also would like to thank to Wolfram Schultz, Yael Niv, Nathaniel Daw, Tim Behrens, Michael Frank, Mathias Pessiglione whom I got chance to discuss many topics on reinforcement learning during various conferences, personal meetings, and workshops. I especially want to thank to Peter Dayan and Daniel Polani our meetings were very delightful. I am also grateful for my friends Dicle Dovencioglu, Inci Ayhan, Thiago Matos, Jane Garrison, Lindsey Hudges, Louis, Hema and Supri for their support during these three years and thanks to my friends in the Biocomputation and Adaptive Robotics group.

Finally, the biggest thank will go to my mother and father whom I dedicate this thesis.

Table of Contents

List of Figures

List of Tables

Chapter 1

Overview of the Thesis

1.1 Motivation

Stimuli that we encounter in everyday life get much of their meaning from the associations that have been learned from previous experiences. These stimuli may trigger associations of rewards and punishments or in certain circumstances they may trigger courses of actions including simple stimulus-response mappings (e.g., green light means pass in the traffic) or even more abstract socio-cultural rules (e.g., not to start eating before the guest sits at the dinner table). As well as being affected from previously taught associations there is also a need to learn new associations from scratch, or overwrite new rules over the previously learned ones (e.g., reversal learning). In certain circumstances, there is also the need to switch between these associative rules in order to better adapt to the environment. Given the complexity of the spectrum of associative mappings, the capacity to predict and learn these mappings of associations in a flexible way is likely to increase the chance of survival of the organism. For this reason one of the most important research questions that behavioural and cognitive neuroscience are concerned with is how the brains of animals and humans learn and predict outcomes of rewards and punishments, and make decisions based on these prediction in order to avoid punishments and to obtain rewards.

In behavioural psychology, this question has been investigated with Pavlovian (classical) and instrumental (operant) conditioning, and much evidence has accumulated regarding different aspectsof the associative learning mechanisms (e.g., goal directed and habit mechanisms). On the other hand electrophysiological and anatomical studies in animals suggested that these associative learning mechanisms likely to involve separate brain regions (Frank, Cohen, Sanfey; 2009), and transitions might occur between these regions in different learning contexts (Packard & Knowlton, 2002; Yin & Knowlton, 2006; Gaybiel, 2008). Given that evidence from lesion and neuroimaging studies suggests that there are multiple learning mechanisms, each utilizing rewards and punishments in different ways (Packard & Knowlton, 2002).

This thesis aims to answer two specific but interrelated questions of associative learning. Firstly, what are the neural correlates of monetary gains and losses during reinforcement learning? Secondly, what are the neural correlates of goal-directed and habitual learning systems (automated and controlled) in relation to novel and familiar stimuli? The thesis examines these two questions with two functional imaging studies. Furthermore, by using a model-based functional neuroimaging technique this thesis also aims to provide answers for where in the brain prediction errors and predicted values coded.

1.2 Structure of the Thesis

- The literature related to the neurobiological basis of rewards and punishments and their affect on learning stimulus-response associations are reviewed in **Chapter 2**.

- **Chapter 3** reviews the literature on the neural mechanisms of associative learning including a review of the neuroanatomy of basal-ganglia, and a discussion of how novelty influence perception of stimuli and action sets during reinforcement learning.

- In **Chapter 4**, formal models of reinforcement learning are reviewed. The models included in this chapter are the linear learning rule of Bush and Mosteller (1955), Rescorla-Wagner (1972), Pearce-Hall (1980). In addition to that Temporal Difference learning (Sutton & Barto, 1988), Q-learning (Sutton & Barto, 1988) and several adaptive learning rate models also reviewed.

- **Chapter 5** summarizes general methodology of fMRI including pre-processing and statistical analysis. In addition to that it reviews how computational models of reinforcement learning are utilized in model-based fMRI.

- **Chapter 6** reports the results of the first fMRI Experiment that examined the neural correlates of monetary gains and losses with a binary choice probabilistic-learning task. A related hypothesis concerning the neural correlates of gains and losses, which refers to the opponent relationship between successful avoidance of losses versus monetary gains was also tested in this chapter. Additional model-based analyses were carried out in this chapter, which tested the neural correlates of prediction errors for gains and losses and of expected value.

- **Chapter 7** is a continuation of **Chapter 6** where additional hypothesis were tested based on the evidence that suggest a shift of activity in the brain (in the rosto-caudal axis) during associative-learning. According to previous work, it was hypothesized that neural activity from rostral to caudal brain regions reflect a shift from goal directed to habitual learning (Graybiel, 2008). This hypothesis is tested by analysing anticipatory responses for predicted values of CS for early versus late learning trials, and reported in **Chapter 7**. It is important however, to mention that the additional analyses reported in **Chapter 7** have been interpreted within a model-based fMRI framework, in terms of the computations that are carried out in interpreting fMRI data.

- In order to examine the difference in early versus late learning trials in more detail, a *second experiment* is designed and this is revealed in **Chapter 8**. More specifically, in the *second experiment* the participants were pre-trained before the scanning session with an instrumental conditioning task utilizing the reward value of a set of abstract symbols where at the end of training they reach the asymptotic levels of the learning curve. After the pre-training session, the fMRI session was carried out and during the fMRI session the familiar stimulus set were intermixed with a set of novel stimuli in order to identify the brain regions that respond more to novel stimulus set than to the familiar sets. The analysis in this chapter looked at differences in brain region for novelty and familiarity with a uni-variate statistical analysis. An additional functional connectivity analysis was also carried out in order to determine which parts of the stimulus novelty signal is broadcasted and received.

- In **Chapter 9**, model-based analysis was performed for novel and familiar stimulus sets in order to find the neural correlates of predicted value, prediction error and adaptive learning rate. This analysis was carried out in order to further test the hypothesis proposed in **Chapter 7**.

- Finally **Chapter 10** includes a general discussion and conclusion section. The contributions of this thesis to the field of decision neuroscience also summarized in this chapter with additional suggestion for future research.

Chapter 2

Neural Correlates of Gains and Losses

2.1 Brain Mechanisms of Rewards & Punishments

2.1.1 Rewards and Punishments as Reinforcers

In neuroscience there are different definitions for rewards, which are based on different properties of it such as its reinforcing features (that make us work for more) or hedonistic characteristics (that make us like rewards) (Berridge & Robinson, 2003; Berridge & Kringelbach, 2008; Berridge, Robinson, Alridge, 2009). In this thesis, I will be primarily concerned with the reinforcer definition of rewards.

In his book on animal intelligence written in 1911, Edward Thorndike showed that responses that are followed by satisfactory outcomes are more likely to be repeated again when the animals' are faced with a similar situation, and responses that produce discomforting outcomes are less likely to be repeated in a similar situation. This associative mechanism later known as the 'law of effect' holds that animals repeat certain responses because those responses end up with pleasurable outcomes (Thorndike, 1911). According to Thorndike's definition of the 'law of effect' rewards are the objects or events that make us come back for more, which makes the organism increase the probability of repeating certain behaviours (Thorndike, 1927). These behavioural changes, induced by rewards, usually occur through instrumental (operant) conditioning and form the basis for theories of reinforcement learning (see **Section 2.1.3.2** for a more detailed account of learning paradigms). According to this definition, pleasurable stimuli are positive reinforcers and elicit approach behaviour in the organism (Schultz, 2007c).

Unlike rewards, punishments, or so called aversive stimuli, work through unpleasant objects and events that make the organism avoid certain circumstances and reduce the frequency of repeating certain behaviours (Estes, 1967). Aversive stimuli can take various forms, for example physical pain (e.g., thermal or electrical stimulation of skin) is considered to be a common aversive stimulus used in various animal and human experimental settings (Seymour et a., 2007). Stimuli that cause a bitter taste are also used as aversive stimuli in neurophysiological experiments (Zald et al., 1998).

2.1.2 Differences in Primary and Secondary Reinforcers

Neuroscientists commonly talk about two types of reinforcers: unconditioned rewards that are accepted as primary reinforcers such as the food or water, whereas conditioned rewards like money are considered as secondary reinforcers (Grabenhost & Rolls, 2011). As with secondary rewards, punishment can also take secondary forms like social exclusion (Eisenberg et al., 2003; Eisenberg & Liberman 2004; McDonald & Leary, 2005) or monetary loss (O'Doherty, et al., 2001; Yacubian et al., 2006) and even regret and envy (Camille et al*.,* 2004; Shamay-Tsoory et al*.,*2007).

In order to satisfy the vegetative needs most animals and humans search for rewarding reinforcers, which is known as foraging behaviour (Steps and Krebs, 1986). The reason for this is that mammalian brains are equipped with certain brain regions, which motivate the organism in the absence of available reinforcers to search for rewarding stimuli (Rolls, 2005). For example, in the absence of certain primary reinforcers such as food and water, brains of many mammalian species are equipped with neural structures that trigger specific hormones and neurotransmitters that create the feeling of thirst and hunger, which can make the organism search for these missing rewards. These internal self-generated signals are the reason why the term "unlearned" is used for primary reinforcers (Rolls, 2005). In fact, there are certain brain regions that are specialized for these internal self-generated responses. For example, the lateral hypothalamus produces hunger signals, which control food intake, energy expenditure (Gao & Horvath, 2007) and body weight (Rolls, 1999). The level of these internal drive signals such as metabolic hormones like leptins cause hunger, and evoke changes in blood glucose level of the organism (Davis, Choi, Benoit, 2010; Parylak, Koob, Zorilla, 2011). Understanding which brain regions are involved in the generation of these internal signals and how they interact with the reward centers like orbitofrontal cortex are considered to be crucial for understanding craving related brain circuits (Siep et al., 2009; Rolls and McCabe, 2007; Pelchat et al., 2004) and reward value coding circuits that are very much associated with different eating disorders (Palmiter, 2007; Grill, Skibicka, Hayes, 2007; Berridge et al., 2010; Wagner et al., 2007; Kaye, Fudge, Paulos; 2009; Fladung et al., 2010). Moreover, primary reinforcers like water, food or sex have direct evolutionary benefits for the organism, for instance they balance the internal homeostasis of the organism and are crucial for survival and reproduction (Rolls, 2005). In addition to that, these internal self-generated signals might be species specific and differ depending on the evolutionary history of that species (Watson, Shephard, Platt, 2009; Watson & Platt, 2008). A good example can be found in North American minks (Watson et al., 2008). North American minks find splashing in water pools to be as rewarding as food rewards and when deprived of water pools they show increased cortisol levels, a response similar to being deprived of food resources (Mason et al., 2001). Another good example for a species-specific response to primary rewards is domestic cats. Domestic cats are devoid of sweet-taste receptors and show no reward responses to orally applied sucrose solutions whereas most species such as homo-sapiens find sweet drinks pleasurable (Li et al, 2005).

On the other hand, secondary reinforcers are conceived of as having no value by themselves (Huettel, Song & McCharthy, 2004), but they are used to obtain primary reinforcers as suggested by the token theory of reinforcement learning[1] (Wolfe, 1936; Cowles, 1937). Money for example is a secondary reinforcer (Breiter et al., 2001; Delgado et al., 2000; Elliott et al., 2000; Knutson et al., 2001; Haruno et al., 2004;Tanaka et al., 2004), used as a tool to obtain food and other primary reinforcers (see for a discussion, Lea & Webley, 2006; Aydınonat & Erdeniz, 2008). Secondary reinforcers might also take various abstract forms like playing computer games (Erickson et al., 2010; Vo et al., 2011), charitable donations (Izuma et al., 2008, Carter et al., 2009), cultural rewards like art or certain brands (Erk et al., 2002; Kawabata & Zeki 2004; McClure et al., 2004), or even more abstract forms like humour (Mobbs et al., 2003) or academic success (Mizuna et al., 2008). It is important to note that, although there is an on-going debate on whether primary and secondary rewards activate overlapping or segregated limbic and cortical-regions (Lea & Webley, 2006; Delgado et al., 2011) providing either a positive or a negative reinforcer in learning situations can generate a greater neural response in the limbic structures compared to presenting either informative (Ghahramani & Poldrack, 2009) or non-informative feedback (Bischoff-Grether et al., 2009).

2.1.3 Value Representations for Primary and Secondary Reinforcers

[1] According to token theory of reinforcement learning monkeys can learn to receive fruit juice by exchangingtokens.

One of the most important concepts related to reinforcers is the goal value representations. Goal values reflect the motivational significance of the unconditional stimuli and are calculated by the brain's value system during the experience of outcomes. In addition to that one might imagine not all positive or all negative reinforcers have equal goal values and their values may change according to various factors (e.g., one might like grape juice more than apple juice). In fact electrophysiology studies in monkeys showed that neurons in the orbitofrontal cortex not only differentiate between these objects (Tremblay & Schultz, 1999; Padoa-Schioppa & Assad, 2006) but also calculate their economic value in a common value scale (see for a discussion Wallis, 2006; Stuphorn, 2006; Padoa-Schioppa, 2011). These findings have been replicated in humans and showed that the orbitofrontal cortex calculates the value of two incommensurable goods such as a box of chocolates and a 2 gigabyte usb disk on a common scale (Fitzgerald et al., 2009; Chib et al., 2009). Moreover, orbitofrontal cortex not only uses a common value scale between two primary reinforcers, or two secondary reinforcers, but it uses a common value scale for comparing both primary (such as juice) and secondary rewards (such as money) (Chib et al., 2009) and even the value of two fundamentally different primary rewards: taste in the mouth and warmth in the hand (Grabenhorst et al., 2010). Likewise certain neurons in the orbito-frontal cortex can also calculate the value of negative reinforcers such as electric shock (Hosakawa et al., 2007).

Furthermore, certain dopamine neurons in the brain adapt their firing rate to the level of predicted-rewards and fine-tune their firing rate according to contextual expectations (Tobler et al., 2005). This adaptive scaling of reward value allows efficient firing rate for different amounts of rewards. For example, firing of my dopamine neurons are equal between two situations where I am expecting 100 ml of fruit juice but receiving 50 ml and where I am expecting 200 ml but receiving 150 ml of fruit juice.

Finally, it is important to note that the goal values of primary rewards are relative to the satiation level of the organism (Rolls et al., 1983; Colwill & Rescorla, 1985; Balleine & Dickinson, 1998; O'Doherty et al., 2000; Gottfried et al., 2003). Therefore, the goal value coding regions only light up as active in imaging studies if the participant is deprived (Pickens et al., 2005, de Aruja et al., 2006). For this reason, it is important to be aware of differences between primary and secondary reinforcers, because as mentioned above rewards have separable hedonic (e.g. subjective experience of taste) and quantitative (e.g., reward value) components that can be identifiable by different neuronal spiking characteristics in different brain regions (Pecina et al., 2006). A major advantage of using monetary reinforcers is that individuals are motivated to gain more money most of the time (Hertwig & Ortmann, 2001).

2.1.4 Anatomy and Neuropharmacology of Rewards and Punishments

2.1.4.1 Involvement of Dopamine in Reward Processing

Until the 1950's the dominant view of rewards was that of Hull's drive reduction theory (Hull, 1943). According to this theory newborn infants have innate drives like hunger or thirst, which make them attach to their caregivers to satisfy their needs (Wolpe, 1950). Although later this theory was challenged by Harlow (1953), until that time it was accepted that when newborn infants are deprived then particular drives are activated which then control the searching behaviour for rewards (Harlow, 1953). Therefore according to the drive reduction theory the primary motivation for searching rewards is punishment avoidance (Wolpe, 1950; Brown, 1955). However, in 1954 Olds and Milner published a remarkable paper, in which they showed that rats prefer to press a lever in a Skinner box in order to get electrical stimulation of their brains, rather than choose food or mating with a female rat although they were starving (Olds & Milner, 1954). Their study clearly showed that stimulating certain brain regions is experienced as more valuable than satisfying certain primary needs. This was especially found when stimulating regions involving ventral tegmental area, nucleus accumbens, hypothalamus, thalamus but was not found in all brain regions. For example when the septal and midbrain regions were stimulated they pressed the lever several thousand times but when cortical regions were stimulated they pressed the lever for approximately a couple of hundred times and most of the time their lever pressing was at the chance level (Olds & Milner, 1954). What Old's and Milner showed was when the brain regions on the dopamine pathway were stimulated rats increased the lever press dramatically not in order to compensate the lack of primary rewards but for pure pleasure. Because if rats were pressing the lever to compensate for their need of water or food, they would have stopped after getting enough stimulation (ie respond as if satiated), but they pressed the button until they died (Olds, 1956; 1958). Later on a similar experiment was repeated with human participants and monkeys, which revealed the first evidence of a link between the dopamine system and reward mechanisms (see for a review Delgado, 1969; Routtenberg, 1978). Since that time dopamine has been found to be the most important neuromodulator for reward processing (see for a review Schultz, 2002; 2007a; 2007b).

Dopamine neurons have their cell bodies in the substantia nigra pars compacta (A9 cell group), and ventral tegmental area (A10 cell group) (Anden et al., 1966). Compared to the total number of neurons in the brain (estimated to be 120 * 10^9, Herculano-Houzel, 2009), there are only a few hundred thousand dopamine neurons in the substantia nigra and only a few thousand in the ventral tegmental area (Kreitzer, 2009). Midbrain dopamine neurons have long axons and have terminals in various parts of the cortex and subcortical regions (Prensa & Parent 2001). The four major axonal pathways for dopamine neurons are the mesocortical pathway that connects the ventral tegmental area to prefrontal cortex, the mesolimbic pathway which connects the ventral tegmental area to the nucleus accumbens, amygdala and hippocampus, the nigrostriatal pathway connects the substantia nigra to neostriatum (putamen and caudate nucleus), and the tuberoinfundibular pathway connects the arcuate nucleus (mediobasal hypothalamus) to the pituitary gland (Björklund and Lindvall, 1984; Berger and Gaspar, 1994). For example, in the neostriatum, dopaminergic boutons are very dense (Arbuthnott & Wickens 2007), and they account for nearly 10% of all synapses in the striatum (Groves et al. 1994).

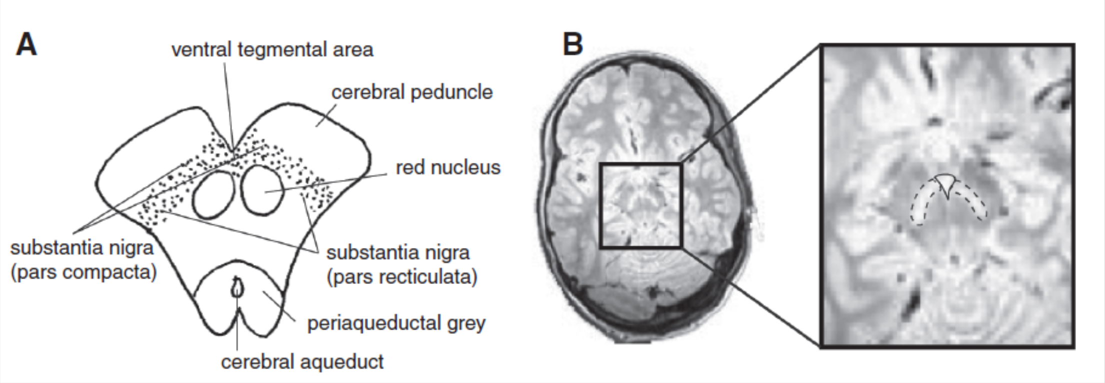


Figure 2.1 a) Illustration of dopaminergic nuclei in a horizontal section of the brainstem: ventral tegmental area and subtantia nigra. b) Midbrain dopaminergic nuclei from a Proton weighted MRI image. Substantia Nigra and Ventral Tegmental area are highlighted in the rectangular box. Both Figures taken from D'Ardenne et al., (2008).

Moreover pharmacological studies suggest that certain drugs like cocaine and amphetamine affect reward processing via dopaminergic mediation (Koop, 1992).Cocaine for example is a dopamine re-uptake blocker and binds to DA transporters (DAT) (Hyman et al., 2006). This neuropharmocological action of cocaine keeps the dopamine in the synaptic gap (Hyman et al., 2006). It has been thought that this action gives rise to the feelings of pleasure and relief at the time of committing drug administration (Hyman et al., 2006). Similar to cocaine, amphetamines work through dopaminergic systems (Koop, 1996; Robinson and Berridge, 2003; Hyman et al., 2006). However, because the molecular structure of amphetamine is very similar to that of dopamine, amphetamine passes through the synaptic membrane by using dopamine transporters and replenishes vesicles where dopamine is stored (Hyman, 1996; Hyman and Malenka, 2001). Studies using self-administration of cocaine and amphetamine in rats showed similar findings to electrical self-stimulation studies, as rats pressed the lever to receive cocaine until they die, which supports additional evidence on the role of dopamine in reward processing (Hyman and Malenka, 2001). Moreover, with the developments of optogenetic techniques the role of dopamine in reward processing becomes more prominent. Tsai et al (2009) showed that optogenetic stimulation of VTA dopamine circuitry causes rats to look for previously learned conditioned place references. Moreover by using a similar technique Adamantidis et al., (2011) showed that phasic activation of dopamine neurons in the VTA causes rats to look for previously extinguished food seeking behaviour.

2.1.5 Associative Learning and The Role of Dopamine in Calculating Neuronal Prediction Errors

Over the past century associative learning and more particularly the role of dopamine in associative learning has been studied using mainly two behavioural paradigms: Pavlovian learning and instrumental learning. The best way to explain Pavlovian learning is the example of Pavlov's dog. It was thought that Ivan Pavlov first realized the effect of rewards when he saw that the dogs in his lab started salivating when they saw the lab worker who was serving food for them (Bouton, 2007). In this example a neutral stimulus such as the lab workers coat was systematically followed by the presentation of an appetitive event ie the food (unconditional stimulus, US+). Over time the animal predicted the occurrence of outcomes (appetitive or aversive) by just seeing the neutral stimulus (conditional stimulus, CS). The pairing of CS and US causes the animal to learn the structure of the environment, which then causes behavioural responses (the conditional response, CR) to occur during presentation of the conditional stimulus (Pavlov, 1927). Moreover a conditional stimulus can predict the occurrence of another conditional stimulus that predicts reward through higher-order conditioning, which will increase the range of associative chains and make conditioning plausible for sequences of events (e.g., Holland & Rescorla, 1975a, 1975b; Rescorla, 1982). On the other hand in instrumental conditioning the outcome (such as food) is contingent on the animal's behaviour. The animal has to perform an action to receive a reward (maximize the amount of food) or to avoid a punishment (minimize the amount of foot shock). Therefore, in instrumental learning the unconditional stimulus becomes a reinforcer to motivate the animal to perform certain behaviours and will give the animal some control over the environment. The distinction between instrumental and Pavlovian learning paradigms is important. The first difference between Pavlovian and Instrumental conditioning is that in the Pavlovian conditioning paradigm the animal only observes the relationships between events that occur in the environment (conditional and unconditional stimuli) and through regular association learns the predictive relationship between the CS and the US. In instrumental learning the animal has influence over the environment (e.g., can control the occurrences of events) and can learn the predictions about the outcomes to guide his choices. Therefore, in the instrumental learning the associations between the actions and outcomes are crucial and changing the contingency of the outcomes will affect the animal's choices. In computational models of reinforcement learning this dichotomy between Pavlovian and instrumental learning in animal learning is captured by open and close loop environments (this is covered in detail in **Chapter 4**). Moreover, in the Pavlovian conditioning animals usually produce almost automatic conditional responses (e.g., salivating) to conditional stimuli whereas in instrumental conditioning these responses are thought to be controlled by higher-level cognitive processes ie are planned rather than automated.

In many of the experiments that involve learning stimulus reward associations, learning depends not only on the joint occurrence of conditional and unconditional stimuli but depends on the discrepancy between the actual occurrence and the predicted occurrence of rewards (Schultz, 2000). A good example for this is perhaps Kamin's (1969) blocking effect, which suggests that a new CS_x-US association can be "blocked" by an already existing association between the CS and the same US.

In the blocking paradigm a neutral stimulus A is paired with a reinforcer and another neutral stimulus let's say B is paired with nothing. In the next stage, A and B are paired with two other stimulus; *X* and *Y* to form compound stimuli ie AX and BY, which are also paired with rewards. In the test phase when *X* is presented alone it predicts nothing and Y presented alone predicts reward. This is because in the *AX* pair reward is predicted by A alone making *X* redundant or blocked. Similarly in the *BY* pair Y is free to associate with any reinforcer because it is not blocked by B and hence comes to predict reward. The blocking effect in both humans (Martin & Levey, 1991) and animals (Kamin, 1969) has shown us that simply pairing two stimuli (e.g. A&X or B&Y) with reward is not sufficient for learning that both of these stimuli predict reward, instead the reinforcer should be unpredictable if learning is to occur.

Although much evidence has accumulated on the role of dopamine in reward processing in the last 50 years, the dopamine hypothesis of reward has undergone refinement several times (Schultz et al., 1997; Schultz, 2000; Schultz & Dickinson 2000; Schultz, 2006, Berridge et al., 2009; Bromberg-Martin et al., 2011). These refinements suggests that the specific role of mesolimbic dopamine neurons may be more important for the acquisition of the reward-related behaviours than for subjective responses to rewards (Schultz & Dickinson 2000). A well-established influential theory about the role of dopamine in learning is that of Schultz (1998, 2000; Bayer & Glimcher, 2005). This theory is called the reward prediction-error theory and has its roots in the Rescorla-Wagner learning rule (Rescorla and Wagner, 1972) and more particularly in the temporal-difference reinforcement-learning model of Sutton and Barto (1998), both of which are summarized in **Chapter 3**. Indeed, Montague, Dayan and Sejnowski (1996) were the first to propose a relationship between the activity of dopamine neurons and prediction errors in reinforcement learning, as observed by the group of Wolfram Schultz. In their paper they stated "*the fluctuating delivery of dopamine from the VTA to cortical and subcortical target structures in part delivers information about prediction errors between the expected amount of reward and the actual reward*" (page 1944). In a series of Pavlovian conditioning experiments Schultz and colleagues investigate the nature of dopamine neurons firing properties. Dopamine is modulated by two mechanisms, tonic (single spike) and phasic (spike burts) dopamine release (Goto et al., 2007). Several studies suggested that the source of the tonic dopamine signalling is regulated by the activity of collinergic interneurons (for a discussion see Wickens et al., 2003) and others argued that tonic dopamine activity is regulated either by the prefrontal cortical afferents (Grace 1991) or by the hippocampus (Grace et al., 2007) also see Alcaro et al, (2007) for an alternative interpretation. Phasic dopamine release results from the activity of the dopamine-containing cells themselves (Schultz, 2002). This activity is characterized by either irregular single spikes, or rapid bursts (2-6 spikes) for about 100-500ms (Schultz, 2002; also Kroner et al., 2009). In their experiment Schultz and colleagues showed that when an arbitrary stimulus is paired with a reward (in this case it is sip of fruit juice) monkeys dopaminergic neurons in the ventral tegmental area fire phasically during the occurrence of reward. However, once learning was complete ie when the monkeys can fully predict the occurrence of rewards, the dopamine neurons did not fire above the baseline levels on receipt of reward. On the other hand when the dopamine neurons fully predicted reward but no reward was provided dopamine neurons fired were below the baseline levels at the time when the reward was expected (see **Figure 2.2a** top image). In fact most of the dopamine neurons produce a phasic prediction-error signal when the reward prediction, based on previous experience, was too low and conversely suppress firing when a prediction of an expected reward was not met by the actual reward (see **Figure 2.2a** bottom image). Hence, during the early training cycles, these neurons are activated by the onset of the unconditional stimulus, which comes as an unexpected reward. After learning, they are activated by the conditional stimulus, which becomes a predictor of future reward. In contrast, the response to the actual rewards vanishes after learning, when the conditional stimulus had already correctly predicted the reward. Therefore, the amount of phasic signalling is directly dependent on the level of surprise during the acquisition of the conditional stimuli. These dopamine activations to reward-predicting stimuli occur in almost 80% of dopamine neurons in the substantia nigra and in the ventral tegmental area (Schultz, 2007). It was argued that several regions receive this prediction-error signal, including nucleus accumbens, medial frontal cortex, the dorsolateral prefrontal cortex, and the amygdala (Schultz and Dickinson, 2000). Schultz and colleagues proposed that learning occurs by sending back and forth the error-signal between different regions. According to Schultz and Dickinson (2000), it is possible that dopamine neurons use the information about predicted rewards for the control of goal directed behaviour, and they suggested that this information helps to construct reward expectations in the form of predicted values whereas the prediction-error signal generated by dopamine neurons used to update the predicted-values associated with states and actions and these predicted-values mightpossibly stored in cortical and subcortical regions (see **Chapter 3** for discussions).

However the evidence for the prediction error signals in humans comes only recently with the advances in human brain imaging techniques.fMRI studies and their combination with computational models allow researchers to test specific hypotheses about the prediction errors. For example, similar to the Schultz Pavlovian learning experiment in monkeys, homologous experiments have been repeated in humans using functional magnetic resonance imaging. In his study, O'Doherty et al. (2003) used a Pavlovian appetitive conditioning experiment to monitor the brain regions involved in learning by prediction errors. The conditional stimuli were fractal pictures, each paired to a different flavour of juice. To enhance prediction errors, neutral juice was delivered on a fraction of 'sweet' trials, and sweet juice on a fraction of neutral trials. Based on electrophysiological studies in monkeys (Schultz, 1997), where dopamine neurons in the midbrain were observed to represent the prediction error signal, the reward expectation (and concomitantly with it the error prediction signal) was expected to shift over time from the unconditional stimulus to the conditional stimulus (see **Figure 2.2**). Similarly the omission of an expected reward is monitored by decrease in activity. The authors used a computational model of the temporal difference learning algorithm (see **Chapter 4**) to assess the magnitude of the prediction error. The prediction error signal, calculated by the TD algorithm, was used as an independent variable in a regression model with fMRI measurements as the dependent variable for each time a conditional or unconditional stimulus was presented (see **Chapter 5** for how similar model fitting procedures are applied in general to fMRI data sets). Significant activity was found mainly in the ventral striatum for the prediction error signal at the time of both conditional and unconditional stimuli (see **Figure 2.2**). In addition, significant effects were also found in the ventral globus pallidus, orbito-frontal cortex and dorsalateral prefrontal cortex, including the inferior and middle frontal gyri. The bilateral cerebellum also showed significant activity for prediction errors at both time points, especially when a slower learning rate was used in the model.

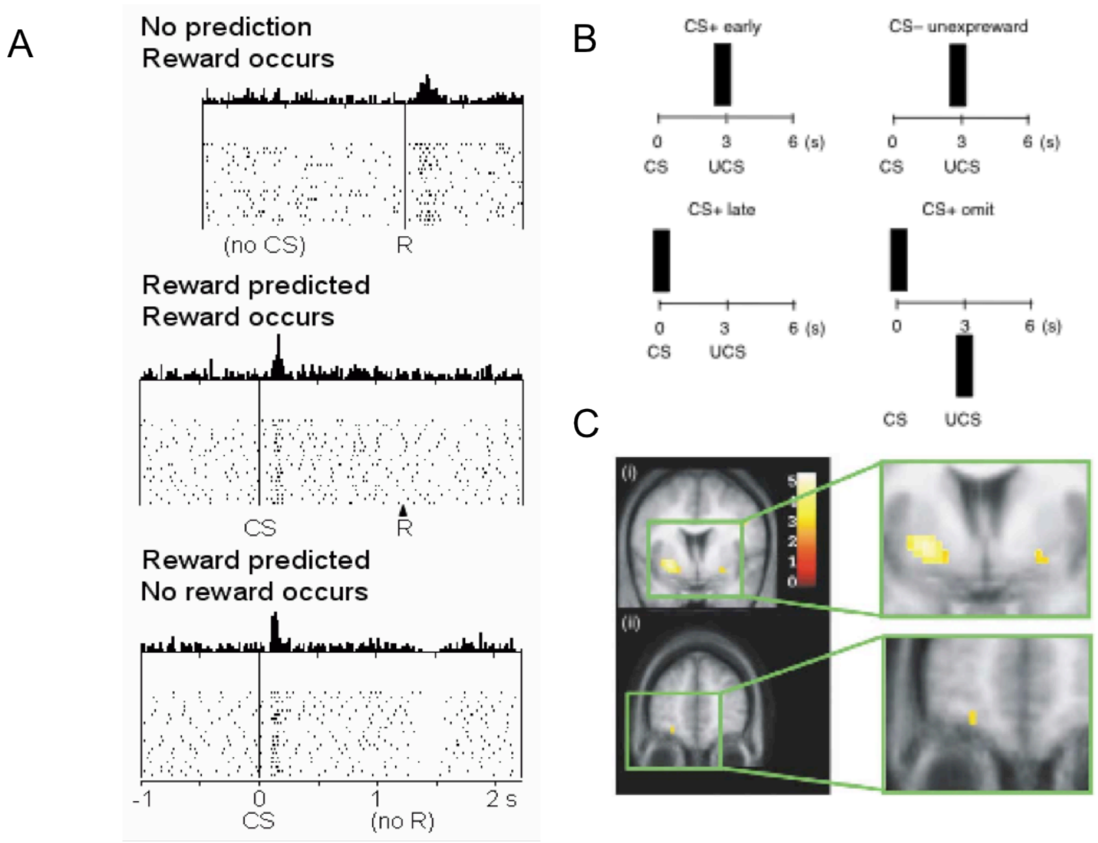


Figure 2.2 Shows prediction error activity in the dopaminergic neurons observed in behaving monkeys (on the left side) and prediction error signal observed in human brain imaging experiment (on the right side). a) Monkey electrophysiology experiment shows phasic excitation of dopamine neurons (positive prediction error) for an unpredicted reward outcome at the onset of US before learning takes place (top figure). The phasic excitation appears at the onset of the CS after learning (middle figure) and when the monkey expects a reward but receive nothing the activity occurs below the baseline (negative prediction error) (bottom figure). Figure 1a is taken from Schultz et al., (1997) b) Positive and negative prediction-error contrasts for a human fMRI study of O'Doherty et al., (2003). c) The results of the O'Doherty et al., (2003) experiment shows prediction-errors in the human striatum and orbito-frontal cortex. Figure 2.2b and Figure 2,2c are adapted from O'Doherty (2004).

Later on, several researchers concerned about the existence of prediction errors proposed two challenging questions (Pessiglione et al., 2006; Schonberg et al., 2007). Firstly(i) if the observed changes in the BOLD activity in imaging studies are actually caused by dopamine neurons during learning then the the BOLD signal *shouldn't change* if the participant already knows the relationship between stimulus and outcome(due to the Rescorla-Wagner prediction error having zero value since the outcomes are fully predicted)and secondly (ii) dopamine neurons spiking rate should increase or decrease with manipulation of the dopamine levels of the participants. Firstly, to answer the initial question Schonberg et al. (2007) scanned 29 participants separated into two groups depending on their behavioural performance in the reinforcement-learning task and they showed that the group that perform better in task showed a remarkable increase in the prediction error activity but the group with the lower behavioural performance showed lesser activity. Additionally, to find an answer to the second question that is related to the BOLD activation and dopamine administration Pessiglione et al., (2006) examined whether, during instrumental learning, the magnitude of the reward prediction error expressed in the striatum is affected by the administration of dopamine modulating drugs (L-DOPA, haloperidol). According to their hypothesis, subjects treated with L-DOPA should show a greater propensity to choose the rewarding action relative to subjects treated with haloperidol in a probabilistic 'go-no go' task. Similarly, they expected subjects treated with L-DOPA to show greater BOLD activation for prediction errors. In the experiment, subjects had to choose between two stimuli in which one of the stimuli had a rewarding outcome with a probability of 0.8 and the other a rewarding outcome with a probability of 0.2. They also used another set of stimuli for negative and neutral outcomes with the same probability values. The experimental procedure was such that if the participant pressed the button the program would accept this as the go response, or selection of the lower symbol, otherwise the program automatically chose the upper symbol as no-go response. In the analysis, outcome prediction errors were calculated for each group of subjects (placebo, L-DOPA and haloperidol) with a standard action-value learning algorithm. These learning models also perfectly modelled the drug enhancement effect. More particularly, the model parameters (e.g. learning rate) were adjusted to maximize the likelihood of the subjects' choices under the model. To this end, the behavioural learning curve was fitted with an action selection function parameterized as the value of the learning rate. Once the learning rate had been determined, the value of the action chosen at each trial could be updated in proportion to the prediction error it caused, defined as the difference between expected value and actual outcome. The imaging results revealed that subjects under L-DOPA showed greater activation in the striatum for reward prediction error signals than subjects under haloperidol (see **Figure 2.3a**). On the other hand, the insular cortex showed greater activation for the punishment prediction error during the avoidance-learning condition (see **Figure 2.3a**and **Figure 2.13c** and alsosection 2.1.5.1.3 for a detailed discussion of the role of insular cortex in avoidance learning).

Moreover, until now only the prediction errors in humans that are observed by functional imaging studies are reported. It is important to note that all of these reports are indirect measurements (even the DA agonist administration study of Pessiglione et al., 2006) due to the physical properties of the BOLD signal (see **Chapter 5**). Perhaps the first direct electrophysiological evidence of prediction errors in humans comes from deep brain stimulation studies in Parkinson'spatients. Zaghloul et al., (2009) tested Parkinson patients' (PD) responses to unexpected rewards where they used intracranial microelectrode recordings from substantia nigra in a group of PD patients who underwent deep brain stimulation surgery. All of the participants engaged in a probability learning experiment using financial rewards where half of the conditional stimuli predict reward with high probability and other half predicts reward with low-probability. During learning neurons in the substantia nigra showed a unique firing pattern that is increased for unexpected rewards and decreased firing for expected reward omissions (see **Figure 2.3b**bottom figure on the left). However, this activity disappeared after participants learned the probabilistic reward contingency (see **Figure 2.3b**bottom figure on the right).

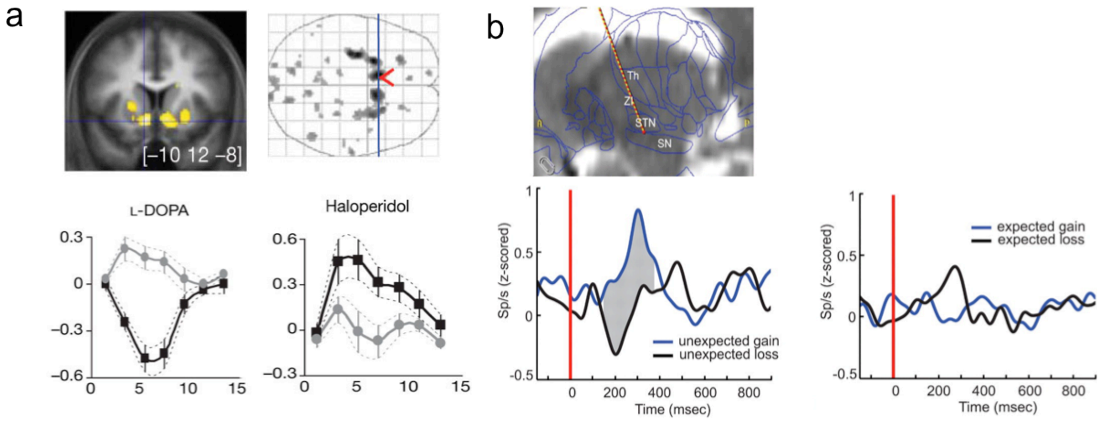


Figure 2.3 a) Figure on the top leftthebrainactivations seen as yellow on the coronal slice (left) and the grey areas in the axial slice of the glass brain (right) show significant effects of *reward prediction error* signal in the striatum. The bottom figures on the left side show the effect of L-DOPA in ventral striatum (at the peak voxel) for the *reward prediction error*and the effect of dopamine antagonist (Haloperidol) forthe *punishment prediction error* in the Insula (at the peak voxel). Figure*2.3a* is adapted from Pessiglione et al., (2006).b) Anatomical Microelectrode recording sites in the substantia nigra (SN) (Top image). Bottom graph on the left shows the mean firing rate of SN neurons for unexpected monetary gains and unexpected monetary losses and the graph on the bottom right shows the mean firing rate of SN neurons for expected monetary gains and expected monetary losses. Figure*2.3b* is adapted from Zaghloul et al (2009).

In summary most of the time in the fMRI studies reward prediction errors in humans are observed in striatal regions and this activity interpreted as a proxy measure of dopaminergic activity for reward prediction errors (Montague, King-Casas, et al. (2006)). However in the last decade several imaging and electrophysiology studies showed that not only striatal regions but also other regions such as cingulate cortex and medial-frontal cortex (Oya et al., 2005; Amiez et al., 2006; Matsumoto et al., 2007) show significant activation for prediction-errors. In addition several studies showed that insular cortex and amygdala are also involved in coding prediction-errors that are involved in learning the value of aversive stimuli (Yacubian et al., 2006; Pessiglione et al., 2006). In the following sections, I reviewed these other regions that are involved in coding prediction errors for rewards and punishments. Moreover dopamine is not only involved in reward related activity but also it is involved in punishments. These functions of dopamine in punishments are also reviewed in the following sections.

2.1.6 Involvement of Dopamine in Punishment

Until now I reported evidence about the role of dopaminergic neurons in processing reward information and more specifically their role in learning from unexpected reward outcomes. However, recent findings showed that induction of painful stimuli also activates dopaminergic neurons across a variety of different species (Ungless, 2004). For example, Brischoux et al. (2009) showed that giving foot shock to rats increase the firing rate of dopaminergic neurons (see also, Mantz et al., 1981; Becerra et al., 2001; Coizet et al., 2006). Moreover, dopamine neurons in monkeys show phasic excitation for aversive cues that predict punishments (Schultz & Romo, 1987; Matsumoto & Hikosaka, 2009, Joshua et al., 2009). In humans various imaging studies showed activity related to pain in the main projection cites of dopaminergic neurons in the ventral striatum (Scott et al., 2006) and dorsal striatum (Bingel et al., 2004; Scott et al., 2006).

Altogether the evidence suggests an increase in dopamine firing for aversive outcomes and stimuli that predict aversive outcomes. However, recent discussions put together with the classic role of dopamine neurons in punishments suggest that increased firing in dopaminergic neurons happens only in minority of dopamine neurons 3-49 % where most dopaminergic neurons usually show a decrease in activity for aversive CS and aversive US and ithasn't been clear which aspect of rewards and punishments cause differences in firing (see for a discussion see, Ungless et al., 2004; Schultz, 2010; Frank & Surmeir, 2009).

2.1.6.1 The Role of Dopamine in the Interaction Between Reward and Punishment Processing

It is hard to classify rewards and punishments as always objects of desire or objects of aversion. There are countless types of rewards and punishments which have many facets for humans where some of them have tremendous consequences for some people and some are crucial for those peoples' survival (Kringelbach & Berridge, 2009; 2011). Certain addictive drugs are good examples for this dichotomy. Opiates like morphine are both deteriorative objects of pleasure, which can cause addiction in the organism, and they are also strong pain killer (analgesics) used in certain medical settings. For example, mice whose μ-opioid[2] receptors are genetically knocked out lackboth the analgesic effects of opioids and the effects of physical dependence (Matthess et al., 1996). This suggests that most of the time there is a trade-off between the cost of pain and the benefit of pleasure, making reward and punishment processing a complicated subject (Leknes & Tracey, 2008; Talmi et al., 2009). In fact several studies suggested that aversive stimuli and rewards are not completely different from each other but they interact in various psychological and neural levels (Konorski, 1967; Grossberg, 1984). For example in rare conditions people might experience pleasure and pain simultaneously. In certain contexts such as when people are engaging in sadomasochistic sex (Stark et al., 2005; Georgiadis & Kortekaas, 2009) or when eating spicy food (Grabenhorst et al., 2007; Rolls, 2009) they might experience both pain and pleasure. This is because rewards and punishments have a shared component that is called motivational saliency (Bromberg-Martin et al., 2010) because rewards and punishments are both salient objects in the way they deserve attention (Jenson et al., 2006; Cooper & Knutson, 2008). However, they are different from each other in the way they affect learning behaviour such that punishers decrease the probability of behaviour (e.g., lever press) and rewards increase the probability of behaviour. Then how is it possible that dopamine neurons are coding both the salient properties of aversive stimuli as phasic excitations and the value properties as below baseline activity?

[2] This receptor class is one of the main opioid receptor classes. The other two-receptor classes are δand κ opioid receptors.

To resolve this debate about the role of dopamine in appetitive and aversive conditioning Bromberg-Martin et al., (2010) proposed the existence of two functional distinct types of dopamine neurons. These two distinct types of dopamine neurons process two different aspects of rewards and punishments. The first type of dopamine neurons encodes *motivational-salience* properties, which modulates cognitive processing and motivational drives and the second type encodes a learning signal such as a prediction error signal referred as the *motivational-value* coding dopamine neurons. They suggested that dopamine neurons, which encode motivational-saliencyrespond to both rewarding and aversive events but dopamine neurons which encode the motivational-value respond only to rewarding events and are inhibited by aversive events (see **Figure2.4a** and **Figure2.4d).** Bromberg-Martin et al., (2010) also suggested that these dopamine neurons, which code motivational-saliency, are located in different anatomical regions to motivational-saliency coding dopamine neurons. They argued that motivational-saliency coding dopamine neurons are located mostly in the dorsolateral substantia nigra and medial VTA and motivational-value coding dopamine neurons are more densely located in ventromedial substantia nigra and VTA.

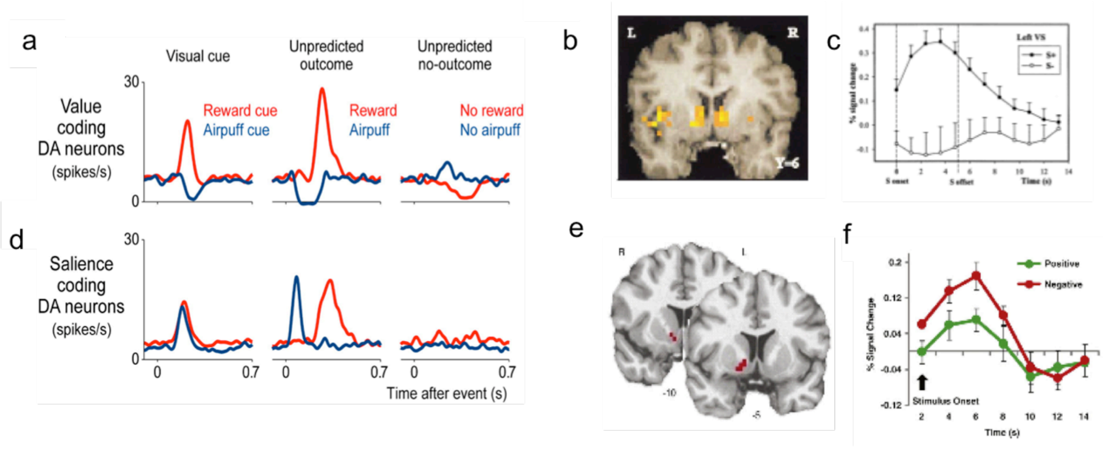


Figure 2.4 a) Monkey electrophysiology studies showed that value coding dopaminergic neurons increase firing rate for visual cues which predicts rewards and unexpected reward outcomes but they show below baseline activity for expected aversive cues and unexpected aversive outcomes. d) A separate group of dopaminergic neurons shows above baseline activity for both visual cues that predict rewards and punishments and outcomes of rewards and punishment. *Figure 2.4a* and

Figure 2d is taken from Bloomberg-Martin et al., (2010) b-c) On the left side coronal slice shows significant activity of NAcc for cues that predict aversive and rewarding outcome. On the right side graph shows time course of heamodynamic response changes as increase for cues that predict reward and decrease for cues that predict electric shock. Figure 2.4b and Figure2.4c are taken from Jenson et al., (2003). d) NAcc activation map in the coronal plane ($p<0.05$, corrected). Time course for the hemodynamic response in the right NAcc cluster. Figure 2.4d is taken from Levita et al., (2009). Experiments used in *Figure 2.3c* and *Figure2.3d* are very different in design with the former includes learning and with the latter excludes learning and any motor responses.

Altough the hypothesis of Bromberg-Martin et al., (2010) is hard to test with current imaging studies due to inherent nature of the physics of BOLD signal and slow spatial resolution of fMRI studies (it is not possible to identify whether same or different DA neurons in the VTA are involved in processing motivational-salience or motivational-value because the BOLD signal is correlated with local field potentials rather than single neuron activity), several lines of study havereported parallel findings in humans for the motivational-value coding dopamine neurons for aversive stimuli in the striatum. Jenson et al. (2003) reported motivational- value coding in ventral striatum. In an aversive conditioning experiment they found that BOLD response in the ventral striatum increased for conditional stimuli that predict reward but decreased for conditional stimuli that predict punishment (see **Figure 2.4b**). However, Seymour et al. (2004) in a second-order Pavlovian aversive conditioning experiment showed that ventral striatum and insular cortex code punishment prediction errors supporting the role of dopamine neurons in coding aversive values. Moreover in a recent aversive conditioning study where participants were randomly allocated to either placebo, amphetamine (DA agonist) or haloperidal (D_2 antagonist), Diaconescu and colleagues (2010) showed that blocking dopamine transmission via haloperidol was associated with significant increase in the functional connectivity between theamygdala and the insula making it's role critical for the areas involved in the avoidance circuitry (see **Section 2.1.5.1.3** for the role of insular cortex in prediction errors). In addition Pessiglione et al., (2006), in an fMRI experiment, showed that administration of D_2 receptor antagonist haloperidol, affected the performance of learning from negative outcomes and increased the BOLD signal for negative prediction errors in the insula. In the light of these studies, administration of dopamine agonist seems to impair learning from bad outcomes (decreasing the effects of punishments) and vice versa, that is to say, administration of dopamine antagonist should improve learning performance from bad outcomes. Actually, this is exactly what several studies have found. Administration of D_2 agonist prepemoxil, and other agents acting upon D_2 receptors, impaired the signalling of bad outcomes (Frank and O'Reilly, 2006; Santesso et al, 2009; van Eimeren et al, 2009) and caused an increase in loss-chasing behavior[3] (Campbell-Meiklejohn, et al., 2008; 2011). These results are also supported by animal experiments where D_2 receptor agonists suppressed pain related responses anddopamine D_2 receptor antagonistsenhanced pain-related responses (Ben-Sreti et al., 1983; Lin et al., 1981; Magnusson and Fisher, 2000). Overall these studies suggest that dopamine neurons carry useful information not only about rewarding outcomes but also involved in processing of aversive stimuli that predict future punishments. Furthermore, it was suggested that dopamine-D_2 agonists impair learning from aversive outcomes because it decreases the effectiveness of painful stimuli and dopamine D_2 antagonists improve learning with aversive outcomes perhaps by increasing the value of punishments.

For the motivational-salience part of the hypothesis several lines of evidence suggest that even in the absence of learning, VTA and nucleus accumbens, the main projection site of DA neurons from VTA, shows increases in BOLD response (Zink et al., 2004). Levita et al., (2009) tested whether nucleus accumbens responds to unpleasant auditory stimulation in the absence of learning. They showed increased BOLD response

[3] Loss chasing is commonly observed in people who have gambling addiction, where they play again to cover up their losses. It was considered to be an indication of impaired learning from negative outcomes.

at the onset of both pleasant and unpleasant sound in the right ventral striatum (see **Figure 2.3e**).

Moreover, Carretie et al., (2009) showed similar findings for unpleasant pictures in the caudate nucleus and prefrontal cortex. Finally, in a monetary incentive delay task where no learning is involved, Carter et al., (2009) showed increased BOLD response in the VTA and nucleus accumbens for stimuli that predict self and charity related monetary donation. Bromberg-Martin and Hikosaka's proposal also fits well with non-reward activity of DA neurons observed in several other experimental settings which related to surprising, novel, salient, and aversive experiences (Redgrave et al., 1999; Horvitz, 2000; Di Chiara, 2002; Joseph et al., 2003; Pezze and Feldon, 2004; Lisman and Grace, 2005; Redgrave and Gurney, 2006). However, as mentioned in **section 2.2** prediction-errors for aversive outcomes might be represented in different brain regions such as insular cortex (Pessiglione et al., 2006; Preuschoff et al., 2008) and habenula (Matsumto and Hikosaka, 2007; Salas et al., 2010; Ide and Li, 2011) and might be influenced more by the activity of serotonergic neurotransmitter system (see **Section 2.1.3.5**) than the dopaminergic neurotransmitter system.

2.1.7 Involvement of Serotonin in Reward Processing

Like the complexity of dopamine in reward and punishment processing serotonins' role in rewards and punishments seems equally complicated due to controversial findings (Cools et al., 2008). Over the last years different proposals for the role of serotonin in reinforcement learning have been made (Daw, Kakade and Dayan, 2002; Dayan and Huys, 2008 Rogers, 2010; Boureau and Dayan, 2010; Cools, Nakamura, Daw, 2010; Kranz, Kasper, Lanzenberger, 2010). Some of the early evidence from electrical stimulation studies of rat serotonergic nuclei (medial raphe nuclei) showed reinforcing effects for lever pressing similar to that of stimulation of ventral tegmental (VTA) dopamine circuitry (Miliaressis et al., 1975; Rompré & Miliaressis, 1985; Rompre and Boye, 1989). Moreover, these common reinforcing effects of serotonin and VTA stimulation could be explained by the neuroanatomical evidence of direct projections from serotonin neurons to dopamine neurons (Parent et al., 1981), which might regulate dopamine neurotransmission and cause these reinforcing effects (Kapur & Remington, 1996; Alex and Pehek, 2007). However, it is not perfectly clear whether serotonin has a direct role in reinforcing behaviour or it influences reward processing indirectly via its connections with dopaminergic and GABAergic neurons (Liu and Ikemoto, 2007).

On the other hand, the serotonin systems' inhibitory effects on behaviour are well studied in the literature (Soubrie, 1986; Dayan and Huys, 2008) and shown to be opposite to that of dopamine, in that administration of dopamine increases appetitive learning, whereas administration of 5-HT serotonin agonist decreases motivation for appetitive learning (Curzon, 1990; Cools et al., 2008). In psychiatry for example, selective serotonin reuptake blockers (SSRI) are commonly used for the treatment of impulse control disorders (Buhot, 1997; Robbins, 2000). More recently, fMRI studies showed that increase in 5-HT levels causes participants to increase the amount of delay gratification and make the participants choose the option that delivers bigger later rewards rather than small immediate rewards(Tanaka et al., 2007; Schweighofer et al, 2007) and oppositely decreased 5-HT levels makes the participant's more impulsive (Wogar et al., 1993; Bizot et al., 1999). The opponency of dopamine and serotonin can be well explained by the studies, which showed that stimulation of raphe nuclei have inhibitory effects on substantia nigra dopamine neurons (Drayet al, 1976; Tsai, 1989; Trent and Tepper, 1991) which might be involved in delay gratification. Additionally a very important empirical finding in the last few years is that of Nakamura et al., (2008) which compared firing patterns of neurons in the dorsal raphe nucleus (DRN) and substantania nigra in primates during a learning task. They showed that serotonin neurons calculate the size of the reward outcomes by spiking in tonic fashion whereas dopamine neurons calculate a reward prediction error signal by making phasic firings as predicted by the prediction error learning model (see, Cools et al., 2011 for a discussion). This study in summary showed direct evidence that monkey dorsal raphe nucleus is involved in reward processing by coding the expected and received rewards with a quite different pattern of firing than that of dopamine neurons.

2.1.8 Involvement of Serotonin in Punishment

Serotoninergic activity has long been considered to be crucial for aversive systems and thought to mediate learning about negative events (Bari et al, 2010; Daw et al, 2002; Deakin and Graeff, 1991; Evers et al, 2005). Several studies suggested a specific role for 5-HT in punishment prediction error (Daw et al., 2002). Based on their computational models Daw and his colleagues suggested that tonic and phasic serotonin have different functions. They proposed that a tonic serotonergic signal might report a long-run average reward rate, which was by the later evidence of Nakamura et al (2008) and a phasic serotonin signal might report prediction-errors for future punishments. Moreover a later study by Dayan and Huys (2008) suggested that reductions in serotonin activity (experimental or clinical) could produce increases in the size of negative prediction errors, which might be effective in people who experience major depressive disorder. Thus evidence exists for two different roles for serotonin in aversive processing. Firstly, there is an extensive literature on the role of selective serotonin reuptake inhibitors (SSRIs) in the management of clinical depression (Deakin and Graeff, 1991; Cools et al, 2008; Esher and Roiser, 2010) as well as chronic and neuropathic pain (Sawynok et al. 2001, Sommer 2004) where both of them are characterized by increased aversive processing of negative stimuli (Clark et al., 2009; Crockett et al., 2009; 2012). Secondly, tryptophan depletion can improve the accuracy of predictions of negative or punishing outcomes in healthy adults (Cools et al, 2008). Moreover, Evers et al (2005) showed that tryptophan depletion enhances neural activity in response to errors during reversal learning within the anterior cingulate region, an area that is activated while making decisions to stop chasing losses (Campbell-Meiklejohn et al, 2008). Thus, it was shown that tryptophan depletion in healthy adults enhances the salience of bad outcomes during gambling tasks (Cools et al., 2008). As a summary, in appetitive learning the serotonin system seems to work as a motivational opponent to the dopamine system, but its role in aversion behaviour is in fact similar to that of dopamine D_2 receptor agonists. Such that decreasing the amount of serotonin by selective trytophan depletion increases the effectiveness of punishments and punishment prediction errors and increasing the serotonin decreases punishment sensitivity.

2.1.9 The Extended Reward and Punishment Networks in Humans

Until this section I reviewed the role of dopamine and serotonin in processing rewards and punishments and summarized their role in calculating computational learning signals. In the following sections, I will summarize the role of gross neuroanatomical structures that are involved in reward and punishment processing.

Functional neuroimaging studies involved in winning or losing monetary outcomes become an important research topic for understanding decision making and reinforcement learning in humans. There is a wide range of experiments conducted to understand the brain areas that are involved in neural processing of monetary rewards and punishments. Generally, this aspect has been described in terms of the outcome valence (Knutson and Greer, 2008). For example monetary gains have positive valence and monetary losses have negative valence. Recent studies showed controversial findings regarding gain and loss related activity in the human brain (for a review see, Knutson and Greer, 2008). Some studies showed that monetary gains and losses activate a similar fronto-striatal network (Dreher, 2007; Gottfried et al., 2003; Marco-Pallares et al., 2007; Nieuwenhuis et al., 2005; Tom et al., 2007; van Veen et al., 2004) whereas other studies suggested that gains and losses are processed by a more distributed network (Frank et al., 2004; Wrase et al., 2007; Yacubian et al., 2006). The proponents of the latter theory suggest that positive outcomes are correlated with prefrontal and basal ganglia and negative outcomes are correlated mostly with amygdala and insula (Frank et al., 2004; Wrase et al., 2007; Yacubian et al., 2006). To resolve this debate Liu et al., (2011) performed a meta-analysis of 140 studies that included anticipation and retrieval of reward and punishing feedback. They found that during the outcome or the feedback period NAcc was activated by both positive and negative outcomes across various stages of a task (e.g., anticipation and outcome). However, the medial OFC and PCC responded more to the reward outcomes, whereas the ACC, bilateral anterior insula, and lateral PFC selectively responded to negative outcomes (see **Figure 2.5a**). Additionally when they categorized the activations based on anticipation and outcome they showed that during the anticipation period the activity mostly occurs in the cingulate cortex regardless of the outcome valence but during the outcome period the activity occurs in ventro-medial prefrontal cortex (see **Figure 2.5b)**.

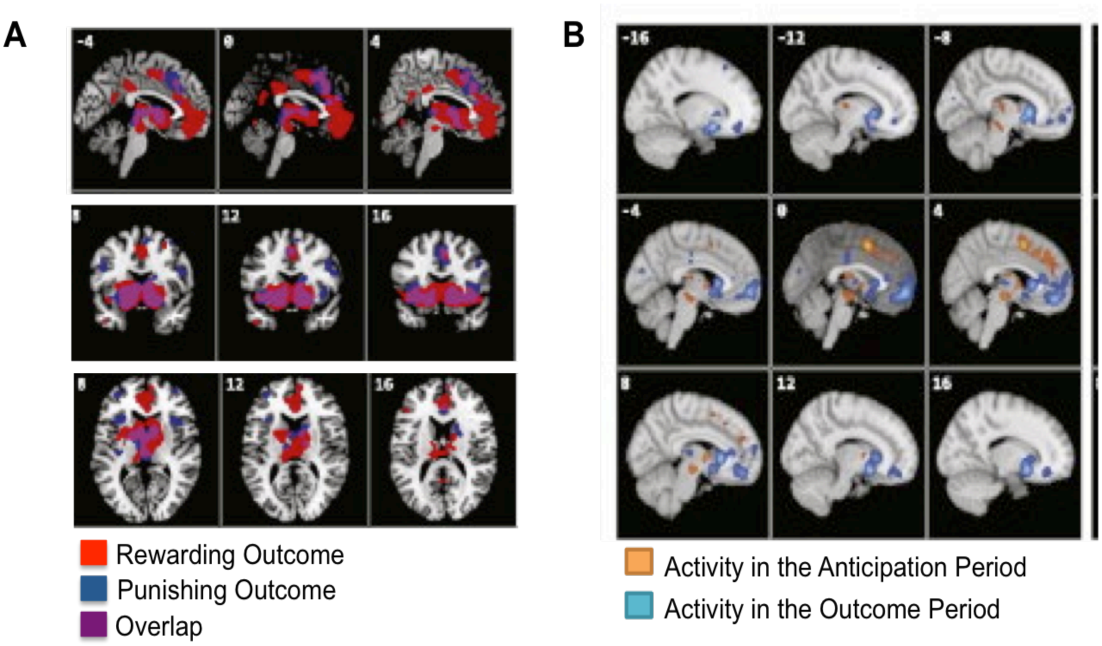


Figure 2.5 Results of the fMRI meta-analysis of Liu et al., (2010). a) Bold activations for reward outcomes (red) and loss outcomes (blue) shown in sagittal, coronal and axial views. Regions indicated as Purple are overlapping regions for gain and loss outcomes. b) Activations shown in blue are involved in outcome processing and activations shown in orange involved in anticipatory processes. Both Figures are adapted from Liu et al., (2011).

As the Liu et al., (2011) study showed, there is substantial mount of segregation between the processing of reward and punishment information. In the following sections the role of these distinct regions are reviewed in terms of their possible role in reward and punishment processing.

2.1.9.1 The Ventral Valuation Network

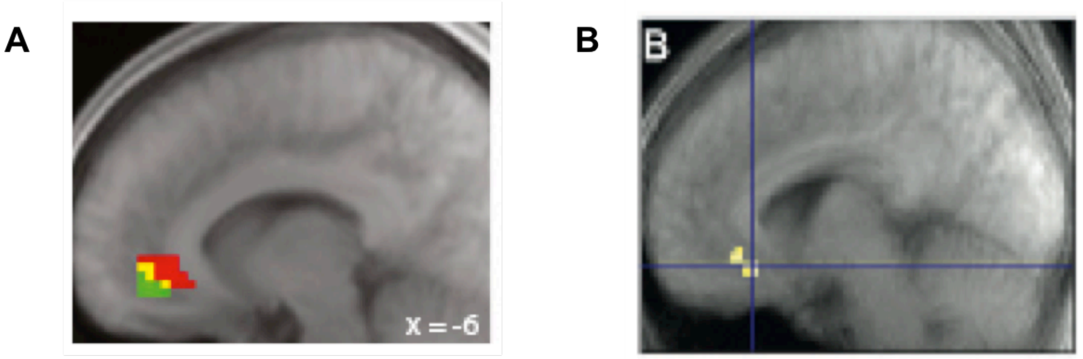
According to Brodmann (1909) the human map of the medial prefrontal cortex is occupied mainly by the Brodmann areas 24, 25, 32 and 10, where as the Brodmann areas including 11, 13 and 14 are considered as ventro-medial frontal cortex (see **Figure 2.8a** and **Figure 2.8b**). In this section and until the rest of this chapter medial and ventro-medial frontal cortex refers to the region encompassing both medial OFC and adjacent ventral medial PFC and they are used inter-changeably (see, Ongur et al., 2000 and

Fellows, 2007 for a discussion). Moreover Petrides and Pandya (1994) classified Brodmann areas 47/12 as lateral-orbito frontal cortex. In the studies reported herewe used the term lateral-orbito frontal cortex inter-changeably with ventro-lateral prefrontal cortex, which refers to Brodmann areas 47/12. It is important to note that both the medial frontal cortex and the adjacent-lateral frontal regions are thought to play crucial roles in choice behaviour and these regions affect particular aspects of reinforcement learning as discussed in detail below.

The general term used to refer to the brain areas that process the reward values of reinforcers (both primary and secondary) are called the 'ventral valuation network' (VVN) (Montague & King-Casas, 2006). Montague and colleagues proposed this term to refer to orbito-frontal cortex, striatum and ventro-medial prefrontal cortex. In a meta-analysis of imaging studies, when participants were instructed to maximize their outcome (e.g., money) in a learning experiment, significant activity is found in the brain regions that are involved in the ventral valuation network (Liu et al., 2011). One of the brain regions in the ventral valuation network, the ventral striatum contains the main projection sites of the midbrain dopaminergic neurons and show activation for almost all types of salient stimuli such as pleasant odors (Blood et al., 1999), sport cars (Erk et al., 2002), or pleasant music (Blood & Zatorre, 2001). In addition to that striatum is also activated by the feelings of love and trust (Bartels & Zeki 2004, King-Casas & Tomlin, 2005).

Even though I discussed earlier in this chapter that the striatum responds similarly to rewards and punishments due to dopamine neurons' excitatory response to both positive and negative reinforcers (e.g., the neural signals for the motivationally saliency as discussed in the earlier section which refers to Bromberg-Martin et al., 2010) the story for the orbito-frontal cortex is a bit more complicated because controversial findings have been observed in the orbito-frontal cortex for the outcome related activity.

We think that the reason for this controversy is because during the outcome phase the neural responses in the orbito-frontal cortex might differ before and after learning due to the presence or absence of goal values, predicted values and prediction error signals. For example it has been shown that orbito-frontal cortex is actived by both reward predicting CS and the actual US (O'Doherty, 2011) and it is not only involved in outcome related activity (goal values) as suggested by the Liu et al., (2011) meta-analysis. For the CS related activity, human brain imaging studies showed that medial-frontal cortex increased activity for coding the reward value of primary and secondary reinforcers during the time of decision making and it was suggested that the type of computation is closer to value predictions than prediction errors (Mainen and Kepecs, 2009). Supporting this idea several studies showed activity in the ventromedial frontal cortex for value prediction. For example, Wunderlich et al., (2010) showed activity in the orbito-frontal cortex for the predicted value signal of the chosen stimulus (a post-decision signal) before the actual action was taken (see **Figure 2.6a**). In addition activation caused by a predicted value signal in OFC seems unaffected by the type of reinforcer (primary or secondary), which was further suggested by Chib et al., (2009) (see **Figure 2.6b**).



Chosen stimulus value (before action)

Conjunction for the expectation of fruit juice and money .

Figure 2.6 Predicted value coding in the VMPFC correlates of the chosen stimulus value during CS presentation. a) Activity in red shows predicted value when no action information is available (a type of Pavlovian predicted value signal) and activity in green shows when the action

information is available (a type of instrumental predicted value signal) and yellow shows the overlap. Figure is taken from Wunderlich et al., (2010). b) Predicted value of the CS that is predicted by a primary reward (juice outcome) and a secondary reward (the monetary outcome). The result of the conjunction analysis (yellow regions) shows that expecting either juice reward or monetaryreward both activates the medial frontal cortex during CS presentation. Figure is taken from Kim et al., (2011).

In addition to the findings reported above, orbito-frontal cortex and striatum also code the value of more than one conditional stimulus in the form of an expected value signal. For example if there are tworeward-predicting stimuli presented then striatum will elicit neural activity for both of these conditional stimuli. This was calculated as the weighted sum of the values for all choice options (e.g., Expected Value = Probability of the occurrence of CS$_1$* Value of CS$_1$ + Probability of the occurrence of CS$_2$* Value of CS$_2$. Note that expected values and also predicted values of chosen options are usually calculated with reference to a computational model and inserted as a parametric variable, at the time of cue onset, in regression analysis used in fMRI studies (see, **Chapter 5** for details). From a theoretical perspective, expected values are considered in psychology as being the Pavlovian response to the CS in instrumental studies and in computational models (see **Chapter 4**) as state values. Nevertheless, recent studies showed that expected value signalsare correlated with the activity in the ventro-medial frontal cortex and the striatum (see **Figure 2.7**).

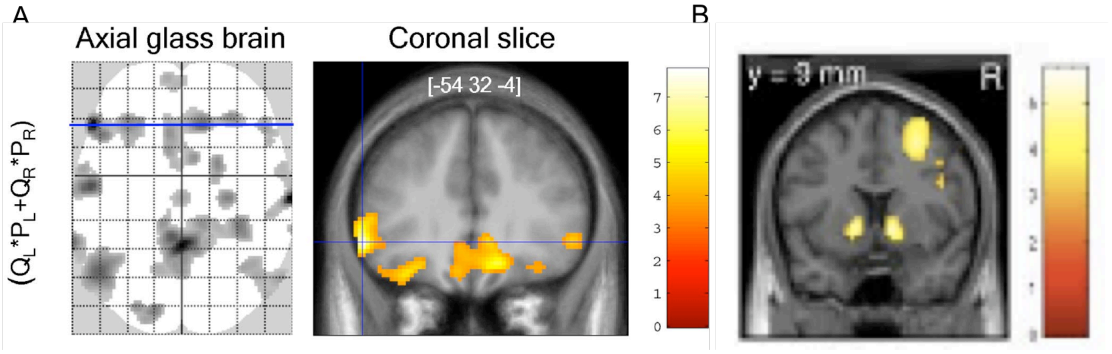


Figure 2.7 Statistical parametric maps of the activity that correlates with expected value signals. a) Expected value during the presentation of CS that indicates two available options in an instrumental conditioning task. Activity for expected value (Q$_L$ refers to value of the left option and Q$_R$ refers to the value of right option where as P$_L$ refers to the reward probability of the left option, P$_R$ refers to reward probability of right option) that correlates with the ventro-medial and

the lateral-frontal regions. *Figure 2.6a*taken from Palminteri et al., (2009). b) Activity correlates with expected value of four options in a decision making task. Figure is taken from Yacubian et al., (2006).

2.1.9.2 Reward and Punishment in Medial and Ventro-Lateral-Prefrontal Cortex

Previous studies showed that rewarding outcomes primarily activate regions of the medial prefrontal cortex and orbitofrontal cortex in general (Haber & Knutson, 2009; Liu et al., 2011). This view was supported by the electrophysiological studies in monkeys, which showed spiking activity in various frontal regions. For example, during the evaluation of rewarding outcomes, spiking neural activations were mainly found in the monkey anterior cingulate cortex (ACC), orbitofrontal cortex (OFC) and the lateral-prefrontal cortex (LPFC) (Niki & Watanabe, 1979; Hikosaka & Watanabe, 2000; Tremblay & Schultz, 2000; Walton et al., 2004; Sallet et al., 2007). These studies showed that frontal neurons are sensitive to the value of outcomes across multiple decision variables (Hikosaka & Watanabe, 2000; Rolls, 2000; O'Doherty et al., 2001; Amiez et al., 2006; Roesch et al., 2006). Moreover, neuropsychological studies also showed that damage to either OFC or ACC impairs the ability to utilize the value of an outcome in a decision making task (Shima & Tanji, 1998; Kennerley et al., 2006; Murray et al., 2007).

In two recent reviews the distinction between ventro-lateral and medial-orbito frontal regions were compared in detail in terms of their sensitivity to reward and punishment processing (Grabenhorst and Rolls, 2011; Kringelbach and Rolls, 1994). Both of these studies showed that lateral regions code for unpleasant outcomes such as unpleasant smells or painful stimuli but the medial regions are more sensitive to pleasant outcomes such as nice odours (see **Figure 2.8c** and **Figure2.8d**).

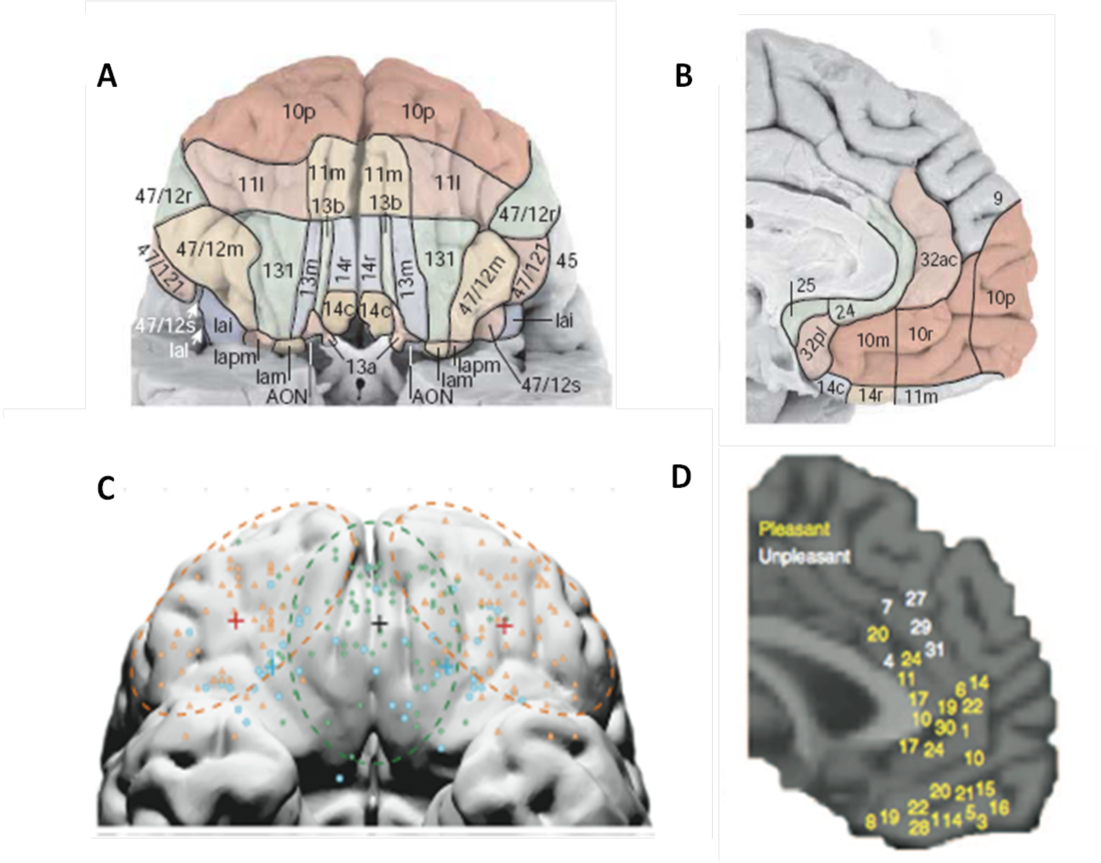


Figure 2.8 a) Broadmann map of the prefrontal cortex (axial view) and b) Broadmann map of the medial frontal cortex (sagittal view). Both Figure 2.7a and Figure 2.7b is taken from Ramnani and Owen (2004) cross citation from Ongur et al. (2003) c) Meta-analytic review of the orbitofrontal cortexfor reward and punishment outcomes. Areas circled with orange colour indicate the regions that are sensitive to punishments**.** Areas circled with green colour indicate the regions sensitive to rewards. Figure 2.7c is taken from Kringelbach (2005) and refers to the original study of Kringelbach and Rolls (2004) by cross citation. d) Numbers shown in yellow represents the studies, which found significant activation for pleasant outcomes, numbers shown in white represents the studies, which found significant activation for unpleasant outcomes in those particular brain regions. *Figure 2.7d* is taken from Grabenhorst and Rolls, (2011).

Although fMRI studies (Grabenhorst and Rolls, 2011; Kringelback and Rolls, 2004) have shown that lateral regions of the orbitofrontal cortex are more engaged with losses whereas medial regions are engaged with rewards it is important to mention that in the meta-analyses of both Kringelbach and Rolls (2004) and Grabenhorst and Rolls, (2011) the activations are not sensitive to the sign of the bold signal (e.g., increase or decrease of activity. These arguments lead to the suggestion that information about rewards and punishments is either encoded in a common striato-frontal circuitry albeit with increased/decreased activity respectively (Elliott et al*.,* 2003; Knutson et al*.,* 2007; Tom et al.,2007), or that gains and losses are processed in different brain regions as mentioned earlier (Yacubian et al.,2006; Liu et al., 2007; Seymour et al.,2007).

The functional connectivity analysis using the ventral striatum as a seed region revealed a topographically overlapping subcortical-limbic-anterior prefrontal network for monetary gains and losses (Camara et al., 2009a). In their functional connectivity analysis study Camara et al., (2009a) showedthat a network including the OFC, the insular cortex, the amygdala, and the hippocampus correlated with activity observed in the seed region (ventral striatum) for processing of gains and losses. However, although they showed that both gains and losses correlate with overlapping regions in the connectivity analysis (see **Figure2.9a**), when they performed a region of interest (ROI) analysis they showed that OFC decrease activity for losses (see **Figure2.9b**). This indicates that a region might be involved in a task by showing decreased activity.

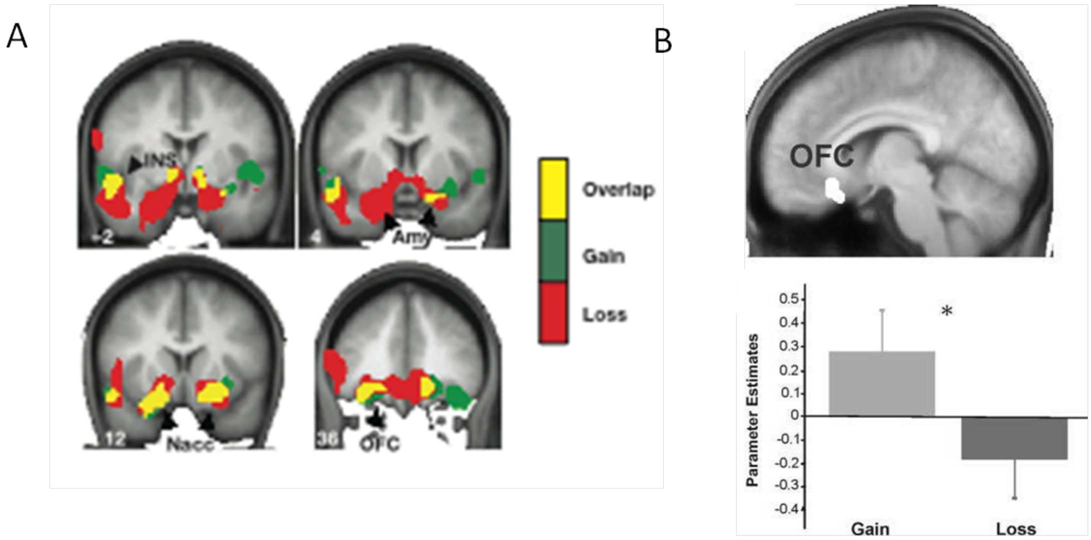


Figure 2.9 a) The results of the functional connectivity analysis of Camara et al., (2009a). Areas show significant connectivity with ventral striatum during the processing of monetary gains (green regions) and monetary losses (red regions) and yellow regions are where both monetary gains and losses are overlapped. The results suggest that processing of monetary losses also engaged with medial orbitofrontal structures. Figure taken from Camara et al., (2009a). b)The results of the univariate analysisof Camara et a., (2008). When they compared gain and the loss condition in the OFC cluster identified in the functional connectivity analysis. They foundincreased BOLD signal for gains, and a decreasedbold signal for losses. Figure is taken from Camara et al., (2009a).

In fact the results of Camara et al., (2009a) is well supported by the early orbito-frontal lesion patients. Bechara *et al.*, (2000) refer to those patients, as having 'insensitive to punishment' and use this to explain their poor performance in the Iowa gambling task (also see for a discussion, Wheeler and Fellows, 2008). Orbito-frontal patients in an Iowa gambling task tend to perseveratively choose the high-risk option despite their higher risk of loss. Further follow up studies with variations of the Iowa gambling task showed that orbito-frontal patients suffer from failing to accurately predict the outcome of a decision (general myopia for the future) rather than evaluating the outcome itself (Bechara *et al.*, 1997; Fellows and Farah, 2005). It was suggested that they may be able to predict unexpected outcomes, but they are unable to predict whether the outcome is going to be positive or negative (Fellows and Farah, 2005; Fellows, 2006). **Figure 2.10a** shows the region of the lesion in orbito-frontal cortex of the famous patient Phineas Cage and **Figure 2.10b** shows the lesion sites from another study of 10 subjects with ventromedial frontal damage who are impaired in learning with negative feedback (Wheeler & Fellows, 2008).

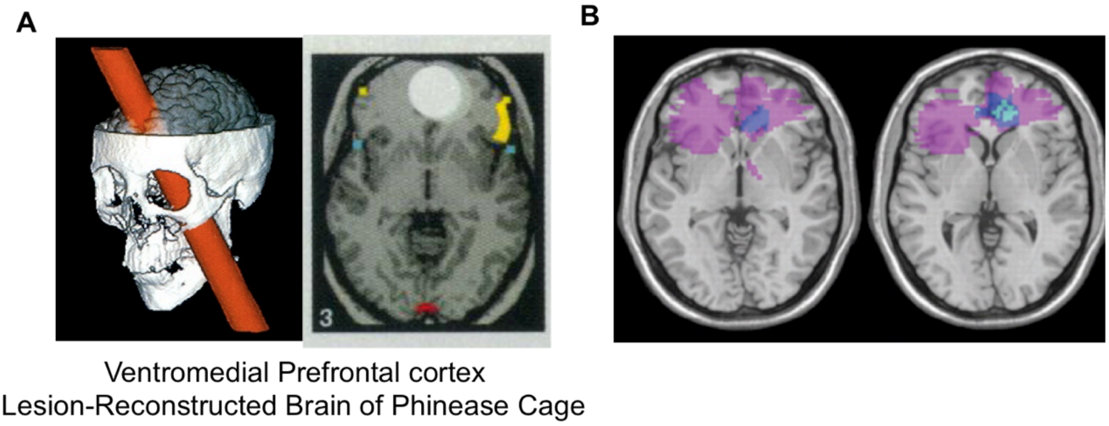

A
B
3
Ventromedial Prefrontal cortex
Lesion-Reconstructed Brain of Phinease Cage

Figure 2.10 a) MRI reconstruction of the famous patient Phineas Cage shows that the main lesion was in the medial-frontal cortex. Figure 2.9a is taken from Damasio et al., (1994) b)Figure shows the lesion sites of 10 patients with ventromedial frontal damage projected on the same axial slices of the standard Montreal Neurological Institute brain. Figure 2.9b is adapted from Wheeler & Fellows (2008).

These studies suggest that both medial and lateral structures are involved in processing loss information but to what extend they contribute to decision making is not well understood. It seems that when the reward learning data are analysed using a model-based way, the results show reward predicted-value activity during the presentation of CS (see, Figure 2.6) but when it is analysed with a standard GLM analysis CS related activity is observed in the cingulate cortex. Furthermore, it is important to note that not finding a positive activity in orbito-frontal cortex in an fMRI study shouldn't lead to the conclusion that it is not involved in processing the loss information but a region might contribute to processing by significantly decreased activity. As it will be revealed in the next section it is possible that during the outcome processing other regions might be involved in processing reward and loss information as suggested by Camara et al., (2009b).

2.1.10 Avoidance Circuitry

2.1.10.1 Reward and Punishment in the Cingulate Cortex

Early studies on the functional segregation of anterior cingulate cortex argued that there is a separation of anterior and posterior regions in terms of emotional and cognitive processing (Devinsky et al., 1995 Bush et al., 2000) However, recent studies based on the cyto-architectonic microanatomy of cingulate cortex suggested that there are four distinct regions (Vogt, 2009). According to this cyto-architectonic division anterior and midcingulate cortex are divided into two further categories. The midcingulate cortex is divided into two parts, anterior (aMCC) and posterior mid-cingulate (pMCC) and anterior cingulate cortex is divided into perigenual (pgACC) and subgenual (sgACC) (see **Figure 2.6a**). Recent fMRI studies showed that aMCC is consistently active for negative affect(Price, 2000), pain(Peyron et al., 2000; Vogt, 2005, Farrell et al., 2005; Rainville et al., 2010) and cognitive control (Yeung et. al., 2004; Cole et al., 2009). Given that aMCC receives nociceptive information from the spinothalamic system (Dum et al., 2009) and has reciprocal connections with amygdala (Ghashghaei et al., 2007; Morecraft et al., 2007) and insula (Mesulam, M. M. & Mufson, 1982; Cauda et al., 2011) it is meaningful to find aMCC activity for anticipation of aversive stimuli (Peyron et al., 2005; Vogt, 2005). Moreover aMCC receives dopaminergic inputs from substantia nigra and ventral tegmenral area (VTA) making its role important in processing of negative reinforcers as well as cognitive control (Williams & Goldman-Rakic, 1998).

According to adaptive control hypothesis of Shakman et al., (2011) cognitive control is an early warning system that allows the subject to proactively alter attention and actions to avoid future errors and perhaps make the aMCC involve in coding punishment prediction errors. They suggested that cognitive control and negative affect show high functional similarity and are both coded in aMCC (Shakman et al., 2011). A recent fMRI meta-analysis showed that aMCC is involved in processing not only negative affective situations such as anger, sadness, fear or painful stimulation of skin but also monetary loss (Liu et al., 2011). Moreover, Fujiwara et al., (2009) showed that there is overlap between regions coding monetary reward outcomes and states of happiness (see **Figure 2.11b** and **Figure 2.11c**) suggesting that pgACC is involved in positive affect but aMCC is involved in negative coding monetary losses and negative affect.

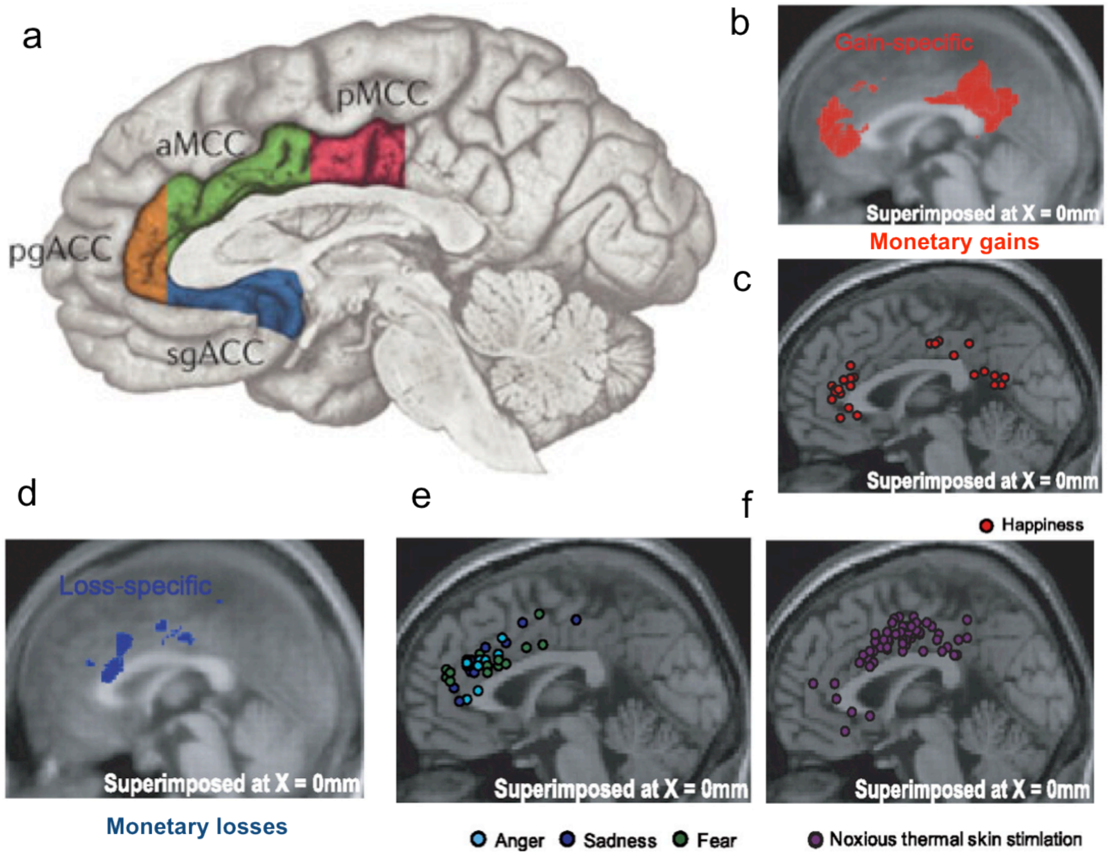


Figure 2.11 a) Four major sub-divisions of the rostral cingulate cortex. Blue regions show subgenual anterior cingulate cortex (sgACC), an orange region shows pregenual-anterior cingulate cortex (pgACC), green region shows anterior medial-cingulate cortex (aMCC), and magenta shows posterior medial-cingulate cortex (pMCC). Figure 2.11a is taken from Shackman et al., (2011). b-c) Meta-analysis of fMRI studies that found activity for (b) monetary gain outcome and (c)happiness, respectively. d-f) Meta-analysis of fMRI studies that found activity for(d) monetary loss, (e)anger, sadness, fear and (f)noxious thermal skin stimulation. Figures from (b) to (f) taken from Fujiwara et al., (2009).

2.1.10.2 The Role of Amygdala in Reward and Punishment

Many studies recognize amygdala as one of the most important brain structures associated with fear and aversive learning (Gallagher and Chiba, 1996; Klüver and Bucy, 1939; LeDoux, 2000; Maren and Quirk, 2004; Murray, 2007 Buchelet al., 1998; LaBaret al., 1998; Schilleret al., 2008; Daviset al., 2010). Anatomically, it consists of at least three anatomically distinct nuclei that comprise two central nuclei and a basolateral nucleus (Alheid, 2006, Olmos & Heimer, 2006). Several imaging studies showed that amygdala is involved in loss aversion (Tom et al., 2007; Dreher, 2007). In fact one recent study showed in two rare patients,with bilateral amygdala lesions, significant impairment in processing loss aversion compared to matched controls in a monetary gambling task (De Martino et al., 2010).Moreover in recent high-resolution fMRI studies the contributions of amygdala subregions have been shown in associative learning. During reward and avoidance learning Prevost et al., (2011) showed that there is a dissociation between the contributions of the basolateral and centromedial during learning, with the basolateral complex contributing to reward learning by coding action values for reward outcome, and the centromedial complex more to action values for avoidance learning (Prevost et al., 2011) (see **Figure 2.12**).

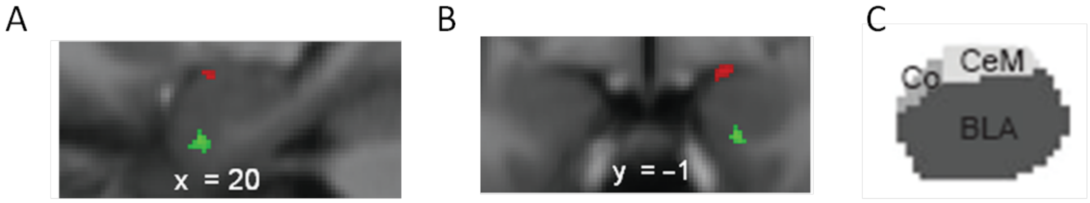


Figure 2.12 a-b) BOLD signals correlating with the expected reward value of the chosenactions.Basolateral complex shows significant activity for action values in the reward condition(in green) and the centromedial complex for action-values in the avoidance condition (in red). c) Anatomical ROI's for the separation of centromedial (Co & CeM) and basolaeral amygdala that is used by the study of Prevost et al, (2011). Both Figures were adapted from Prevost et al., (2011).

Previous fMRI studies have also reported computational learning signals in amygdala such as expected outcome, prediction errors and learning rate during performance of similar tasks (Elliottet al., 2004; Seymour et al.,2005; Yacubianet al., 2006; Hampton et al., 2007; Li et al., 2011). These studies suggest that amygdala as a whole can contribute to both aversive and appetitive learning with its distinct anatomical subregions it can carry out multiple computations (Baxter and Murray, 2002).

2.1.10.3 The Role of Insular Cortex in Pain and Avoidance Learning

Insular cortex is broadly acknowledged as viscerosensory cortex, and implicated in mapping internal bodily states (including pain and taste) and in representing emotional arousal and feelings (Price, 2000; Critchley et al., 2004; Craig, 2003, 2009). Also for many years insular cortex was considered to be included as a major part of the pain pathway (Peyron et al., 2000). According to this view when there is an injury in the body peripheral nerves send signals to the central nervous system through spinal cord with specific neurons via the dorsal horn of the spinal cord (Craig, 2002). These nociceptive signals then ascend to various brainstem nuclei and mainly to the thalamus (see **Figure 2.13a**). After thalamus, many areas of the brain receive this nociceptive information that is coming from the peripheral system and therefore a variety of brain regions are involved in pain processing. These regions mainly include somato-sensory cortex, insular cortex, anterior cingulate and orbito-frontal cortex and subcortical areas including amygdala, and hippocampus (Jones et al., 1992; Hutchison et al., 1999; Leknes & Tracey, 2008).

Accordingly there have been several suggestions that insular cortex supports different levels of representation of both current and predictedsubjective states and that it calculates error-based learning signals for the feeling states (Singer et al., 2009). In fact several studies did shown a risk prediction error signal about the outcomes (Preuschoff et al., 2008) (see **Figure 2.13b**) and punishment prediction error (Pessiglione et al., 2006) (see **Figure 2.13c**).

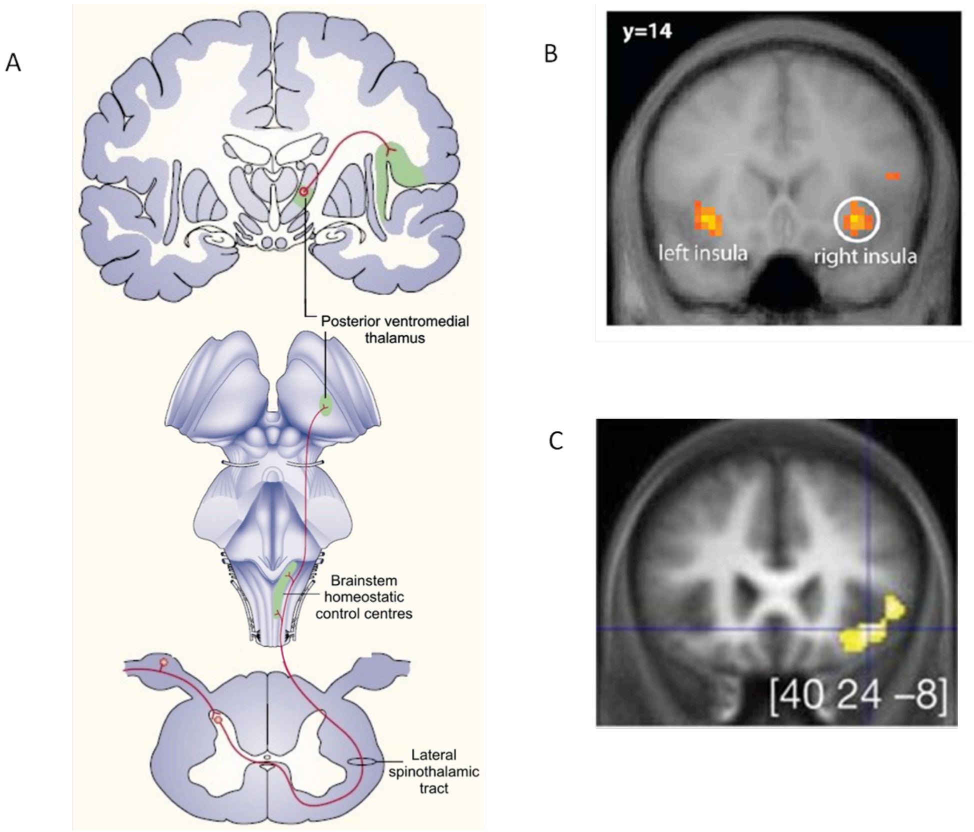

A
Posterior ventromedial thalamus
Brainstem homeostatic control centres
Lateral spinothalamic tract
B
y=14
left insula
right insula
C
[40 24 –8]

Figure 2.13 a) Nociceptive signals that ascend from spinal cord of lamina-1 and projects to posterior ventromedial thalamusand from there they ascend to insular cortex and other cortical regions. Figure is taken from Gray and Critchley (2007) b) Activation in bilateral insula correlates positively with risk prediction errors in a gambling task. Figure 2.13b is from Preuschoff et al., (2008) c) Activation of the insula for the punishment prediction error. Figure 2.13c is taken from Pessiglione et al., (2006).

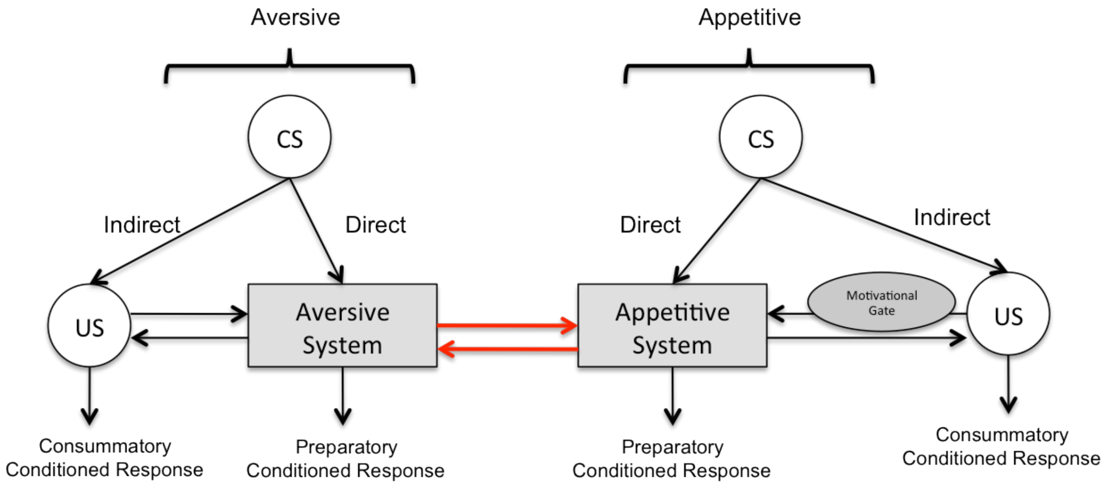
Finally, neuroimaging studies showed that activity in anterior insula cortex and anteriorcingulate cortex in response to the perception of unpleasant stimuli reflects the experience of expecting an aversive events (Brownet al., 2008;Wager et al., 2004; Hester et al., 2010).

2.1.11 Opponent Process Theory

The two main theoretical frameworks, which suggest an explanation for the connection between avoidance and approach behavior, are two-factor theories learning (Mowrer, 1939; Mowrer & Lamoreaux, 1942) and opponent process theories (Solomon, 1974; Dickenson & Dearing, 1979) of conditioning. In this thesis I am mainly concerned with opponent process theory.

Opponency has an important history in psychology and neuroscience (Grossberg,1988). In the classic model of Opponent Process Theory there are two systems for representing affective events (Dickinson and Balleine, 2002). The first system is responsible from appetitive events (positive affects), and the second system is responsible from aversive events (negative affects), and there is a link between those two systems that inhibits each other. According to opponent process model of Konorski the appetitive and aversive systems work oppositely to each other and it has been stated that this opposition "gives rise to four basic categories of motivation" which are "prediction of reward (hope), prediction of aversive events (fear), omission of reward (frustration) and omission of aversive events (relief)" (Seymour et al., 2003, p: 18).

In the Konorskian model, there are two types of representations (i) a stimulus non-specific representation and (ii) a stimulus specific representation. In the stimulus non-specific representation the identity of the outcome (e.g., shock or airpuff) is irrelevant and only the valence of the outcome (e.g., aversive or appetitive) is important. On the other hand in the stimulus specific representations the natureof the outcome is crucial (e.g., air-puff to the eye is crucial for eyeblink conditioning but not for electric shock). The basic architecture of these distinct representations is shown in **Figure 2.14**.



Aversive
Appetitive
CS
CS
Indirect
Direct
Direct
Indirect
US
Aversive System
Appetitive System
Motivational Gate
US
Consummatory Conditioned Response
Preparatory Conditioned Response
Preparatory Conditioned Response
Consummatory Conditioned Response

Figure 2.14 Konorskian model of Pavlovian appetitive conditioning, showing direct andindirect pathways mediating representation of conditioned stimuli (CS) and unconditioned stimuli (US). Figure 2.14 is adapted from Dickinson and Balleine, 2002).

2.2 *Interim Summary*

In this chapter I showed that there are various types of rewards where someof them are species-specific primary reinforcers (e.g., taste of sweetness) where as others are species-specific secondary reinforcers (e.g., money for humans). In humans various brain regions engagein learning the value ofthese reinforcers, which are usually either appetitive or aversive and only few regions like striatum and medial orbitofrontal cortex seem to be involved in both rewards and punishments. In short, these regions include medial frontal cortex, which process both reward and punishment expectation and retrieval, striatum, which processes reward and punishment expectation and retrieval, amygdala, processes mostly punishment expectation and retrieval, insula, processes mostly punishment expectation and retrieval and finally anterior cingulate cortex is involved in processing mostly punishment expectation. This suggests that not a single brain region is involved in processing rewards and punishments but various brain regions process various aspects of rewards and punishments and perhaps in synchrony. It is also important to note that other brain regions such as frontal eye fields (FEF) (see for a review Kable & Glimcher, 2009; Glimcher 2009a, 2009b) and lateral intraparietal cortex

(LIT) (Platt & Glimcher, 1999; Gold & Shadlen, 2007) also engage in decision making and reinforcement learning with saccadic choices but they were excluded from this review.

Furthermore, I summarized the role of dopamine and serotonin in reward and punishment processing. According to that dopamine is not only involved in processing reward information (Schultz, 2006) but also information about the painful stimuli (Wood et al., 2007; Wood et al., 2009). The best example for this in humans is the involvement of dopamine in a chronic pain disease called Fibromyalgia, which is characterized by widespread pain and bodily tenderness in the body (Wood et al., 2007; Wood et al., 2009). Accumulating evidence indicates that fibromyalgia in humans is caused by the decrease of dopamine levels in the brain (Hagelberg et al., 2004; Pertovaara et al., 2004; Wood et al., 2007; Wood et al., 2009). These studies suggest that during the experience of painful events release of dopamine might decrease the experience of pain in individuals. Beyond dopamine and serotonin, a wider network of neurotransmitters are also involved in learning and prediction of rewards, which includes networks of epinephrinergic neuromodulators (see for review, Doya, 2008) and oxytonergic system (Zak, 2007) excluded from the review because the computational underpinnings of such modulators are not well understood and beyond the scope of this thesis.

All sciences are now under the obligation to prepare the ground for the future task of the philosopher, which is to solve the problem of value, to determine the true hierarchy of values.

Friedrich Wilhelm Nietzsche

Chapter 3

Neural Mechanisms of Associative-Learning

3.1 Functional Neuroanatomy of the Basal Ganglia and Cortico-Striatal Maps

The basal ganglia is the name given to a collection of limbic structures and it is one of the most important of all brain structuresfor understanding complex reward and motor processing.Over many years basal ganglia has been studied extensively in order to understand the source of variousbrain disorders. The diversity of psychiatric and neurological disorders that basal ganglia is involved are huge in quantity such as schizophrenia (Carlsson, 1988; Kapur, 2003, Kapur et al., 2005; Howes and Kapur, 2009), attention deficit hyperactivity disorder (ADHD) (Volkow et al., 2009) as well as Parkinson's disease, Tourette's syndrome and Huntington's disease (Bhatia and Marsden, 1994;Albin and Mink, 2006; Gerfer and Surmeier, 2011). It has also been shown that basal ganglia is related to many executive, motor functions that have direct influence on our daily lives (Yin & Knowlton, 2006). Moreover, because the basal ganglia's unique anatomical position with its connections to almost the entire cortex, thebasal-ganglia are central to many cognitive processes including decision making, sequence learning, category learning, and probabilistic learning.

Anatomically, the basal ganglia are made up of different structures; the striatum, pallidum, subthalamic nucleus (STN) and substantia nigra. One of the most important sub-compartments of the basal ganglia for procedural learning is the striatum, which is crucial for theoretical themes as explained later in this chapter. The striatum is also composed of subcompartments: the caudate nucleus, putamen, and the nucleus accumbens. The caudate nucleus and the putamen are one structure at birth and are separated during development by the internal capsule (Holt et al., 1972).

In fact most of our knowledge about the functional neuroanatomy of the basal-ganglia and its structural connectivity to the cortex and thalamic structures are based on animal studies of *monkeys* and *rats*. Researchers used various techniques to understand the cyto-architecture of basal ganglia and cortex such as antero or retrograde neuronal tracers (e.g., rabie-virus) (Strick *et al.,* 1995; Middleton & Strick, 2000). More recently our knowledge about the structure of the *human* basal-ganglia and its internal loop architecture (cortico-striatal, striato-pallidal, striato-cerebellar) increased tremendously based on a the development of a broad range of techniques coming from various neuroimaging studies which used advanced statistical analysis (e.g., effective or functional connectivity analysis). In addition to that post-mortem dissection studies provide invaluable inputs to our understanding of the human basal ganglia and thalamus, which will be revealed in the following section (Morel, 2007).

Unlike prefrontal cortex, which consists of mostly glutamatergic neurons, the basal ganglia are mostly composed of inhibitory GABA (g-aminobutyric acid) containing neurons that are spiny in striatum (input) and aspiny in pallidum (Yin & Knowlton, 2006). These spiny neurons also contain different amounts of neuropeptites such as substance P and dynorphin or enkephalin where their relative amount to each other correlates with two types of dopamine receptor that are of type D_1 and D_2 (see for a discussion, Nadjar et al., 2006). It has been suggested that these dopaminergic neurons reduce the pallidal output, and facilitate the target motor networks by disinhibition (Deniau & Chevalier, 1985; Chevalier & Deniau, 1990). Moreover, the striatum controls motor movements through its connections with thalamus and cortex via direct and indirect pathways, which are facilitated by D_1 and D_2 receptors respectively (Albin et al, 1998). It has been suggested that the direct pathway contains the D_1 receptor subtype and facilitates approach behaviour whereas the indirect pathway that contains the D_2 receptor subtype facilitates avoidance behaviour (Albin et al., 1989; Gerfen et al., 1990; Kravitz et al., 2010; Hikida et al., 2010; Bromberg-Martin et al., 2010) (see **Figure 3.1**).

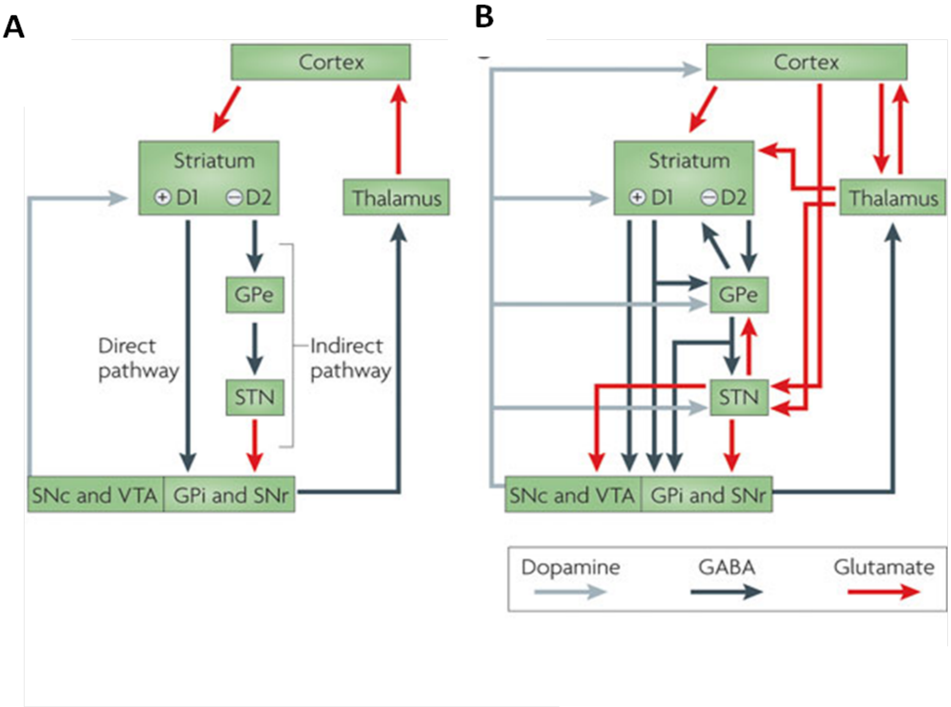


Figure 3.1 Box-and-arrow diagrams showing direct and indirect connections of the basal ganglia to cortex and thalamus. a) The classic basal ganglia architecture of Albin et al., (1989) shows the striato-midbrain-thalamo-corticol loop. **b)** A more recent update on the striatal microanatomy. It represents connections between sub-nuclei of the basal ganglia, the cerebral cortex and the thalamus that is based on Bolam and Bennett (1995). Both Figures are taken from Redgrave et al., (2010)

Recent theories on reinforcement learning suggest that phasic DA bursts during instrumental learning produce conditions of high DA, activate D_1 receptors, and cause the direct pathway to select high-value movements, whereas DA pauses produce conditions of low DA, inhibit D_2 receptors, and cause the indirect pathway to suppress low-value movements (Frank, 2005; Hikosaka, 2007). Consistent with this hypothesis, Bromberg-Martin et al., (2010) suggested that high DA receptor activation promotes potentiation of corticostriatal synapses onto the direct pathway and learning from positive outcomes (Shen et al., 2008, Frank et al., 2004; Voon et al., 2010), whereas the inhibition of D_1 receptor release impairs those movements with rewards (Nakamura and Hikosaka, 2006). Similarly low DA receptor activation promotes potentiation of corticostriatal synapses onto the indirect pathway and learning from negative outcomes (Shen et al., 2008; Frank et al., 2004; Voon et al., 2010), whereas blockation of striatal D_2 receptors suppresses movements to nonrewarded targets (Nakamura and Hikosaka, 2006). This division of D_1 and D_2 receptor compartmantalization in the basal ganglia explains many of the effects of dopamine on reinforcement learning (but see also **Chapter 2**, section *2.1.3.4* for a discussion on the effect of D_1 and D_2 receptor on learning with rewards and punishments).

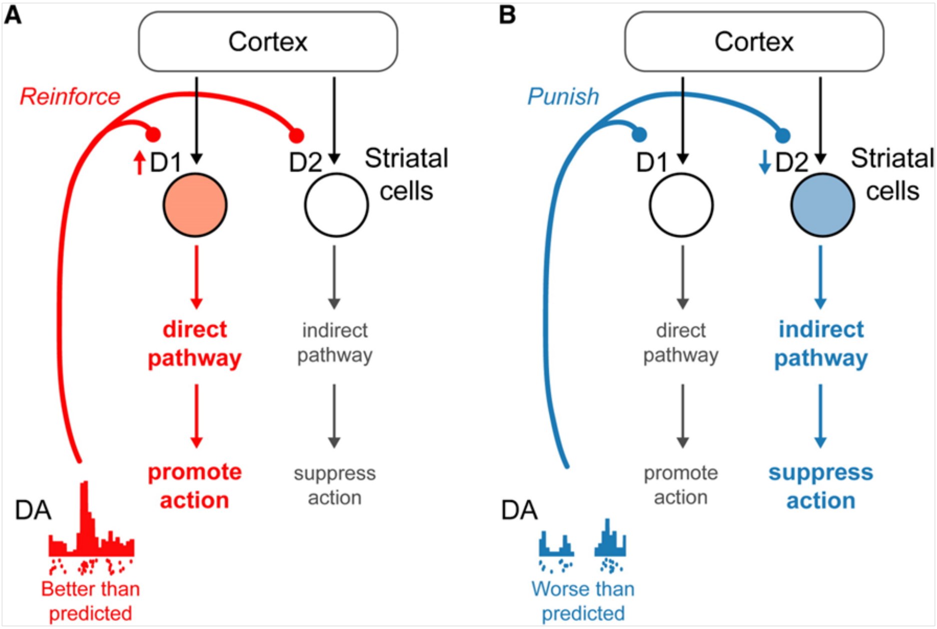


Figure 3.2 a) DA neurons fire a burst of spikes (positive prediction error) activate D1 receptors on direct pathway neurons, promoting selection of that action. b) DA neurons pause their spiking activity during negative prediction error promote suppression actions. Figure taken from Bromberg-Martin et al., (2010).

Our knowledge about the cyto-architecture of the basal ganglia and itsfunctional neuroanatomyhas advanced dramatically over the last couple of decades. Early studies showed that the basal ganglia coordinate the link between motivation and motor processes (Mogenson et al, 1980). In fact an early suggestion by Mogenson et al, (1980) proposed that the nucleus accumbens (an important structure in the striatum) works as the 'limbic-motor gateway' where motivational signals meet motor-actions. On the other hand there are a few, rare clinical cases, which also support this limbic-motor-gateway hypothesis. For example, a clinical syndrome called "apathy" (Marin, 1991; Chase, 2011) can be seen in patients who have significant loss of dopamine neurons in the basal ganglia (Starkstein et al., 1992; Isella et al., 2002; Pluck and Brown, 2002; Aarsland et al., 2005; Kirsch-Darrow et al., 2006; Levy & Czernecki, 2006), and this causes significant decrease in goal directed behaviours in these patients. There is also a more extreme form of apathy called PAP syndrome (Psychic-autoactivation-deficit coming from french perte d'auto-activation psychique, or loss of psychic auto-activation) which is characterized by the loss of will to execute motor actions and is directly linked with the effect of reduced levels of motivational signals triggered by the midbrain dopaminergic neurons. In the late, 90s French neurologist Laplane showed that PAP syndrome is commonly seen in patients whose major axonal terminals were damaged in the nigro-striatal pathway (Laplane et al., 1989; Laplane & Dubois, 2001; Levy & Dupois, 2006). Since these findings, the striatum has been studied extensively in many vertebrate species in order to understand how the basal ganglia with its cortical connections promote learning of rewards and actions in the instrumental conditioning context (Yin et al., 2005; 2006).

On the other hand a different body of research suggested that rewarding outcomes of actions and the actual motor processes that causes actions could be coded separately due to the neuroanatomical division of labour between dorsal (caudate & putamen) and ventral striatum (nucleus accumbens) (these suggestions being mostly based on the assumption that cortico-striatal loops are closed, meaning that the inputs from distinct sites of the cortex to the striatum do not converge) (Haber and Knutson, 2009). These studies suggested that the dorsal and ventral striatum projects to different cortical regions;the dorsal striatum projects to premotor and somato-sensory cortex, whereas the ventral striatum projects to orbito-frontal and anterior cingulate. Aside from the anatomical issues there were also several good reasons to believe that ventral and dorsal striatum are involved in segregated functionally independent territories (such as ventral-motivation dorsal-sensory motor). Firstly, it was already known that the ventral striatum has a direct and reciprocal connection with the dopaminergic neurons located in the substantia nigra pars compacta (SNc) and the ventral tegmental area (VTA) (Nicola et al., 2005; Nicola, 2007). Secondly, human neuropsychological studies and animal lesion studies showed that learning of stimulus–action associations is directly affected by activity in the dorsal striatum (Graybiel, 1998; Packard and Knowlton 2002). Thirdly electrophysiological studies in animals showed that dorsal-striatal neurons represent learned stimulus–action associations (for a review, see Schultz & Dickinson, 2001).

Altogether these studies lead to the proposal that cortical regions project to the striatum with an organization that can be identified based on different functional territories, that is, the ventral striatum (limbic-territory), the caudate nucleus (associative territory) and the posterior putamen (sensori-motor territory) (Mchaffie et al., 2005). Even though these studies suggested that ventral and dorsal mechanisms are functionally separate based on the topographical projections, the boundaries with respect to functional roles are imprecise (for a discussion, see Haber, 2003). For example, an increasing number of electrophysiological and functional imaging studies have shown that the ventral striatum (nucleus accumbens) is also activated at the time of motor preparation in procedural learning tasks (for discussion, see Nicola, 2007, Humphries and Prescott, 2010; Van der Meer and Reddish, 2010; 2011).

Putting these debates aside we know that the dissociation between these functional territories for the caudate nucleus, with its links to associative cortex and the putamen, with its links to sensorimotor cortex are over simplistic and hence no longer valid. More recent neuroanatomical studies based on human post-mortem data, which used immuno-reactivity of calcium-binding proteins have shown that caudate and putamen are made up of different functional sub-territories as shown in **Figure 3.3** (Morel, 2007). Based on these data it is hard to say for example if the putamen is only involved in sensori-motor territory or the caudate is only involved in associative territory.

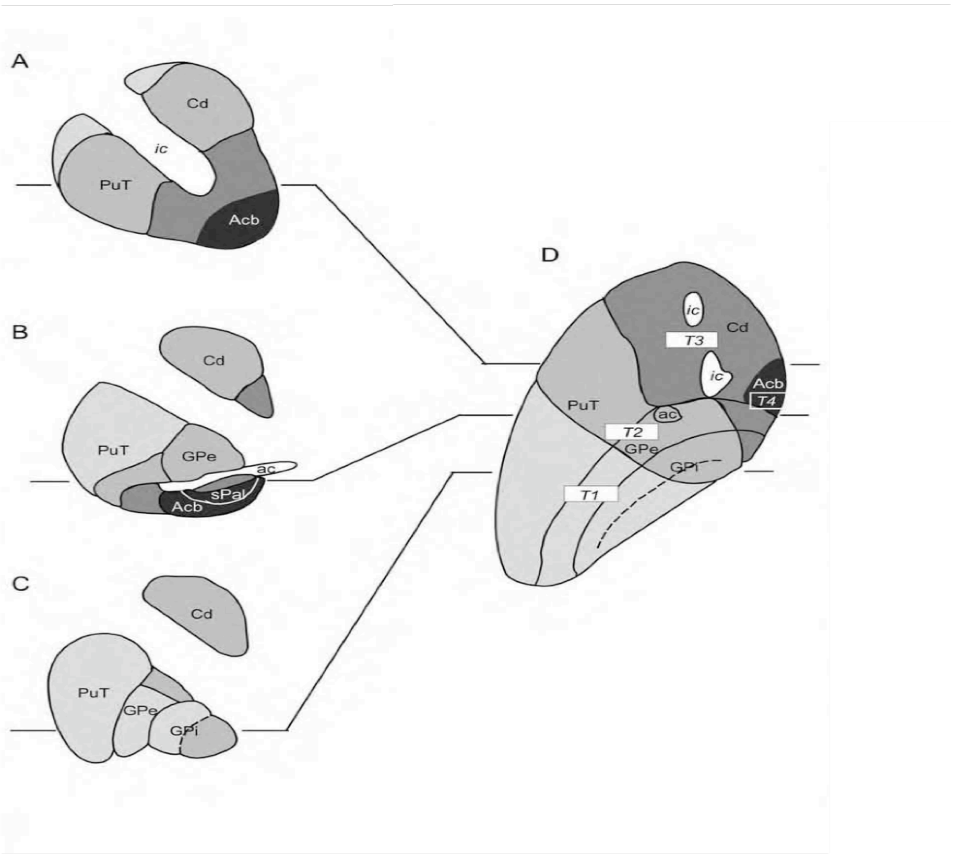


Figure 3.3 A diagram of the human striato-pallidal anatomy with different functional territories based on the photomicrographs of Morel et al. (2002). (A) to (C) represent sagital sections of the left hemisphere human basal ganglia and (D) represent axial section of the left human basal ganglia. The four territories (sensorimotor, T1; associative, T2; paralimbic, T3; and limbic, T4) are indicated by different grey levels. Abbreviations are (Cd) Caudate nucleus, (Put) putamen, (GPe) globus pallidus external segment. Figure taken from Morel (2007).

Moreover the basal ganglia are not only organized functionally but also topographically. The processing of motor information flows through a segregated sensorimotor loop connecting the primary motor cortex (MI), the supplementary motor area (SMA), the premotor cortex (PMC), and cingulate motor areas (CMA) with the sensorimotor areas of the basal ganglia and the thalamus. An ordered somatotopic distribution of motor inputs along the sensorimotor areas of the basal ganglia and thalamus has been consistently described in primates as shown by the **Figure 3.4**. According to Figure 3.4 lateral putamen of the primates receives inputs from primary motor cortex whereas medial parts of the putamen receive input from the supplementary motor area (SMA).

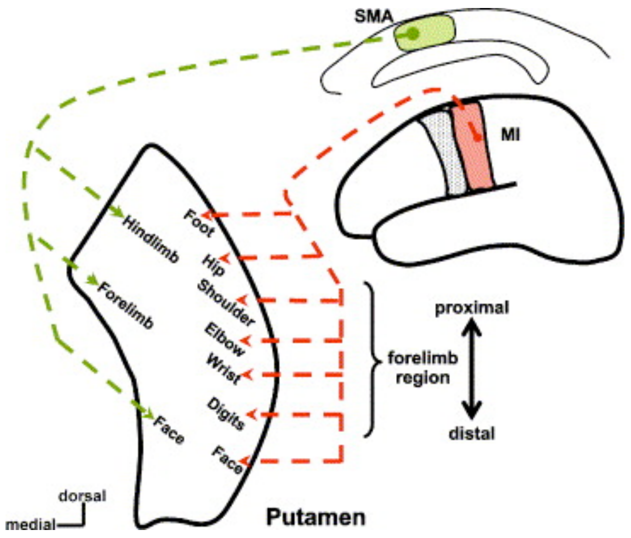


Figure 3.4 Somatotopic map of the primate striatum and cortical projections. Motor cortex (MI) and Supplementary motor areas (SMA) projection to the lateral and medial parts of the putamen in a somatotopically organized fashion. Figure Taken from Romanelli *et al.* (2005)

Furthermore in humans, recent fMRI studies of somatotopic representations ofthe foot, hand, face and eye areas in the basal ganglia have shown that each of these effectors are coded inpartially segregated regions between these somatotopic territories (Gerardin et al., 2003). Gerardin et al. (2003) showed that within the putamen, regions activated during movements of thefoot were located in the dorsal part of the structure, whereas the regions activated during lipmovements were located more ventrally and medially, and regions activated during handmovements inbetween. As we discuss in more detail in the next section, this somatotopic organization of the basal ganglia with its funnel architecture (reduction in the number of neurons from cortex to striatum through midbrain) may allow for convergence of predicted-values for stimulus-action pairs from many modalities, making the basal ganglia a effector independent action-value calculating system (see, **Figure 3.5**).

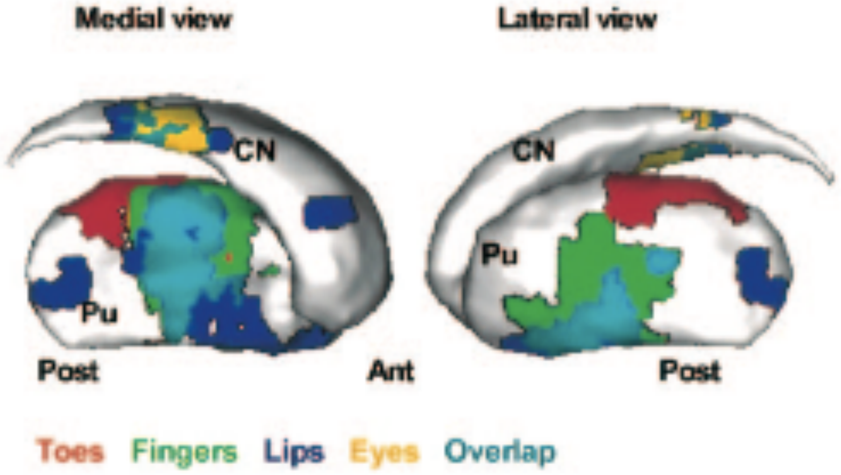


Figure 3.5 Somatotopic body representation in the medial (on the left) and lateral (on the right) striatum during toe (red), finger (light green), lip (dark blue) and eye (yellow) movements (group analysis). Abbreviations: Ant, anterior; CN, caudate nucleus; Post, posterior; Pu, putamen. Taken from Gerardin et al. (2003).

3.1.1 Multiple Cortico-Striatal Loops

The Basal ganglia play a key role in learning new procedures and action-outcome associations, implying the necessity for integrative processing by the reward and the motor circuitry. Hence, one of the most important research questions is how different cortico–basal ganglia loops integrate information in order to promote learning of reward values and actions. It has been shown that the striatum interacts with the cortex through cortico-striatal loops. Initial neuroanatomical studies suggested that these cortico-striatal loops process information in a parallel and segregated way complementing those of the cortical areas they interact with (Alexandre et al., 1986). Some of the loops are crucial for learning about cognitive (executive), motor, motivational and visual information (Alexandre et al., 1986). In the light of recent studies it seems that these multiple loops allow processing of information both in parallel and in an integrated way. In this section we review the recent findings on the cortico-striatal loops and shed light on how the spiral-loop architecture (Haber et al., 2000) of the cortico-striatum connection promotes integration of information from motivational and motor loops in order to promote coding of action values.

Recent studies suggest that the basal ganglia have a "funnel-shaped architecture" a concept used to explain the reduction in the number of cells along the dorsal-ventral axis of cortico-striatal and striato-pallidal pathways (Bar-Gad et al., 2003). This reduction in the number of neurons where each striatal neuron receives input from approximately 10,000 cortical neurons, lead to the suggestion that there is functional convergence of information from different cortical regions on the same striatal neurons. Bar-gad et al. (2000) suggested a functional description for this process, which is called dimensionality reduction. They have argued that cortico-striatal loops can do dimensionality reduction, which means that the information from cortex is being compressed towards output structures of the striatum (Bar-Gad, Morris, Bergman, 2003). This idea is in fact similar to that of Graybiel (1998) who proposed that actions towards automatization become motor chunks, which are coded in sensori-motor cortico-striatal loop (for a discussion refer to, Graybiel, 2008).

In agreement with the funnel-shaped architecture, Haber & Knutson (2010) proposed a spiral structure, both between prefrontal structures and striatum and between striatum and midbrain. Haber et al. (2000) showed that projections from the orbito-frontal cortex, ventromedial prefrontal cortex and dorsal anterior cingulate cortex converge largely in the ventral striatum and weakly in dorsal striatum, whereas the projections from the dorso-lateral prefrontal cortex and dorsal anterior cingulate cortex converge largely in dorsal striatum and weakly in ventral-striatum (see, **Figure 3.5**). This suggests that projections from the cortex are not completely segregated but overlap. Moreover, it has been shown that the spiral loop continues from striatum to midbrain and back to striatum making a three-layer cortico-striato-midbrain pathway. Hence, this cortico-striato-midbrain pathway is very important because it describes the shifts of activity towards caudal regions.

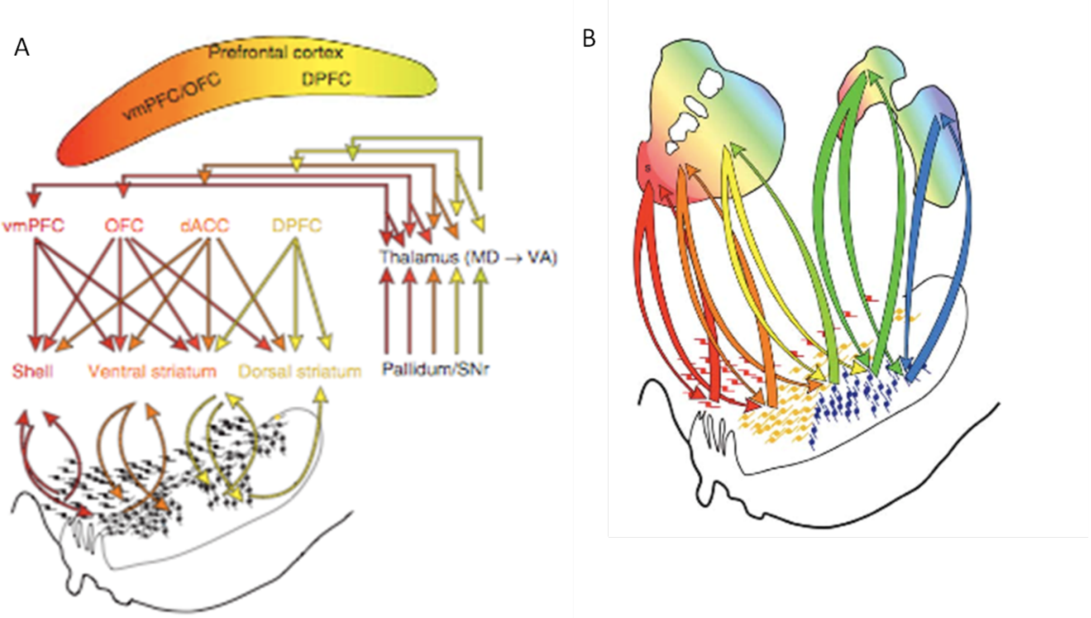


Figure 3.6 Neural circuits consisting of cortico-striatal and striato-nigral loops. a) DLPFC: dorsolateral prefrontal cortex, OFC: orbitofrontal cortex, ACC: anterior cingulate cortex, CMr: rostral cingulate motor area, PMd: dorsal premotor cortex, PMv: ventral premotor cortex, MI: primary motor cortex. Fibers from different prefrontal cortical areas converge within subregions of the striatum and from striatum they project to midbrain. b) Projections from the VTA to the shell form a "closed," reciprocal loop, but also project more laterally to impact on DA cells projecting the rest of the ventral striatum, forming the first part of a feed-forward loop (or spiral). The spiral continues through the SNS projections, through which the ventral striatum impacts on cognitive and motor striatal areas via the midbrain DA cells. Both Figures are taken from Haber & Knutson, (2010).

Moreover techniques based on probabilistic diffusion tractography and resting state functional connectivity MRI (rs-fcMRI) analysis provide means to assess functional and anatomical connectivity non-invasively in humans (Draganski et al., 2008). Overall, these studies showed that these multiple loops have overlapping projection sites in the basal ganglia and provide additional means of evidence supporting a spiral-loop architecture (Postuma & Dagher, 2006; Draganski et al., 2008). The anatomical architecture of basal ganglia should readily permit thecalculation of predicted-value signals in various parts of the brain in a flexible way. This flexibility can be seen as a shift of activation in the cortico-striatal loop from rostral parts of the cortico-striatal loops to caudal parts where highly learned actions are processed and the role of these regions are reviewed in the following section.

3.2 The Fusion of Instrumental Learning and Pavlovian Conditioning

Early studies showed that Pavlovian learning in animals elicit more than Pavlovian conditional responses but also elicit instrumental responses. An example for this is in animals' reactions to aversive cues or omission of rewards where they show freezing, running, or fightingresponses (Ulrich and Azrin; 1962; Hutchinson et al., 1968). In addition to that instrumental learning also includes certain Pavlovian responses (e.g., pupil dilation, skin conductance responses), whichallows learning of potentially highly adaptivephysiological responses beyond the restrictive set of instrumental actions. The following section reviews the brain regions that are involved in CS specific predictive responses about the outcomes of certain stimuli and actions in the context of instrumental conditioning.

3.2.1 Neural Correlates of Predicted Values

As we discussed early on and briefly in **Chapter 2 section 2.1.4.1**, previous studies showed that reward-predicting stimuli by themselves elicit neural activity about expected rewards (Schultz et al., 2003). However, in instrumental conditioning, the CS stimuli determine not only the expected reward predicted by the stimuli but also the action required by the subjects (see for a discussion, O'Doherty, 2011; Fellows, 2011; Schoenbaum et al., 2011). Because the action selection requires motor preparation and movement execution, it has been argued that these types of processes usually comprise of neuronal activity that occurs at the same time as viewing the decision cues (Schultz et al., 2003). Perhaps for this reason the predicted-value codingregions of the brain not only respond to reward predicting features of stimuli but they also respond to preparatory features of motor actions (see for discussion O'Doherty, 2011). It is important to note that sometimes the "predicted-values" might directly relate to specific actions which refer to the future rewards that are expected to be obtained after taking a specific action (e.g., if red light turns on always press the button with the right hand) and can be used interchangeably with action-values (Morita et al., 2012) or it might prefer to the expected future rewards for a particular conditional stimulus in the instrumental conditioning where the motor action can be arbitrary (e.g., chose right or left response if red light turns on) in which case it has been called predicted-value of chosen option whereas in computational neuroscience the predicted value of chosen options and actions are represented by the letter Q (e.g., Q_{CSx} for the predicted value of choosing a particular condition stimulus CS_x or Q_{left} predicted value of performing an action with the left finger which is coming from Q-learning algorithm see **Chapter 4** for details). In fact both of these response specific actions and stimulus specific actions are the source of predicted value signals, which are very hard to separate in experimental conditions and might involve orbito-frontal cortex (see for a discussion, Gerardin et al., 2001; Wunderlich et al., 2010; O'Doherty 2011).

It is important to note that we used predicted-value to refer to the second definition unless otherwise mentioned (e.g. action-value). In the following section, we review some of the studies that showed correlates of predictive-values in cortical and sub-cortical regions. These studies mainly showed predicted value and action-value activity in the striatum and various cortical regions. Some of the striatal regions are located in the input structures of the basal ganglia (Kawagoe et al., 1998; Hassani et al., 2001), and some of them are located in the output structures (Sato and Hikosaka, 2002). Moreover various cortical areas beyond medial orbito-frontal cortex (as reviewed in **Chapter 2 section 2.1.4.1**) also showed predicted-value activity. These regions include dorso-lateral prefrontal cortex, ventro-medial frontal cortex, lateral-intra parietal cortex, motor cortex, and supplemetary motor areas (Watanabe, 1996; Shima and Tanji, 1998; Leon and Shadlen, 1999; Platt and Glimcher, 1999; Wallis and Miller, 2003) which will be reviewed below in detail.

3.2.2 Predicted Values in the Striatum

Many studies suggested that pre movement firing of striatal neurons is usually influenced by reward predicting cues (see for a review Nicola, 2007). For example, earlier studies showed that neurons that fired to initiate movement showed greater excitation when the instruction indicated that the movement was to be rewarded (Hollerman & Schultz 1998; Kawagoe et al. 1998). These early studies showed that before the motor actions took place striatal neurons enhanced their firing rate by the information that movement will result in a rewarding outcome (Schultz, 2003). This enhancement probably serves to increase the probability of movement in the direction that maximizes reward (e.g., predicted-value of the chosen option or action).

For many researchers the dorsal striatum is the key area for coding predicted-values of options and actions. Perhaps because it has been thought that the main function of the dorsal striatum is related to the preparation and execution of movements (Yin et al., 2005; Balleine, Delgado & Hikosaka, 2007). More recently Hori et al. (2009) studied how dorsal-striatal neurons code for action-values by recording from the putamen of monkeys before and after action execution in a go-nogo task. They showed that most of the neurons (~50%) in the putamen code for action-values before and after action execution. Also in another reward based choice task with monkeys, Samejima et al. (2005) found that during the delay period before action execution more than one third of striate projection neurons (43/142) code the reward value in the direction ofaction that is going to be taken. In another experiment Pasquereau et al. (2007) compared action-value (ie the action values prior to action execution) and chosen-value (values at the time of action execution) in the putamen and globus pallidus internal segment (Gpi). They showed that in the period of learning the number of action value neurons in the Gpi increased and both of the structures are influenced by incentive value during the execution of motor responses. For some researchers the increase in the number of neurons that discharge for action value is an indication of automatization during learning (for a discussion, see Graybiel, 2005, 2008). Moreover a study by Arkadir et al. (2004), who recorded from globus pallidus (GPe), showed that only a small percentage of neurons code for both reward values and actions. They showed that in the initial trials the activity was modulated by both reward predictions and movement direction but later the activity was modulated mainly by movement direction. Arkadir et al. (2004) argued that most of the neurons dynamically changed their response properties as learning proceeds. Human brain imaging experiments also support the findings on predicted values in the basal ganglia. A study by Haruno and Kawato (2006) showed action-value coding activity in the dorsal striatum during a probabilistic reward-learning task.

Overall these studies showed converging evidence for representation of action values in the basal ganglia. Most of these studies report a high degree of overlap for coding reward and action related features (e.g. predicted-values and action values). Although, some of the studies report action-value coding by a much smaller number of neurons, these studies actually showed that these neurons selectively discharge during learning, more particularly in early trials they discharge more for reward-prediction and in later trials more for actions (e.g., Arkadir et al., 2004).

3.2.3 Predicted Values in the Cortex

Electrophysiological studies showed predicted-value type of activity in various cortical regions such as the dorso-lateral-prefrontal cortex (DLPFC), parietal cortex, rostral anterior cingulate cortex, and frontal eye fields (FEF) (for a review, see Sugrue, 2005; Samejima & Doya, 2007). For example, Sugrue et al. (2004) showed that in an oculomotor decision-making task, activity in the lateral intra-parietalregion activity was modulated by both the probability of choosing anaction and the reward value of the outcome. Moreover by using a computational model of the choice behaviour they also showed that this activity was highly influenced by the history of actions andrewards. Other studies also reported predicted-value coding neurons in lateral pre-frontal cortex (Watanabe et al., 2002; Kobayashi et al., 2006, Barraclough et al., 2004), and parietal cortex (Sugrue et al., 2004).

With the advances in human brain imaging recent studies showed predicted-value and action-value activity in various cortical regions. Some of these studies showed activity in ventro-medial prefrontal cortex (VMPFC) (O'Doherty, 2007) and cingulate cortex (Jocham et al., 2009). Another study by Gershman et al. (2009) found action value activity in parietal cortex during a dynamic probabilistic learning task. One recent study by Palminteri et al. (2009) showed that depending on the hand side (left or right) the predited-value activity was located in ipsi-lateral frontal cortex. This study for the first time showed that the expected value for both alternatives is first calculated in the medial frontal cortex but the predicted value of individual symbols that are determined by separate actions (left or right hand responses) are coded in the lateral prefrontal cortices. This finding converges with the earlier electrophysiological studies in monkeys (Wallis, 2007b). For example, Wallis (2007a) showed that orbito-frontal cortex calculates the predictive reward value of the outcome and then passes this information to dorso-lateral prefrontal cortex in order to promote action selection by calculating action-specific values. Moreover, Pessiglione et al. (2008) showed that not only activity in frontal regions but also in visual cortex is modulated by the value of actions. In their task, participants had to make a go/nogo selection from a masked stimulus that is delivered only for very short period of time (33-50ms). They showed that participants learn to choose the rewarding outcome even though they cannot consciously differentiate between the stimuli.Moreover, in order to compare whether pre-motor action values are coded in effector dependent or independent way, Wunderlich et al. (2009) compared the action selection with saccades and hand movements. They showed that when participants made saccadic decision the action values appear in the frontal eye fieldsof prefrontal cortex whereas when they made the decisions with hand movements the action values were found in the supplementary motor area region of prefrontal cortex.

Even though this distributed predicted-action-value network is not easy to interpret, one can easily see that parietal cortex and frontal eye fields can only code for effector-specific predicted values. Also, perhaps because of the somatotopic body representations in the striatum, it is possible that the striatum responds to predicted values from all motor modalities in an effector-dependent way. Samejima and Doya (2007) have argued that due to the hierarchical organization of frontal cortices, it is possible that predicted valuescan be represented differently at each level of this hierarchy. For example, they have argued that VMPFC and MOFC may only do state-value coding and monitoring, but LPFC may be selective both for coding the contextual information (goal or subgoal representation in working memory), as well as state predictions and action values.

3.3 *Cognitive Neuroscience of Automaticity and the Gradual Shift of Activity in the Brain*

3.3.1 Controlled and Automated Mental Processes

It has been showed that much of the human motor behaviour and cognitive processes become automatic after substantial training (Hélie & Cousineau, 2011). A good example of an automatic process in the procedural learning context is driving a car, wherein the manual gear shifting for expert drivers while recognizing road signs at the same time is considered to be an automatic process, unlike novice drivers who can perform only one of these tasks at any one time (Shinar et al., 1998). Shiffrin and Schneider (1977) proposed a dual-process model of information processing. According to their model controlled processes are defined as deliberate and attention requiring whereas the automatic processes are simply internal response chain mechanisms that are activated by the external stimuli and do not require active control or attention. Their model became very inspirational and has been applied in many areas of cognitive psychology such as connectionism (see for a discussion Botvinick and Plaut, 2006). The decisions as to whether a cognitive process is automatic or controlled is usually tested with the dual task paradigm where subjects need to perform two tasks at the same time, for example counting digits backwards and making decisions at the same time (Posner and Snyder, 1975; Logan, 1979). It has been argued that if after training participants don't pay attention to the secondary distractor task then they are making decisions automatically (for a review, Hélie & Cousineau, 2011). However, previous studies showed thatthe amount of practice for a certain skill to become automatic usually varies from individualto individual and also depends on the type of task being used (Doyon et al., 2008; Hélie & Cousineau, 2011). More recently studies have suggested that there is no one strict criterion for automatization but different operational criteria can be used to determine automaticity of motor skills, for example, reduction in the number of errors and speedy reaction times (Doyon et al., 2002; 2005; 2009; Lehéricy et al., 2005; Krakauer & Shadmehr, 2007). Utilizing these the psychological definitions of automaticity, fMRI studies usinga dual task paradigm have shown that certain brain regions show a change in activity during the process of skill automatization (Szameitat et al., 2002; Dreher and Grafman, 2003; Jiang 2004). For example, Poldrack et al., (2005) tested participants with a dual task aftertraining them on a sequential reaction time task (Nissen and Bullemer, 1987). They showed that activity in the bilateral ventralpremotor cortex, right middle frontal gyrus, and the right caudate body decreases over learning trials while performing a dual-task. On the other hand they also showed that activity in the prefrontal and striatal regions decreases equallyfor the dual and single task conditions suggesting that the decrease in activity might be a general practice effect. They concluded that lateral and dorsolateral prefrontal regions together withstriatum could subserve the executive processes involved in novice dual-task performance. Other studies have also shown a gradual decrease in activity in in various cortical and subcortical regions during the practice of complex motor skills (Jueptner et al., 1997; Poldrack et al., 2005, Doyon et al., 2009). But before pursuing automaticity in humans in more detail it is important review a similar approach that is widely used in animal studies to distinguish deliberate and automatic processes, that is the goal directed and habitual actions.

3.3.2 Goal Directed and Habitual Actions

In experimental psychology a similar distinction between the deliberate and automated mental processes can be found for goal-oriented and habitual processes in the animal instrumental learning literature. Studies have shown that there are two types of instrumental actions: habits, and goal-orientated actions where the former refers to more automatic processes whereas the later refers to deliberate calculations of outcome values (Dickinson and Balleine, 2002). Goal-directed and habitual actions are mainly studied by two experimental methods (Balleine & O'Doherty, 2010). The most common technique is the devaluation paradigm. In this method, at the beginning of the experiment rats learn via the goal-directed system because they are hungry and motivated to press the lever, they work for reaching goals (getting food). Rats highly trained to perform an instrumental task while they are hungry (or thirsty) are later subjected to the same task when they are satiated. If their response rate is not significantly different compared to non-satiated rats it has been argued that the behavior becomes a habit, since it is not being driven by goal values (ie hunger or thirst). Similarly in the second technique, extinction, rats are first highly trained to perform an instrumental task and later in the extinction phase the reward is paired with an illness. If the rats even after the extinction phase continue to perform the behaviour it has been argued that the behaviour has become habitual (Dickinson, 1985; Balleine & O'Doherty, 2010).

Computational studies have shown that habit-based learning systems are computationally efficient if the animal is exposed to a familar environment over many trials (Daw et al., 2005). However despite their efficiency habits are inflexible once they have been learned. In other words although during learning the value of the outcomesfacilitate the learningof certain actions, after many trials habits are acquired and these do not involve computations about the outcomes rather the values become fixed at a certain value for that specific outcome and no longer use computational neural resources for learning and calculation. Performance at this stage is often regarded as simple stimulus-response mappings whereas in goal directed actions the outcome is always crucial hence they are referred as action-outcome mappings (Dickinson and Balleine, 2002). Computationally habit-learning algorithms (e.g., see Q-learning algorithm in **Chapter 5**) use thepredicted-values of individual stimuli and determine the relevant actions based on those predicted values (Gläscher et al., 2010; Daw et al., 2011). However, recent studies have shown that it may be inefficient to learn predicted values of individual stimuli or actions, for example if the outcome is delivered after a series of stimulus-response stages forming a chain in which case alternative goal-directed learning algorithms have been proposed (Daw et al., 2005; Daw et al., 2011). In these alternative algorithms, during goal-directed learning an internal representation of the outcome (specific to each action) is always calculated and used to guide actions (Daw et al., 2005).

Although there are important differences between the methods used to study controlled-automatic processes and the goal directed and habit systems, it was suggested that there might be overlapping neural mechanisms controlling both paradigms (for a discussion, see Ashby, Turner, Horvitz, 2010). Recent fMRI studies in humans have shownactivation in the medial prefrontal cortex, whenthey performed goal-directed actions for food rewards (when the participants are hungry) (Valentin et al., 2007) and activation was found in the dorsal striatum the actions are no longer goal-directed (when the participants are satiated) (Tanaka et al., 2008)(see **Figure 3.7a** and **Figure 3.7b**). More recently Tricomi et al., (2009) conducted an fMRI study where some of participants trained for 3 days on an instrumental contingency task. In their experiment the value of the outcome devalued after extensive training and an increased BOLD signal was observed in dorsal striatal in the testing phase (**Figure 3.7c**).

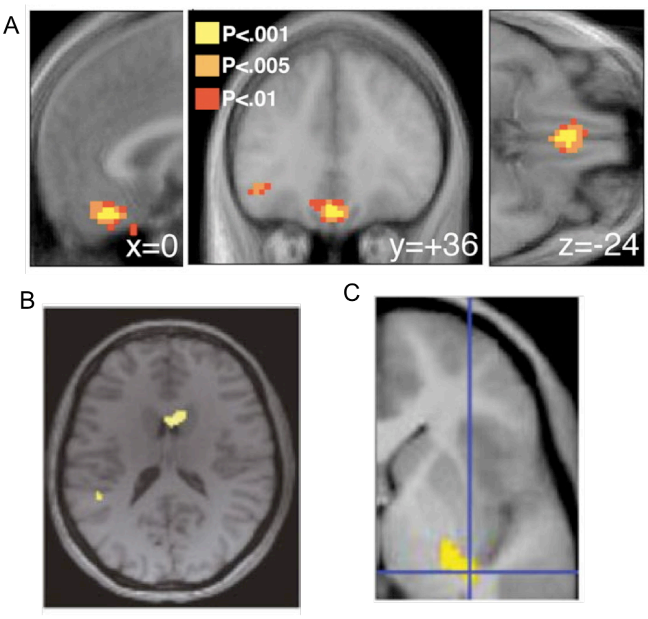


Figure 3.7 a) Region of human medial OFC exhibiting a response profile consistent with the goal-directed system. Activity in this region during action selection for a liquid food reward was sensitive to the current incentive value of the outcome, decreasing in activity during the selection of an action leading to a food reward devalued through selective satiation compared to an action leading to a non-devalued food reward. Figure 3.7a is taken from Valentin *et al* (2007). b) The regions of the human anterior dorsomedial striatum also exhibit sensitivity to instrumental contingency when the paricipants are in non-satiated condition. Figure 3.7b is taken from Tanaka et al (2008). c) The region of the human posterior lateral striatum (posterior putamen) that exhibits a response profile that is consistent with the development of habits in humans. . Figure 3.7c is taken from Tricomi et al, 2009.

It is important to note that the distinction between the goal-directed/habitual control of actions and the deliberate/automatic mental process do not correspond perfectly since the former one is related to acquisition of instrumental behaviour with rewards whereas the latter is not. However recent studies have suggested that the underlying neural mechanism might partially overlap between these two types of processes (see for a discussion, Ashby, Turner, Horvitz, 2010).

3.3.3The General Effects of Practice on the Neural Correlates of Procedural Learning

Although the abovementioned methods are the most generally used techniques to study controlled and automated (or goal-directed and habit) processes, the gradual shift in activity in the brain towards learning new skills are also studied with various motor tasks in humans by comparingeither early and late learning trials or by comparing novel and familiar tasks (Jueptner et al, 1997; Toni & Passingham, 1998; 1999). In fact human fMRI studies have shown that the caudate nucleus and putamen, the two sub-regions of striatum which receive input from different cortical regions are differentially active during early versus late learning trials for various probability learning tasks (Haruno & Kawato, 2006; O'Doherty et al., 2004), motor sequence learning tasks (Jueptner et al, 1997; Toni & Passingham, 1998) and visuo-motor learning tasks (Toni & Passingham, 1999). The activity in the head of the caudate nucleus is correlated with early learning trials, as well as in novel tasks, whereas the putamen is active in late learning trials as well as with familiar task sets (see for a review, Ashby, Turner, Horvitz, 2010).

In a series of electrophysiological recordings in monkeys Hikosaka and colleagues showed that distinct brain regions become preferentially activated in specific stages of learning during a sequential finger movement task (Hikosaka et al., 1995; Hikosaka et al., 1999; Hikosaka et al., 2002). In addition to that, Miyachi et al. (2002) showed that neurons in the rostral-striatum (associative striatum) show preferential movement-related activations while monkeys learn novel movement sequences in a procedural motor learning task. In contrast to the previous studies of Hikosaka and colleagues, Miyachi et al. (2002) showed that the activity in these neurons decreases after the learning phase of the experiment (see, **Figure 3.8**).

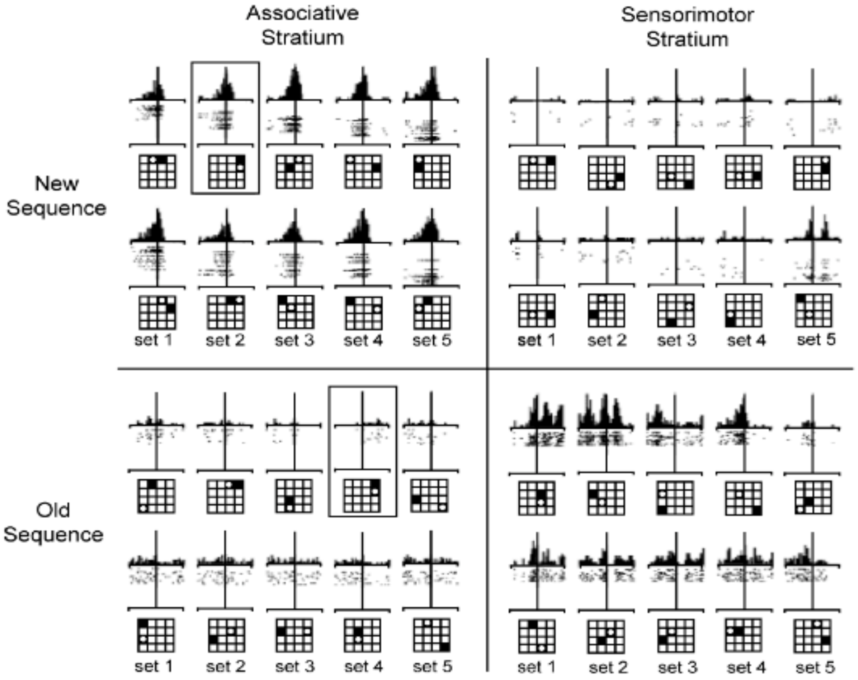


Figure 3.8 Responses of a single neuron in the associative striatum and a single neuron in the sensori-motor striatum of a monkey during the performance of a motor sequence learning task. The recordings shown are recorded during movement initiation. This figure shows that neurons in the associative striatum responded more strongly to the new sequence than the old sequences, where as those in the sensori-motor striatum responded more strongly to old sequences than to new sequences (Taken from Miyachi et al., 2002)

Additionally, differences in activity have been observed in cortical regions for early versus late learning (Kubota & Kamatsu, 1985; Assad et al., 1998). For example, several electrophysiological studies showed that movement related activity in the prefrontal cortex was more sensitive to learning novel stimulus-response pairs than familiar pairs (Kubota & Kamatsu, 1985; Assad et al., 1998). The gradual activation shift in the cortex has been recently reviewed by Badre and D'esposito (2009), where they argued that rostral prefrontal cortex is involved in coding more abstract goals and actions whereas caudal prefrontal regions code more concrete motor actions. Graybiel (2008) also suggested a similar progression of functional activation in cortico-basal ganglia circuits where anterior frontal regions get activated in the early trials whereas when the behaviour becomes highly repetitive in the form of habits, addictions, stereo-types the involvement of caudal prefrontal regions as well as dorsal striatum increases (see, **Figure**

3.9). Moreover, before moving on to the next section, we want to point out that overall these results suggest a similar rostro-caudal shift in activity for predicted-values (Haruno & Kawato, 2006).

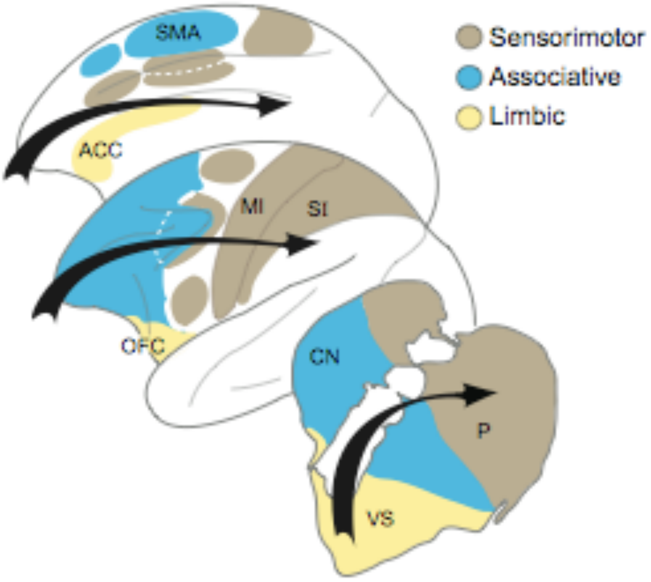


Figure 3.9 An approximate schematic of the dynamic shifts in activity cortical and striatal regions as habits and procedures are learned. Sensorimotor, associative, and limbic regions of the frontal cortex (medial and lateral views) and striatum (single hemisphere) are shown for the monkey. ACC, anterior cingulate cortex; CN, caudate nucleus; CP, caudoputamen; MI, primary motor cortex; OFC, orbitofrontal cortex; P, putamen; SI, primary somatosensory cortex; SMA, supplementary motor area; VS, ventral striatum. Taken from Graybiel (2008).

The main finding of these studies in general is that during the process of habituation of actionsas well as automatization of motor skills activity certain regions such as the prefrontal cortex and limbic striatum decreases but activity in the sensori-motor striatum increases (see for a review, Graybiel, 2008; Ashby, Turner, Horvitz, 2010).

3.4 Novelty Signals as a Possible Source of Explanation for the Shift of Activity in the Brain

We reviewed that there are differences in brain activation during automatization of cognitive processes, motor skills and habits. This change of activation might perhaps be related to the way, the brain deals with novel information. In order to highlight that point in the next several sections the brain areas involved in processing *novel stimuli* and *novel instrumental actions* are reviewed.

3.4.1 Novelty Related Activations in the Dorsal-Prefrontal-Circuitry

Novelty is a puzzling concept and neural correlates of it are still not well understood (Redgrave, Gurney, Reynolds, 2007; Dayan, Kakade, Montague, 2000). Perhaps the reason for this is that there is more than one way to test and describe novelty in experimental settings (Ranganath and Rainer, 2003). Several descriptions used to refer to novelty in experimental settings are known as *stimulus novelty*, *contextual novelty,* and *associative novelty (see* for a recent discussion, Duzel et al., 2004*).* Given the strong diversity of the definitions and usage of novelty there is no consensus on a single strict definition (Ranganath and Rainer, 2003). We think that *stimulus novelty* and *action novelty* are crucial for understanding the rostro-caudal shift in the brain during instrumental conditioning and for that reason, in this thesis "novelty" mostly refers to familiarity of the participanteither with the stimulus or related instrumental actions. Thus during the early learning trials of an instrumental conditioning task novelty signals influence both the stimulus and related actions

Stimulusnovelty implies that the current stimulus and its properties are completely unknown to the subject (e.g., letters in an unfamiliar language). Stimulus novelty in experimental settings is usually tested by showing the participant a set of familiar and novel items and the neural responses to each category are compared (see e.g. Tulving et al., 1996; Kirchhoff et al., 2000). It has been shown that stimulus novelty activates various regions in the brain including lateral prefrontal cortex, hippocampus, basal ganglia, and midbrain (Bunzeck and Duzel, 2006; Wittmann et al., 2007). The most important evidence for neural responses selective for novelty in the prefrontal cortex come from electrophysiological studies (see for a review, Miller & Cohen, 2001). For example, Assad et al. (1998) showed that when monkeys were required to make a saccadic choice for novel pictures in an associative learning task then prefrontal neurons fired significantly more than to familiar pictures (see **Figure 3.10b**). Also similar findings have been replicated in fMRI studies with humans (for review please see, Duncan & Owen, 2000) (see **Figure 3.11b**).

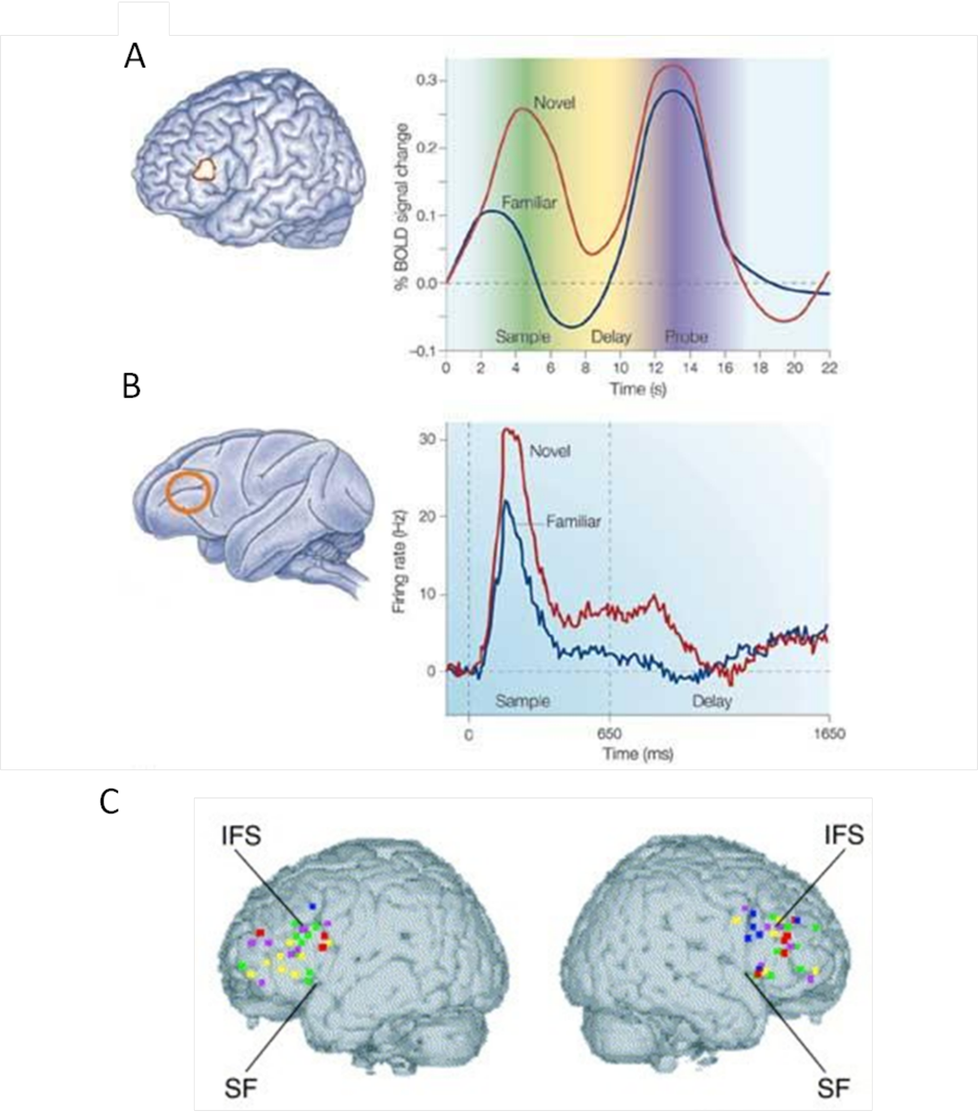


Figure 3.10 a) Bloodoxygenation level-dependent (BOLD) functional magnetic resonance imaging (fMRI) study, inwhich activity during a delayed matching-to-sample task was compared between trials involvingnovel or familiar stimuli. Activition in the lateral prefrontal cortex show increased BOLD response and greater delay for 'novel' trials than during 'familiar' trials. Figure is taken from Ranganath and Rainer, (2003) whereas the original study is based on Ranganath & D'Esposito, (2001). b) Single-unit recordings from the sulcus principalis of themonkey lateral prefrontal cortex during a delayed matching-to-sample task. Average firingrates

across a group of neurons shows that neurons in this region show a greater discharge rate for novel objects compared to familiar objects. Figure is taken from Ranganath and Rainer, (2003) whereas the original study is based Assad et al., (1998).c) Purple regions in the inferior frontal sulcus (IFS) represents brain areas involved in novelty progressing.Figure 3.10c is taken form the meta-analytic review of Duncan & Owen (2000).

On the other hand, *novel actions* are usually studied by asking subjects to learn new motor procedures such as a new finger tapping sequence. Passingham & Rowe (2002) suggested that the role of the lateral frontal region (BA47) is crucial for learning novel actions. In a series of experiments Passingham and his colleagues showed that the inferior frontal cortex is active while participants are learning novel motor skills but this region is inactive when participants perform familiar motor skills. Moreover his follow-up work showed that the activity in the dorso-lateral prefrontal cortex is not caused by the working memory but by directing of attention to novel actions.

Novel information processing in the brain is also studied via response suppression and priming tasks, which are based on the fact that repeated stimuli can cause a reduction in the discharge rate of neurons and consequently cause a decrease in the haemodynamic response (Henson & Rugg, 2003). Even though the exact neuronal mechanism of this phenomenon is not clearly understood, it has been shown that it can occur in various cortical regions and is supposed to be caused by the sharpening of cortical representations due to plasticity (Wiggs & Martin, 1998). Various fMRI studies showed that when the stimuli become familiar there is a decrease in the amplitude and a delay in the peak latency of the heamodynamic response (Henson & Rugg, 2003). As we discuss in the next chapter, it is plausible that the novelty-triggered attention mechanism for stimuli and actions might be controlled by dopamine in the form of adaptive learning rates (see for details, **Chapter 4.3**).

3.4.2 Dopamine and Novelty

In this section we review evidence that novelty and attention is associated with an increase in dopamine release in various brain regions, which facilitates cortico-striatal plasticity. Early single-cell recordings in monkeys reported that midbrain dopamine neurons respond to novel situations (Ljungberg et al., 1992) (see, **Figure 3.11b**). Additionally, Ihalainen and colleagues (1999) showed that hippocampal and prefrontal dopamine release in mice increases while the animals were exposed to new cage environments. Moreover, Legault and Wise (2001) reported that the novelty-based dopamine release in the ventral tegmental area (VTA) can be abolished by interrupting the connection between hippocampus and VTA. This suggests that hippocampal memory mechanisms seem to control novelty tagging to newly arriving stimuli by differentiating familiar and non-familiar items based on memory. Moreover, by performing electrophysiological recordings from cat VTA Horvitz et al., (1997) showed that mesolimbic dopamine neurons fire to non-rewarding salient events and they suggested that dopamine might play a role in attentional processes, rather than a specific role in reward (see, **Figure 3.11a**).

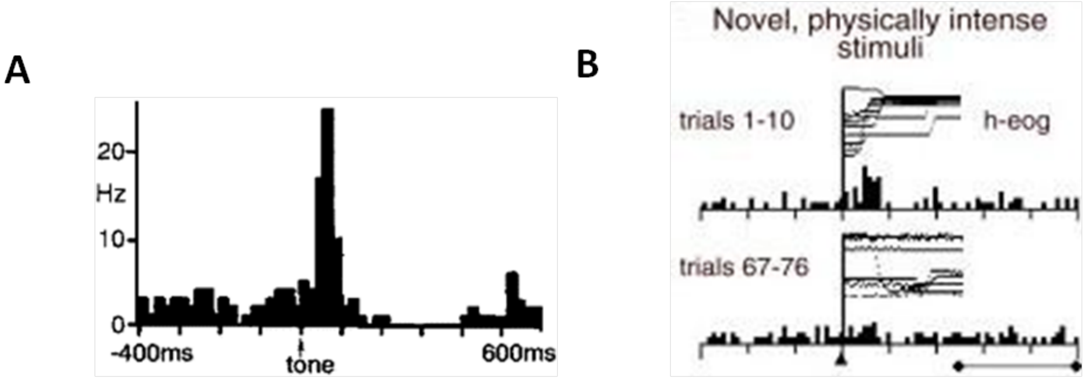


Figure 3.11 Dopaminergic novelty responses with phasic activation. Figure 3.11a shows a histogram of the activity of a single dopamine neuron in cat VTA. This neuron shows increased firing in response to novel stimulus. Figure 3.11a is taken From Horvitz, Steward, and Jacobs (1997). Figure 3.11b showsphasic dopaminergic activations following novel, physically intense stimuli. Overlapping h-eog traces show horizontal eye movements toward the novel stimulus;

unfocused h-eogs after >60 trials indicate familiarity, accompanied by loss of dopamine responses. Data from Ljungberg et al. (1992) published in Kobayashi and Schultz (2010). Figure 3.10btaken from the review of Schultz (2011).

Moreover, in order to study the effect of novelty in computational reinforcement learning, Kakade and Dayan (2002) proposed a model where novelty acts as a "bonus reward". Their model is mainly based on the evidence that in the absence of rewarding stimuli novelty acts like an intrinsic reward and in some cases promotes exploration of novel environments (Bevins and Besheer, 2005). Moreover, based on the working-memory models of Braver & Cohen (2000), Kakade and Dayan (2002) proposed that novelty-based dopamine release maygate stimulus information into working memory to allow for the storage of a new stimulus until its potential rewarding properties are evaluated (Kakade and Dayan, 2002) (see **Chapter 4**, for the details of this computational model).

Given that novelty attracts attention and that novelty and attention both influence processing of actions and objects of perception in similar ways through dopaminergic pathways, several studies suggested that they are crucial in guiding cognitive control in reinforcement learning situations (Hikosaka and Watanabe, 2000; Wittmann et al., 2008). For example, Schultz et al. (1995) and Roelfsema & van Ooyen, (2005) argued that an attentional feedback signal from the output layer of a neural network (e.g., cortex) might limit plasticity in earlier layers (e.g., striatum). They argued that attentional feedback signals mimic the dopamine activity and this may be related to the effect of dopamine release on post-synaptic excitability of striatal neurons. In other words, dopamine release helps to enhance the cortical activity on the strongest striatal activity, and decreases the weaker activity, which in turn facilitates the focusing (or gating) effect (see, **Figure 3.12**).

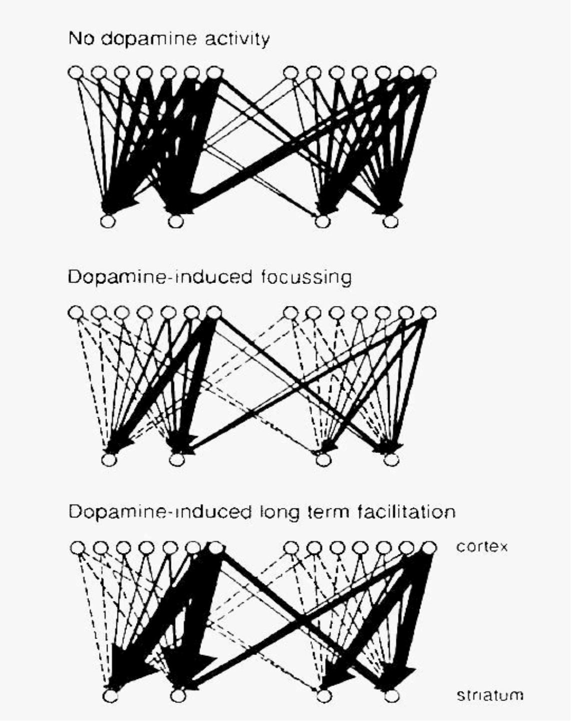

No dopamine activity
Dopamine-induced focussing
Dopamine-induced long term facilitation
cortex
striatum

Figure 3.12 A two-layer neural network model showing the effect of dopamine on the connections weights. The top image shows no dopamine activity, the weights are high for all connections. In the middle image, dopamine increases the relevant weights causing focusing effect. In the bottom image dopamine induce long-term facilitation. Figure taken from Schultz et al. (1995)

3.4.3 Top-Down Influences of Attention on Stimuli and Actions

The brain's ability to selectively allocate resources to certain stimuli, memories, thoughts and actions is called attention (Posner & Petersen, 1990). Studies in neurological patients and physiological studies in humans and animals implicate a distributed network of cortical and subcortical regions for attention (Desimone & Duncan, 1995). Neuroimaging and electrophysiological studies identified attention-related areas in the frontal, parietal, temporal and occipital cortex (see for a review, Corbetta & Shulman, 2002). The limited attention processing capacity of our brain allows filtering out irrelevant sensory information by top-down (goals) and bottom-up (stimulus salience) processes. There are several factors that may affect top-down processing of information such as knowledge, expectation or goals (Corbetta & Shulman, 2002). In addition to those, novelty and unexpectedness are also counted as factors affectingattention (Corbetta & Shulman, 2002). Previous studies have shown that the role of attention in Pavlovian and instrumental conditioning is crucial because it helps the organism to select good predictors and inhibit bad predictors of outcomes (Kakade, Dayan, Montague, 2000). Both bottom-up and top-down attentions on decisions are widely studied in cognitive neuroscience (for a review see, Knudsen, 2007). It has been thought that both forms of attention can enhance decisions. For example, according to Gazzaley & D'Esposito (2007), top-down attention in decision-making refers to selective attention processes on goal-directed decisions. It has been shown that selective attention to goal directed decisions inhibit the effects of exogenous factors like saliency and novelty and improves decision quality by decreasing reaction times. Moreover, the link between top-down attention and working memory has been well studied over the years (Corbetta & Shulman, 2002). It has been argued that useful information for decisions such as the values of options are kept in working memory over periods of seconds (Wallis, 2007a) and top-down attention is responsible for manipulating this information by accessing working memory (Knudsen, 2007).

In the instrumental conditioning context the term "attentionalset" has been generally used to define representations that include both selecting task-relevant stimuli (e.g., symbols representing options) and task-relevant actions (Cools et al., 2010). In the attentional set-shifting paradigm, subjects discriminate between two patterns according to one of two stimulus dimensions (e.g., shapes or lines). It has been shown that PD patients are impaired in the attentional set–shifting paradigm as they are impaired in top–down control attention (Cools et al., 2010).

Moreover, Armel et al., (2008) developed a model to investigate the role of visual attention on binary choice. Their model predicts that by increasing the amount of time spent on a particular option in a multiple-choice task the probability that an item be chosen increases. Their results showed that, in the appetitive condition the attended options are 6% to 11% more likely to be chosen than the unattended options. In contrast, aversive items are 7% less likely to be chosen. In a follow up study using eye-tracking and excluding the fixation, they showed that the participant's choices exactly matched with the drift-diffusion model's predictions (Krajbich et al., 2010). These studies clearly demonstrate the importance of top-down attention on action selection.

Although the effect of attention on procedural learning is well studied, its effects on predicted-values are less clear. For example, Pessoa and Engelmann (2010) suggested that various fronto-parietal attention networks can affect information processing in subcortical motivation networks. They proposed three models where the attention network and the reward systems can modulate actions. The first model considers that attention and motivation don't interact with each other but both influence the actions separately. In the second model, motivation mediates attention in a unidirectional way and then attention and motivation influence behaviour separately. In the third model, attentional and motivational systems integrate information via bidirectional communication and influence the behaviour jointly. Although it is not perfectly clear which model provides better accounts for the relation between attention and motivation one recent study provides valuable evidence. van Schouwenburg, et al. (2010b) using a dynamical causal modelling of functional brain imaging data, demonstrated top-down control by prefrontal cortex and selective gating of basal ganglia to salient events implying that there may be functional connectivity between prefrontal cortex and visual cortex (see, **Figure 3.13**).

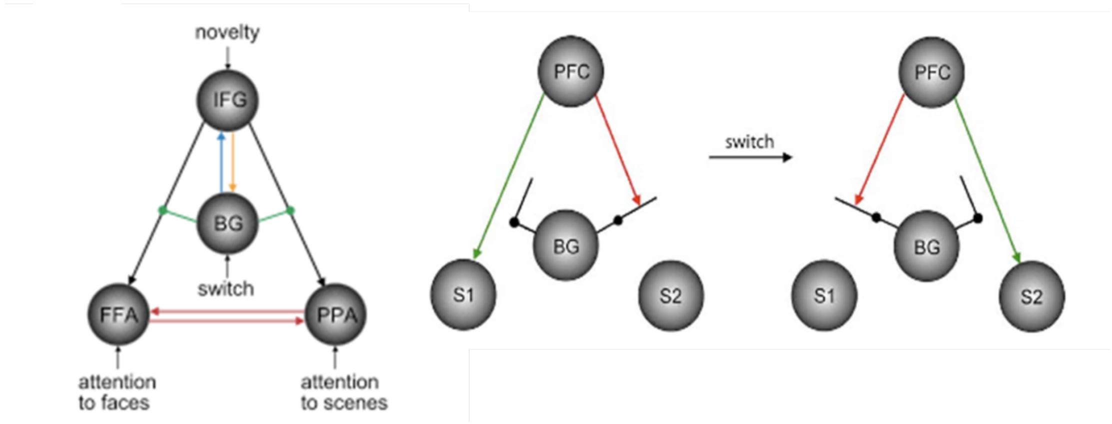


Figure 3.13 The model that best fits to the fMRI model includes connections from the IFG to the FFA and PPA (black) and the following inputs: novelty to the IFG, switch to the BG, attention to faces to the FFA, and attention to scenes to the PPA. Figure 3.13a is taken from van Schouwenburg et al., (2010a). The next two figures show cognitive switching of basal ganglia by regulating top-down projections from prefrontal cortex (PFC) to posterior sensory areas. Figure 3.13b and Figure 3.13c is taken from van Schouwenburg et al., (2010b).

Their study suggests that there is a bidirectional link between inferior frontal cortex, which exerts top-down attention to novel stimuli, and the basal ganglia, which controlsaccess of salient inputs to other cortical regions (for a review, van Schouwenburg et al., 2010a).

3.5 Interim Summary

Many cortical regions are involved in calculating predicted values during associative learning task, especially mPFC which calculates predicted values for goal-directed actions and dorsal-striatum which is involved in habitual actions. The anterior posterior shift in the brain is also not specific to goal directed learning but also other cognitive-motor processes show a similar shift of activation towards caudal regions as they become more automatic. Moreover, we also reviewed that midbrain dopaminergic neurons, ventral striatum and DLPFC (Ljungberg et al., 1992; Duncan & Owen, 2000; Witmann et al., 2008; Bunzeck et al., 2010) are involved in processing novel stimuli perhaps via a bidirectional link (van Schouwenburg et al., 2010a). Based on that we hypothesized that novel stimuli engage with dorsa-lateral frontal cortex and should show increase in the

BOLD responses in midbrain compared to familiar stimuli. Furthermore, if the predicted-values are calculated for a set of stimuli during early versus late trials (same as for novel versus familiar stimuli), the predicted-values for the early trials should engage in MPFC and the predicted values in late trials should engage in dorsal striatal regional.

Error... Therefore I am.

Chapter 4

4 Formal Models of Associative Learning

In this chapter, the historical progress of mathematical models, which leads to computational models of reinforcement learning, is reviewed. Furthermore, the capacity of these models in modelling psychological situations (e.g., Pavlovian and instrumental) is discussed.

4.1 Associative Learning Models that are Inspired from Psychology

4.1.1 Bush and Mosteller Model

It is useful to start this chapter by introducing the linear model of Bush and Mosteller (1955) which is the ancestor of many subsequent models (Bower, 1994). Bush and Mosteller (1955) focused on modelling peoples' reactions to binary choices over many trials. In a repeated binary choice task there are two options say option, A and option B, where choosing one or the other provides a correct feedback with a probability P(A) and P(B). In such a case, what Bush and Mosteller wanted to know was the following, if the participant chooses option A with a certain probability value in their head, say P'(A), what will be the probability of choosing A again in the next trial, and how is the participant's choice affected by the actual probability of A being rewarded, P(A)? They suggested that every time a participant gets rewarded for choosing A there is a small increment in the probability of choosing A again and every time a participant gets nothing there is a small decrement in the probability of choosing A. In their model they assumed that the probability of choosing option A is calculated by the following rule:

$$P_t^{'}(A) = P_{t-1}^{'}(A) + \alpha\left[R - P_{t-1}^{'}(A)\right] \qquad [4.1]$$

In Equation [4.1], α represents the learning rate parameter and R is the magnitude of the reinforcer. *$P`_{t-1}$ (A)* is the participants estimated probability for choosing option A in the prior trial. The idea behind their model was simple and captures the basic choice behaviour or the participants. In each trial, *t* the probability of choosing A is increased or decreased by the difference between the reward outcome and the probability of choosing the same option in a previous trial.

4.1.2 Rescorla-Wagner Model

One of the most influential mathematical models in the history of associative learning is the linear learning rule of Rescorla and Wagner (1972). Earlier studies failed to formulate the learning process as accurately as the Rescorla-Wagner model. For example, in the early twentieth century Thorndike (1911) did not consider the organisms' anticipation of reward in his associative learning theory, which remained the most important missing link for decades. In the Bush and Mosteller model explained in the previous section the update was based on response probabilities, which are purely descriptive explanation of behaviour, however in the Rescorla & Wagner model the update is based on the internal association strengths, or so called weights (Newell, Lagnado, Shanks, 2007). Newell and his colleagues (2007) argue that this modification allows predicting various response strategies such as maximizing outcome (choosing the high-probability reward option in all of the trials), which was not possible before with the Bush and Mosteller model because it learns to match the response probabilities.

Rescorla and Wagner (1972) attempted to shed light on the very basic question, which is under what conditions does the associative strength between an unconditional (US) and conditional stimulus (CS) increase ? According to Rescorla and Wagner, their model "depends not only on the reinforcement itself but upon the relationship between that reinforcement and the reinforcement that the organism anticipated" (Rescorla, 1972, p.11). The Rescorla-Wagner learning rule updates the value attributed to a stimulus after each trial, by the fraction of what is referred to as the "prediction error". The latter is the calculated difference between the value of the reward predicted by the CS and the value of reward received. So this can only be calculated after a CS-US sequence of events. According to the Rescorla-Wagner learning rule, the value of an arbitrary stimulus, is updated as follows:

$$\Delta V_x^t = \alpha_x \beta\left(\lambda - V_x^t\right) \qquad [4.2]$$

and the new value of the stimulus becomes:

$$V_x^{t+1} = \alpha_x \beta\left(\lambda - V_x^t\right) + V_x^t \qquad [4.3]$$

Here, ΔV_x^t is the change in the expected reward value of a particular stimulus x; parameter α_x represents the learning rate for a particular stimulus x; $\beta$is the outcome specific learning rate that defines the saliency of the outcome; variable V_x^t denotes the current value attributed to the stimulus or the reward expected by an organism at trial t; and l is the actual reward that the organism received. The term $(\lambda - V_x)$refers to the difference between the reward gained at the end of the trial and the previous (expected) value of the stimulus. Hence the difference between the actual and expected reward is called the Rescorla-Wagner prediction error signal, and is represented by δ (Niv &Schoenbaum, 2008) where the weight change simply becomes:

$$\Delta V_x^t = \alpha_x \beta \delta \qquad [4.4]$$

The Rescorla-Wagner prediction-error signal represents how surprising (in a positive or negative sense) a particular reward is after the organism receives the outcome. The Rescorla-Wagner learning rule is generally used to explain Pavlovian conditioning, however it can be modified to model instrumental conditioning as well. In that case the level of surprise after choosing a particular option changes the associative strength between the US and the associated option. Note that a high learning rate, α_x assigns greater weight to the prediction error, and can slow down the convergence, or delay learning, when the rewards have a stochastic character.

In fact this fundamental principle was used early on in the least-mean-square error algorithm (LMS) (sometimes referred to as generalized delta-rule in neural networks) where it is used to train single-layer neural networks (Widrow & Hoff, 1960). In such a single-layer neural network learning occurs by updating the weights of a network by the error, δ and its multiplication with the learning rate parameter (see for example in the classic back-propagation algorithm, Rumelhart, Hinton, and Williams, 1985). However, there are two important differences between the Rescorla-Wagner learning rule and the least-mean-square algorithm. Firstly, in the least-mean-square algorithm the outcome is not a reward but a predetermined teacher signal, which corrects errors in a supervised manner. For example, Klopf (1972) suggested that supervised methods lack adaptive behaviour because the trial-and-error behaviour of the agent is missing in the equations of supervised learning algorithms. For that reason, the supervised error-correcting rules like the delta learning rule, the prediction error is calculated as the difference of desired outcome and the observed outcome. In contrast the prediction error in the Rescorla-Wagner learning rule is calculated as the difference between the reward received and the value of the CS. Several researchers proposed that the Rescorla-Wagner learning rule is a special case of least-mean-square algorithm (Sutton and Barto, 1981; Quinlan, 1991). Still, it is important to note that these two learning methods are only equal in an artificial setting because in natural ecological settings the rewards are embedded in the environment where the animal has to learn by trial and error.

Secondly, a more fundamental implementational difference exist between the least-mean-square algorithm and the Rescorla-Wagner learning rule, which is that in neural networks the weight inputs (in this case a CS) are transformed by an activation function (that defines activation of a neuron with a particular input) and then the prediction-error is calculated, which allows learning of non-linear CS-US relations (Dawson, 2008). Having said that Gluck and Bower (1988) used the actual Rescorla-Wagner algorithm in their connectionist models of human probabilistic category learning

4.1.3 Limitations of the Rescorla-Wagner Learning Rule

Although the Rescorla-Wagner model can explain a wide range of behavioural phenomena, including classical conditioning, extinction, or blocking it has several limitations (Miller, Barnet Grahame, 1995). In this section a few of these limitations are reviewed (see for details, Miller, Barnet Grahame, 1995; Pearce and Bouton, 2001).

Firstly, the Rescorla-Wagner learning rule fails to explain certain learning phenomena such as latent inhibition (Gluck, Mercado, Myers, 2008). The latent inhibition paradigm has two phases (Lubow & Moore, 1959, Lubow, 1973). In the first phase animals are pre-exposed to the CS without training and then in the second phase they are trained with the US. When compared with a group of animals that have not been pre-exposed to the CS, the group of animals who are pre-exposed show decreased learning. Therefore, latent inhibition occurs when the animals are pre-exposed to CS. This is problematic for the Rescorla-Wagner learning rule because no unexpected outcome occurs (no prediction error) during the first phase of the experiment (just familiarization with CS) and one should expect no learning to take place. However, it has been shown that this pre-exposure to CS with no feedback affects learning in the second phase of the experiment. For this reason, the Rescorla-Wagner learning rule is considered to be a **US modulation theory** because it cannot take into account the CS novelty and CS familiarity (Gluck, Mercado, Myers, 2008).

Secondly, the Rescorla-Wagner model predicts that the history of conditioning has no influence on its present status; only the current association value is important. However, experimental evidence shows that history of reward effects have significant influence on the value calculation (Kennerly et al., 2004; Amiez et al., 2004; Hampton et al., 2006).

Third, Rescorla-Wagner models cannot deal with higher-order conditioning since during learning of two consecutive conditional stimuli, no unconditional stimulus occurs (Seymour et al., 2004). As explained in **Section 4.2.2** the Temporal-Difference learning algorithm can be seen as a modified Rescorla-Wagner learning rule that is modified to account for modelling higher-order conditioning and within trial temporal dynamics.

4.1.4 Attentional Approaches to Associative Learning

To account for latent inhibition and other phenomena several alternative mathematical models have been proposed (Machintosh, 1975, Grossberg, 1975, Pearce and Hall, 1980; Schmajuk and Moore, 1989). In these models, stimulus novelty is modulated by either the attention paid to the CS, or the US (Schmajuk, 1997), such that during learning novelty increases when an unpredicted CS is presented or when a predicted CS is absent. For this reason these extended associative learning models are usually referred to as **CS modulation theories** (Gluck, Mercado, Myers, 2008). For example, in the Pearce and Hall (1980) learning rule the associability is determined by the following equation,

$$\alpha_x^t = \left|\lambda^{t-1} - V_x^{t-1}\right| \qquad [4.5]$$

In equation 4.5, α_x^t is the learning rate or associability, which is based on the prediction error in the previous trial. Likewise in the Rescorla-Wagner model, λ is the outcome and V_x is the value of stimulus *x*. Hall (1991) introduced a further change and the latest equation became the following:

$$\alpha_x^t = \gamma\left|\lambda^{t-1} - V_x^{t-1}\right| - (1-\gamma)\alpha_x^{t-1} \qquad [4.6]$$

In equation 4.6, γ determines the responsiveness to the associability of the CS that controls the speed of learning and the change each trial. Differences between Rescorla-Wagner learning rule and Pearce Hall learning rule can be easily seen by the following two equations:

$$\text{Rescorla-Wagner:}\ \Delta V_x^t = \alpha_x \beta \left(\lambda - V_x^t\right) \qquad [4.7]$$

$$\text{Pearce-Hall:}\ \Delta V_x^t = \alpha_x^t S_x V_x \qquad [4.8]$$

Similar to the role of β in the Rescorla-Wagner learning rule in equation 4.4, S_x in equation 4.8 refers to stimulus specific saliency. Eventually the most important difference between the Rescorla-Wagner learning rule and Pearce-Hall is that in the latter one the prediction error is calculated with the prediction error in the previous trial and called as the learning rate which models CS novelty or so called attention (Hall, 1991). It is important to note that according to Schmajuk (1997) in much of the models that follow the Pearce-Hall learning rule novelty decreases as learning progress, but it is never totally eliminated.

4.2 Associative Learning in Computational Models of Reinforcement Learning

Reinforcement learning (RL) is a computational framework used to model the behaviour of an artificial agent that receives scalar reward signals. RL has its roots from control theory (Bellman, 1957) and psychology (Watkins, 1989; Suton & Barto, 1998). Over the last two decades research in reinforcement learning promoted fruitful interaction between different disciplines like neuroscience, psychology and artificial intelligence perhaps because most of the time the challenges that reinforcement learning models are exposed to are similar to actual human and animal behaviour in an experimental setting such that when an animal has to learn the environment it needs to select and monitor the consequences of its actions to achieve a goal state. During learning the environment provides immediate or delayed reward and the animal learns the value of intermediate states and to estimate the value of taking particular actions.

From the control theory side of RL, initial steps were taken by Bellman (1957) who tried to find a solution for the optimal policy problem (learning to choose the actions that maximize expected future reward), and came up with a set of equations called the Bellman consistency equation in 4.9.

$$V(S_t) = E\lfloor r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \ldots \mid S_t \rfloor \qquad [4.9]$$

$$= E\left[\sum_{t=1}^{\infty} \gamma^{t-1} r_t \mid S_t\right] \qquad [4.10]$$

Here in equation 4.9, the real world counter part of the definition of a *state* can be a stimulus in a conditioning trial that predicts reward. Then V(S_t) becomes the value that is equal to the cumulative sum of all future rewards, *r*. The discount factor γ ($0 \leq \gamma \leq 1$) dictates the extent to which rewards that arrive earlier in time are favoured over those that arrive later. Equation 4.10 is equal to Equation 4.9 where E[.] denotes the expected value of the sum of future rewards. Similarly the cumulative sum of all future rewards or called the value of the next time step V(S_{t+1}) can be written as:

$$V(S_{t+1}) = E\lfloor r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \ldots \mid S_{t+1} \rfloor \qquad [4.11]$$

Note that, although *t* indicates the trial number it can also indicate any variable that provides a way to order the states visited during the learning process. Then, the current expected return for V(S_t) could be re-written in terms of next states value; V(S_{t+1}) as:

$$V(S_t) = r_t + \gamma V(S_{t+1}) \qquad [4.12]$$

Equation 4.12 describes the optimal value equation and is the essential for most RL models. A similar Bellman equation can also be written for a Markov Decision Process

(MDP) where the outcome and state transitions are stochastic (Doya, 2007). In such a system, transitions from one state to another is predefined with a probability distribution, P(S_{t+1} | S_t) and the recursive relationship for the state values is written as:

$$V(S_t) = P(S_t) + \gamma \sum_{S_{t+1}} P(S_{t+1} \mid S_t) V(S_{t+1}) \qquad [4.13]$$

From that point, the agent can enumerate and store the entire state transition matrix in its memory. Similarly the instrumental counterpart for the state transition matrix can be written as follows:

$$T(s,a,s') = P(s_{t+1} = s' \mid s_t = s, a_t = a) \qquad [4.14]$$

Defining a state transition function is the essence of the markov decision process and takes the form in equation 4.14 where the probability distribution over the new state, (s') is conditional on state (s), action (a) pairs. Such a process involves learning the probabilities associated with the different states in a model and sometimes referred to as the model-based RL or dynamic programming. Action values in model-based RL systems are calculated for different routes by searching forward through the map and evaluating the potential rewards found therein (Daw, Niv & Dayan, 2005; Doya, Samejima, Katagiri & Kawato, 2002). Then, the optimal policy can be found by determining the best states or actions. However, the optimal policy that is explained above are related to two types of problems and directly related with the environmental situations of the agent (see **Section 4.2.1**). The first problem is called the *prediction problem*. Basically, the prediction problem refers to the question, how well the agent evaluates (or predicts) upcoming events derived from its policy after learning the value of each state? A Policy evaluation is used to refer to the prediction problem that is to find, how much reward an agent will get when it follows a certain policy (Sutton and Barto, 1998; Ji and Daw, 2011). On the other hand, the second problem is related with the control theory (which also depends on the prediction problem) and refers to the question how can the agent make planning decisions after finding the optimal policy (Wörgötter & Porr, 2005; 2008). Nonetheless, whether there exists an optimal policy or not the aim of the agent is to increase its long run sum of the expected future rewards (or minimize punishments).

4.2.1 The Effect of Environmental Dynamics in Reinforcement Learning that Determine Pavlovian and Instrumental Conditioning

In a learning situation it is important to define the environment of the agent because it gives the premises of which updating strategy the agent is going to use. The environments in a RL methods are usually non deterministic which indicates that an action in a state on two different occasions may result in different next states (Kaelbing, 1996). There are two types of environments that exist, the open loop and the closed loop environment. In the open loop environment there is no need for an agent to make explicit actions but in the close loop environment, it is necessary for an agent to take actions to learn the value of a particular state. If the environment is in open loop such as in a Pavlovian conditioning situation, the agent will use the state-value updating strategy (for details see temporal difference learning algorithm in the next section). On the other hand, if the environment is close loop such as the instrumental conditioning experiments, the agents has to use an action value updating strategy, where it has to make decisions by taking actions and have to learn how desirable those actions are in order to build a policy. Algorithms used in the close-loop environments will be returned in section 4.2.3.

4.2.2 *Temporal Difference Learning Algorithm*

Besides the Rescorla-Wagner model, there are alternative computational models for Pavlovian conditioning. One alternative is the temporal-difference (TD) learning algorithm (Sutton & Barto, 1998). As reviewed in detail in **Chapter 2**, over the last decade, the TD algorithm has been commonly used to model the prediction error signal of dopamine neurons in various Pavlovian learning experiments. The TD learning algorithm is based on the assumption that state values (expected future rewards) are coded separately from the actions taken to optimize rewards, hence its suitability for modelling Pavlovian conditioning. On the other hand, standard action-value learning algorithms (as reviewed in the following **Section 4.2.3**) are based on the assumption that state-action pairs are coded together, hence their suitability for modeling instrumental conditioning experiments. The TD algorithm is represented by the following equation:

$$\delta_t = r_t + V(S_{t+1}) - V(S_t) \qquad [4.15]$$

Here in equation 4.15 the prediction error is calculated as the difference between the reward in the current trial summed with the value in the next trial minus the value of the current state. Although at first sight the TD algorithm is similar to the Rescorla-Wagner learning rule, TD implements some major generalizations. First, the TD algorithm explicitly represents time during the course of a trial. Second, during the course of a trial the subject can visit multiple states. Third, the TD-associated values are predictions of the *cumulative* future rewards associated with each *state* (rather than predicting the immediate reward associated with the current stimulus as in the Rescorla-Wagner rule). In the state value updating strategy, the stimulus-reward schedule is defined by the experimenter and the agent can only act as the observer not the decision maker. A variant of TD learning, called the actor-critic algorithm, is also used in instrumental learning experiments (Barto, 1995, Joel et al., 2002). Even so, it shares lots of similarities with the TD algorithm that is to sum the state values are coded separately from actions; optimal actions in Actor-Critic architecture are those that follow the trajectory of optimum states.

4.2.3 Computational Models of Action Learning

A relatively recent concept in decision neuroscience, action-value coding, aims to answer this question: how can learning about stimuli that lead to rewarding outcomes guide actions? According to some computational reinforcement learning algorithms, action-value coding involves agents first learning the reward value of state-action pairs by doing trial-and-error learning and then updating the value of actions by using the prediction error, that is the difference between how much reward was expected and how much actual reward received by performing that particular action. The basic idea can be illustrated with the example, that is if I am in a state where I am wondering over possible outcomes then which course of action will make me happier (such as going to the kitchen for a cup of water or reaching to the table for a cup of coffee).

It has been argued that during the time of stimulus presentation depending on the *action selection rule* (policy) it is highly likely that the action with the higher action value will be selected. Over the last twenty years various methods were developed in order to explain the underlying computational mechanisms of this selection procedure; e-greedy4, sofmax5and winner-take-all selection rules are examples of such methods commonly used in reinforcement learning (Kaelbing, 1993; Kaelbing et al., 1996; Sutton & Barto,

[4] In the e-greedy policy most of the time the action with the highest estimated reward is chosen, called the greediest action. However with a small probability e, an action is selected at random. The action is selected uniformly, independent of the action-value estimates. The epsilon greedy method balances the exploration and exploitation behavior of the agent. This method ensures that if enough trials are done, each action will be tried.

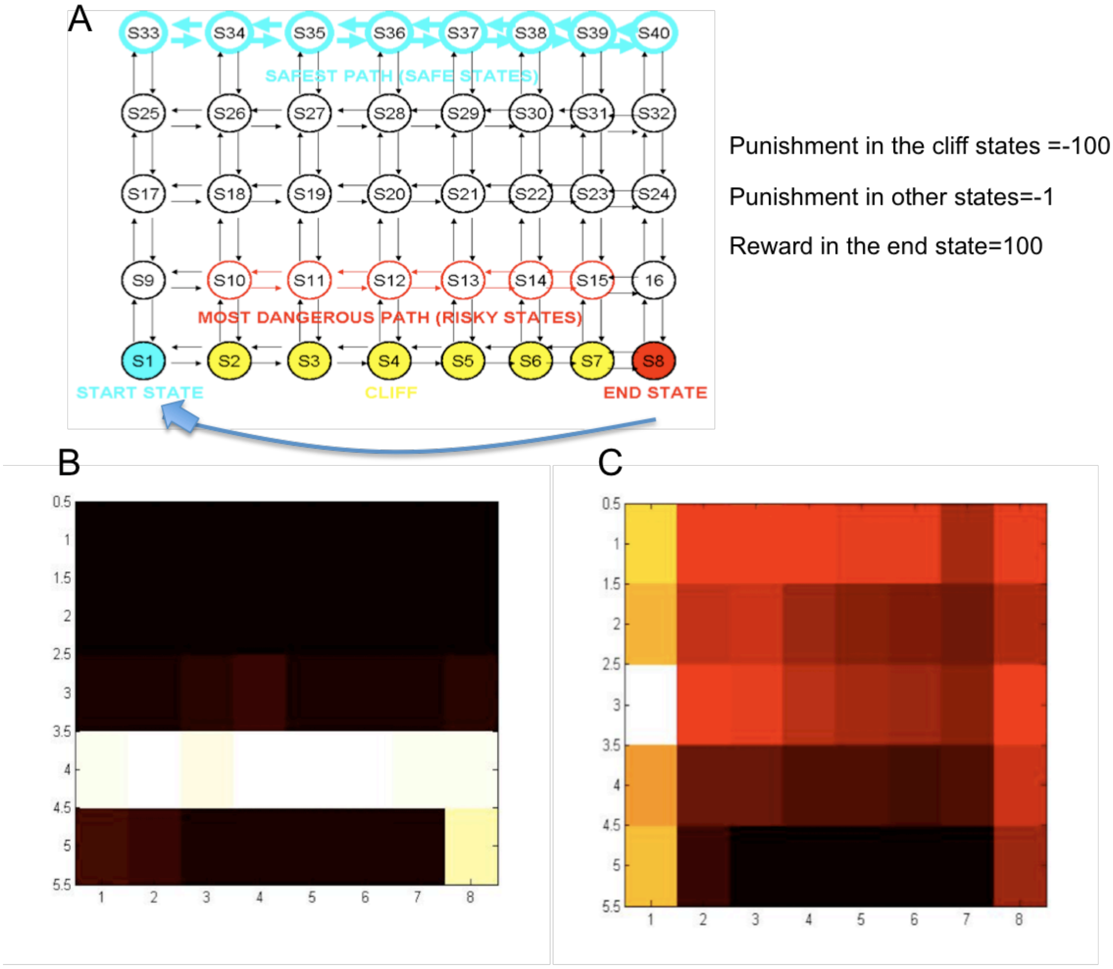
[5]The softmax method assigns weights to each possible action, according to their action-value estimate. Then, a random action is selected taking in to account the weight associated with each action.

1998; Dayan & Abbott, 2001; Suri & Schultz, 1998, 1999). Recent studies in computational neuroscience based on the cortical neuroanatomy suggest that the motor cortex represents actions for all available options in the action preparation period, and these actions compete with each other in order to take control over the desired action, which is in fact similar to unsupervised competitive learning in neural networks (Cisek, 2006, 2007).

A more fundamental question, which is crucial for this thesis, is how the nervous system represents action values and where in the brain these action values are encoded. In order to answer these questions, it is important to understand the historical evolution of the action-value concept in reinforcement learning. The notion of action-value was originally borrowed from computational theories of reinforcement learning and used to find an optimal solution to the spatial credit-assignment problem. Spatial credit-assignment simply refers to efficient allocation of actions in a multi-option choice task (Minsky, 1963). The first steps for understanding the mathematics behind action-values were taken by Richard Bellman (1957). His work showed that it is possible to write a set of value functions recursively, which allows an agent to increase its expected cumulative reward in a learning situation. Similar to equation 4.11 the expected value of reward for a particular action can be represented by the following equation:

$$Q(s,a) = E\left[r_t + r_{t+1} + r_{t+2} + r_{t+3} + \ldots \mid S_t = s, A_t = a\right] \qquad [4.16]$$

Q represents the *state-action value function* and E[.] denotes expectations of reward over a state *s* and action *a*. This major concept not only created new research areas in artificial intelligence and operations research but also later lead to Watkins' (1989) first Q-learning algorithm, that is proved to converge to the optimal solution (Watkins & Dayan, 1992). The Q-learning algorithm is an off-policy reinforcement learning technique which uses the values of state-action pairs represented by Q(s,a) to calculate the optimal solution without needing to store state transition probabilities. This can be simply demonstrated by the cliff-walking task shown in Figure 4.1. In the cliff-walking task, the agent starts at state S1 (initial state), and can move up, down, left, or right until it reaches the terminal state, S8.For each action taken, the agent gets a negative reward of -1.The terminal state S8 gives a reward of +100, and the yellow zone (see **Figure 4.1**) is a "cliff", which gives a negative reward of -100 and sends the agent back to the starting state. It is expected that the agent should learn to reach the goal in the least number of steps. The reward value of each state-action pair is stored in a table, which is usually referred as the Q-table. When the agent in the next trial come to the same state, the algorithm updates this table by changing the value of the old state-action pair Q(s,a) to the new state-action pair Q(s_{t+1}, a_{t+1}) by using Equation 4.16.



Punishment in the cliff states =-100

Punishment in other states=-1

Reward in the end state=100

Figure 4.1 Schematic representation of the cliff-walking task that demonstrates the relationship between an agent and its environment. (A) There is a set of environmental states "s"; a set of actions, represented by black arrows and a set of scalar reward and punishment values at the intermediate and terminal states. (B) The policy of Q-learning algorithm (on the left) and (C) The policy of the Sarsa algorithm (on the right). The color indicates how many times that state is visited (e.g., black-not visited, red-frequently visited, white-most frequently visited. We tested both learning algorithms with the epsilon greedy action selection rule with e = 0.1. The results obtained were averaged from 1000 episodes. The Q-learning algorithm showed less explorative behavior than the Sarsa algorithm with the same exploration rate. The reason for this behavior is because the Q-learning algorithm selects the highest value future actions in the next state $Q_{\max}(s_{t+1}, a_{t+1})$Indeed accidently falling in the cliff does not change the *maximum* Q-value of that state and hence does not diminish Q-values at previous states. The results also showed that the agent that use the Q-learning algorithm choose the risky path more often than the safe path (see Appendix B for implementations of the above figures in Matlab).

It is important to note that in this example the only problem the agents have to deal with is the spatial creditassignment problem, and not the temporal credit-assignment problem. The temporal credit-assignment problem, which is also called the distal reward problem in classical conditioning (Hull, 1943) deals with predicting the timing of events such as reward delivery. It thus appears that the action-value coding framework can only be applicable to actions that lead to immediate rewards, whereas most actions in natural settings don't lead to immediate rewards but serve long-term goals. As such these are related to both the temporal and spatial credit-assignment problems.

The key issue to emphasize with the Q-learning algorithm is the similarity between an artificial agent in a cliff-walking task and a subject in a real world experimental setting and the way in which the problem is approached to find an optimal solution. Perhaps an analogy can be made between the cliff-walking task and a saccadic decision making task. Consider a monkey making a saccadic decision-making task where it gets fruit juice after making a saccade to a symbol on the screen. When the monkey is in the initial state, which refers to the situation when it is fixating a cross on the screen, two symbols appear on the screen either the right or left side and the monkey has to make a saccade. Let's call these alternatives action one and action two. Byrepeated trials and errors through reward processing, the monkey will adjust its behavior by updating the value of actions in order to adapt tothe reinforcement contingency (for a slightly different example from an actual study please refer to Morris et al., 2006). The prediction error here can be represented simply by the difference between reward and the Q-value; $[R - Q(s_t, a_t)]$. In other words, in that kind of experimental setting, subjects will continuously update the values of their actions in order to reach that final rewarding state that increase their satisfaction (or decrease un-satisfaction in the case of avoidance learning). The two algorithms, Q-learning and Sarsa, that were used in Figure 4.1 are shown in Equation 4.17 and Equation 4.18 respectively.

$$\text{QL} \Leftarrow Q(s_{t+1} + a_{t+1}) = Q(s_t, a_t) + \alpha[r + \gamma Q_{\max}(s_{t+1}; a_{t+1}) - Q(s_t, a_t)] \text{ off-policy} \quad [4.17]$$

4.

$$\text{SARSA} \Leftarrow Q(s_{t+1} + a_{t+1}) = Q(s_t, a_t) + \alpha[r + \gamma Q(s_{t+1}; a_{t+1}) - Q(s_t, a_t)] \text{on-policy} \quad [4.18]$$

The only difference between Q-learning and Sarsa is that the former uses an off-policy method and the latter an on-policy method where the off-policy algorithm can update the estimated value functions using hypothetical actions $Q_{\max}(s_{t+1}, a_{t+1})$where as on-policy algorithms update the value function based strictly on experience. This explains the naming of the SARSA algorithm, an abbreviation referring to state, action, reward, state, action. As such, the optimal policy can be found by determining the best state-action pairs that will lead to the optimal path.

4.2.4 General Limitations of the Applications of the Computational Models in Functional Imaging

Although both the Q-learning and SARSA algorithms calculate the action-values and are commonly utilized in electrophysiology and human brain imaging studies (Niv et al., 2006), they are not the only plausible models to for action-value representations in the human brain. The Rescorla-Wagner learning rule is also utilized in electrophysiology and neuroimaging studies in order to help us understand how the values of stimulus-response pairs are coded (Palmineteri et al., 2009; see also **Chapter 5** for a discussion). It is important to note that even though algorithmic representation of Q-values is different from the value assigned to a particular conditional stimulus, V_x as in the Rescorla-Wagner learning rule, both have equivalent potential for representing action values in instrumental conditioning tasks if the subject performs an action based on the stimulus value. This is because these different value representations are usually correlated at the neural level when the task requires a participants' choice from multiple options (ref). Hence, some recent imaging studies used stimulus values (anticipatory value) V_x at the time of motor movement or action selection periods in order to study action values (ref). Even though in these studies they used the term "value of stimulus" we believe that this difference between the value of a state-action pair Q(s,a) and the value of a particular stimulus V_x is only a question of semantics although they are different algorithmically. Therefore, we think that if the task is instrumental, the Rescorla-Wagner learning rule can well be used to study action values. However, one should be cautious in interpreting these variables, for example the term "action-value" is generally used to indicate values that are calculated before the actions and the term "expected value of an action", "reward prediction" or "chosen-value" are calculated during or after the action is executed (Wunderlich et al., 2009). Also in certain occasions the term "expected value" is used for a different meaning. In that context "expected-value" refer to the mathematical operation where one multiplies the value of all possible options with their probability of occurrence ($EV = P_x \times V_x + P_y \times V_y ...$). Hence expected value indicates the expected future sum of all rewards and is usually associated with "state-values" rather than action-value (O'Doherty et al., 2004).

Finally, although, Q-learning, Sarsa and Rescorla-Wagner models provide useful information for understanding action-values, these models fail to explain the effect of novelty on recently learned action values and the effect of familiarity on those action-values in the form of habits. The reason for this is because the effect of novelty is usually captured by the learning rate parameter (except dopamine as novelty bonus models, see Kakade & Dayan, 2002; Witmann et al., 2008), which seems crucial in explaining the differences in activation for learning with familiar and novel stimuli explained in **Chapter 3.** In the next section, examples from adaptive learning rate models are given.

4.3 Adaptive Learning Rate Methods

Similar to the Pearce-Hall learning rule mentioned early in this chapter, which updates associability with a dynamic learning rate, early studies in neural network research showed that using a fixed learning rate in neural networks have certain disadvantages when the step-size of the error surface changes more sharply for one weight than the others (ravines) (Jordan, 1988). Due to this disadvantage various adaptive learning rate methods were developed in order to improve the convergences and speed of learning performance (Jordan, 1988). It has been shown that dynamic learning rate methods are not only perform better than fixed learning rate methods in stationary problems but they are also better in non-stationary problems where the optimal solution to a problem change overtime (Sutton; Behrens; O'Doherty et al., 2006). In fact there are many ways to estimate trial by trialchanges in learning rate. State-space models (Smith et al., 2004), moving average technique (Eichenbaum et al., 1986), information theoretic techniques such as Kullback-Leibler divergence (Haruno et al., 2004), filtering algorithms such as Kalman Filter (Kakade Dayan, 2002), Bayesian learning methods (Fahrmeir and Tutz, 2001, Behrens et al., 2007),fixed-number of consecutive correct responses models (Fox et al., 2003; Stefani et al., 2003) are only few of the dynamic learning rate estimation algorithms. Due to this huge variety in the dynamic learning rate techniques in the next section only two of the most communally used techniques will be summarized Kalman Filtering and Incremental-Delta-Bar-Delta algorithm.

4.3.1Kalman Filter

The Kalman Filter is a powerful mathematical method developed to solve Wiener problems (named in honour of Norbert Wiener) that is to estimate noise in a continuous stochastic-process (Kalman, 1960). In the last ten years a number of studies suggested that the cerebellum and the hippocampus are carrying out computations similar to a

Kalman Filtering algorithm (Paulin, 1986, Bousquet et al., 1998). Moreover evidence suggested that Kalman Filtering might occur in sensory processing and behavioral conditioning (Kakade & Dayan, 2000; Kakade, Dayan, Montague,2001;Dayan & Yu, 2003).The goal of the Kalman filter is to predict the true value of the state (or signal) when the measurements are noisy. The basic idea behind the theory can be demonstrable with a simple example of dead reckoning. Consider, that somebody wants to estimate his precise location by the using global positioning system (GPS) driving a car. In such a case the observations from the GPS will be noisy showing the car a couple of meters away from the place where it actually is. The GPS might give him noisy measurements for a lot of reasons but probably most importantly it will due to driving speed and maneuvers he is making. Since, if we are to estimate the true position of his car by using a Kalman Filter, we need the speed and wheel direction of his car and add this information to the initial noisy position observed from the GPS signal. Daw et al., (2006) applied this simple idea to a multi-arm bandit problem where the participants have to learn to allocate their choices between different bandits in order to earn maximum amount of money. In their experiment the mean payoffs for each bandit is drawn from independent Gaussians with pre-determined mean and variance such that the mean rewards for some bandits are better than others. Secondly, the rewarding outcome from each bandit was diffused with a Gaussian random walk. Given that the mean reward value and the variance in the outcome are assigned a prior, Kalman Filter updates the posterior mean payoff by using the following equation:

$$\mu_t^{post} = \mu_t^{pre} + \kappa_t \delta_t \qquad [4.19]$$

In equation 4.19 μ_t^{post} refers to the updated mean reward of a particular bandit and μ_t^{pre} is the prior mean reward of that bandit with prediction error signal δ_t equals to the difference between the reward outcome in a trial and the mean reward outcome, which is as follows:

$$\delta_t = r_t - \mu_t^{pre} \qquad [4.20]$$

In the Kalman filter the learning rate κ_t which is also called the Kalman gainis calculated by the following equation:

$$\kappa_t = \frac{\sigma_{t,pre}^2}{\sigma_{t,pre}^2 + \sigma_{t,post}^2} \qquad [4.21]$$

Note that while doing parameter estimation these initial mean payoffs μ_t^{pre} and the standard deviation σ_t^{pre} are the first two free parameters in the model that are similar to the initial GPS signal and the speed in the above dead reckoning problem respectively. Also the variance for the payoff of the chosen bandit is updated by separate functions. In addition to that the Kalman Filter makes the assumption that the subject believes that the outcome of the bandits might vary over time and are governed by the Gaussian random walk which adds additional free parameters to the system. Overall this makes six free parameters (Daw et al., 2006). In conclusion the Kalman Filter is a powerful algorithm and had been utilized in fMRI research but due to its high degree of free parameters and initial assumptions we don't think it is suitable for explaining the biological plausibility of all reinforcement learning situations.

4.3.2 Incremental-Delta-Bar-Delta Algorithm

One such algorithm that uses adaptive learning rates is the Incremental-Delta-Bar-Delta (IDBD) algorithm. The IDBD algorithm was first introduced by Sutton (1992) and is an extension of the previous delta-bar-delta learning algorithm (Jordan, 1988). IDBD is a meta-learning algorithm in the sense that it doesn't only learn the weights in a network (such as values of stimuli or actions) but it also learns the learning rate.

In the IDBD algorithm the learning rate is updated by the following equation:

$$\alpha_i(t) = e^{\beta_i(t)} \qquad [4.22]$$

In the above equation α indicates the learning rate and $\beta_i(t)$ is an additional memory parameter that is actually modified using another function $h_i(t)$. The $\beta_i(t)$ term is updated as follows:

$$\beta_i(t+1) = \beta_i(t) + \theta\delta(t)h_i(t) \qquad [4.23]$$

In the above equation 4.23, θ is a positive constant, which is a meta-learning rate, and $h_i(t)$ is a decaying memory trace keeping the records of previous weight changes. The aim of learning is to minimize the squared $h_i$and is thus a decaying trace of the cumulative sum of recent changes to weights $w_i(t)$ and the basic learning rule for updating weights can be calculated as follows:

$$w_t(t+1) = w_i(t) + \alpha_i(t)\delta(t) \qquad [4.24]$$

The advantage of the IDBD algorithm over its predecessors such as the delta-bar-delta-algorithm (Jordan, 1988) is that it has only one free parameter, the meta-learning rate, and it works with incremental training of inputs rather than batch training. It has been shown that the IDBD algorithm shows greater performance than the least-mean-square algorithm (LMS) and is as good as the Kalman Filter algorithm in a benchmark problem (Sutton, 1992). For example, Sutton (1982) suggested an alternative adaptive learning rate framework showing that even though the learning rate in the Rescorla-Wagner learning rule is stimulus specific it is not capable of modeling positive and negative acceleration of learning also Rescorla-Wagner model can't be able to model capture choice switches in a probabilistic reversal learning task (Glascher et al., 2009).

4.4 Modelling Novelty in Reinforcement Learning

According to Kakade and Dayan (2002) novelty in reinforcement learning literature, usually acts like an additional reward given to early trials in a learning situation and promotes exploration of the novel environments. They proposed that this effect of novel stimuli distorts the reward predictions and actions and novel stimuli can come to be treated as if it is rewarding. Also it is important to note that in the normal circumstances in a reinforcement learning the values of states or actions are set to zero at the beginning because the simulated agent doesn't know anything about its environment. However in some cases rather than setting those initial weights to zero some researchers set random initial weights to those novel states. In fact there is neurophysiological evidence, which shows that the monkeys midbrain dopaminergic neurons shows increased spiking for novel stimuli (see Chapter 3 for details). In their reinforcement learning model Kakade and Dayan (2002) call this feature*novelty bonus*. They modelled this by changing the reward *r*(*t*) at time *t* with the following equation:

$$r(t) \rightarrow r(t) + n\big(u(t), T\big) \qquad [4.25]$$

In equation 4.25, *u*(*t*) is the state at time *t* and *n*(*u*(*t*),*T*) is the novelty of this state in trial *t*. According to their proposal *n*(*u*(*t*),*T*) uses information about the novelty of the stimuli associated with state *u*(*t*), and makes the novelty signal decrease over trials as the stimuli become familiar. Therefore, the effect of the novelty bonus on the prediction error signal of the temporal difference equation is written as:

$$\delta(t) = r(t) + n(u(t), T) + v(t+1) - v(t) \qquad [4.26]$$

Kakade and Dayan (2002) also introduced a second modification for modelling the novelty signal that is called the novelty shaping bonuses. According to this modification the value of an initial state *v*(*t*) is derived from a potential function $\varphi(u)$ of the state *u*, so that the estimated value *v*(*t*) at time *t* is replaced by the following equation

$$v(t) = v(t) + \varphi(t) \qquad [4.27]$$

In the above equation $\varphi(t) = \varphi(u(t))$ is the value associated with the state at time *t* and is assumed to be set high for states associated with novel stimuli and that therefore deserve exploration. The temporal difference equation is written as follows:

$$\delta(t) = r(t) + \varphi(u(t+1)) - \varphi(u(t)) + v(t+1) - v(t) \qquad [4.28]$$

In these models described above the novelty signal decays hyperbolically to zero overtime as the stimulus repeated over trials.

4.5 *Interim Summary*

In this chapter various mathematical accounts of Pavlovian and instrumental conditioning were summarized. It was emphasized that several mathematical models of conditioning suggests that learning rates should change overtime as the animals get familiar with the conditional stimulus (CS) or the learning rate should change based on the attention capacity of the animal (Pearce & Hall, 1980). Additionally it was summarized that both in the past and to day using adaptive learning rates also concerns researchers in the computational concerned with the problem related to speed of learning. The common point between psychologist and computer scientists is that over successive trials the learning rate decreases (see for a discussion Schumajuk, 1997) and the values of conditional stimuli don't get affected by the late fluctuations in the prediction error signal as much as early trials. **Figure 4.2** shows the hypothetical trade off between an adaptive learning rate and the stimulus value. On the other hand if a fixed learning rate is chosen the prediction error could have a big effect on the value in the late learning trials (e.g., due to an unexpected negative feedback).

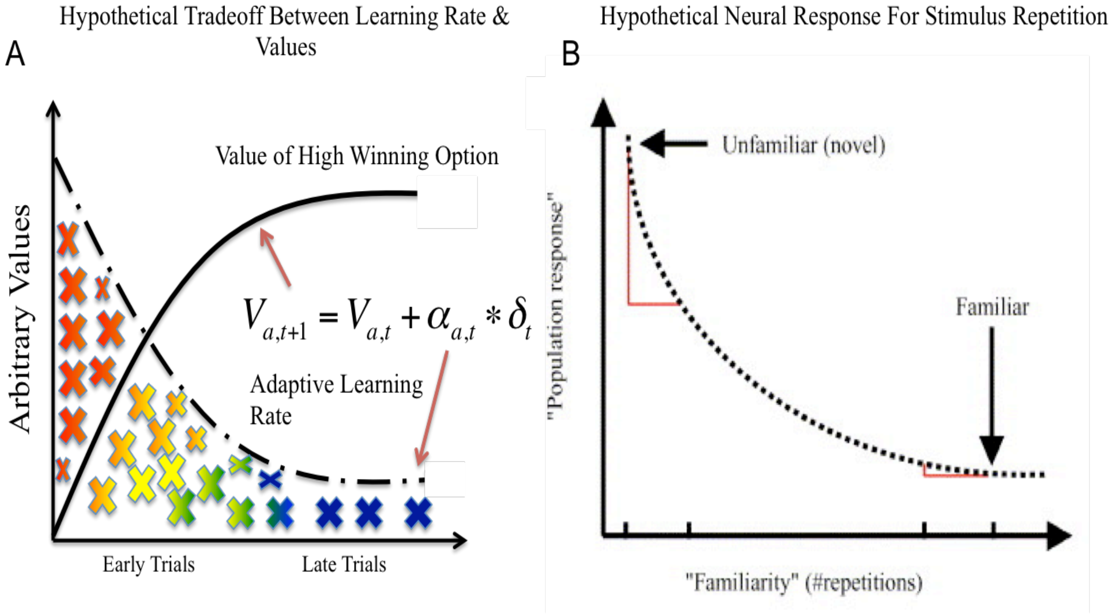

Hypothetical Tradeoff Between Learning Rate & Values
A
Value of High Winning Option
Arbitrary Values
$V_{a,t+1} = V_{a,t} + \alpha_{a,t} * \delta_t$
Adaptive Learning Rate
Early Trials
Late Trials
Hypothetical Neural Response For Stimulus Repetition
B
Unfamiliar (novel)
"Population response"
Familiar
"Familiarity" (#repetitions)

Figure 4.2 (A) The hypothetical trade off between the value of an arbitrary stimulus "a" and the adaptive learning rate. Coloured regions indicate hypothetical activity shift in the brain from more rostral-executive regions towards caudal-motor regions. Red crosses under the curve indicate the larger learning rates and rostral activity in the cortico-striatal loop whereas blue regions indicated small learning rate in the sensori-motor regions. **(B)** Schematic representation of Wiggs and Martin (1998) shows that neural responses decrease over multiple successive repetitions.

Based on the evidence presented above and in the previous chapters, we hypothesized that the novelty signal decreases over time and adaptive learning rates might capture changes in attention and novelty fairly well. Also based on the effect of the learning rate on the updating of the prediction error and action values it's highly plausible that a gradual decrease in learning rate may cause a shift in activity in the rostro-caudal axis found in neurophysiological experiments (Graybiel. 2008).

In order to capture the adaptive learning rate, we used a simple adaptive learning rate updating strategy that is based on the following equation:

If $\delta_t > 0$

$$\alpha_{t+1} = \alpha_t - \alpha_{a,t}\theta \qquad [4.29]$$

if

$$\alpha_{t+1} = \alpha_t + \alpha_{a,t}\theta \qquad [4.30]$$

if $\delta_t = 0$

$$\alpha_{t+1} = \alpha_{a,t} \qquad [4.31]$$

According to the equation 4.29 above, if a prediction error is greater than zero it indicates that something better than expected is happing and the agent should decrease the learning rate to increase the convergence. Here, θ indicates a fixed meta-parameter controls the changes in learning rate. On the other hand according to equation 4.30, if a negative prediction error happens the agent should increase the learning rate because it indicates that something worse than expected is happening whereas if the prediction error is zero the learning rate doesn't changein the following trial. In general the change in weights is implemented by the following rule.

$$\Delta V_{a,t} = \alpha_{a,t}\delta_t \qquad [4.32]$$

The advantage of this updating technique is its simplicity because it includes only two free parameters the initial learning rate α_0 and the meta-parameter θ. Regardless of its simplicity it captures the basic properties of the dynamic learning rate that is it decreases when the stimulus becomes familiar and this is based on the quality of the predictions the organism is making. On the other hand the current learning rule explained in equation 4.32 is different to that of Pearce & Hall's (1980) learning rule because the learning rate in the above equation is based on the signed prediction error rather than the absolute value of the prediction error. This have some neurobiological implications as explained in **Chapter 3** where positive and negative prediction errors project to direct and indirect basal ganglia loops. **Figure 4.3** shows an example of a simulation of the learning rate model described by the equation 4.32 for a binary choice probability learning experiment repeated for 20 trials where one of the options gives a reward with a probability of 0.8 and the other option gives a reward with a probability of 0.2.

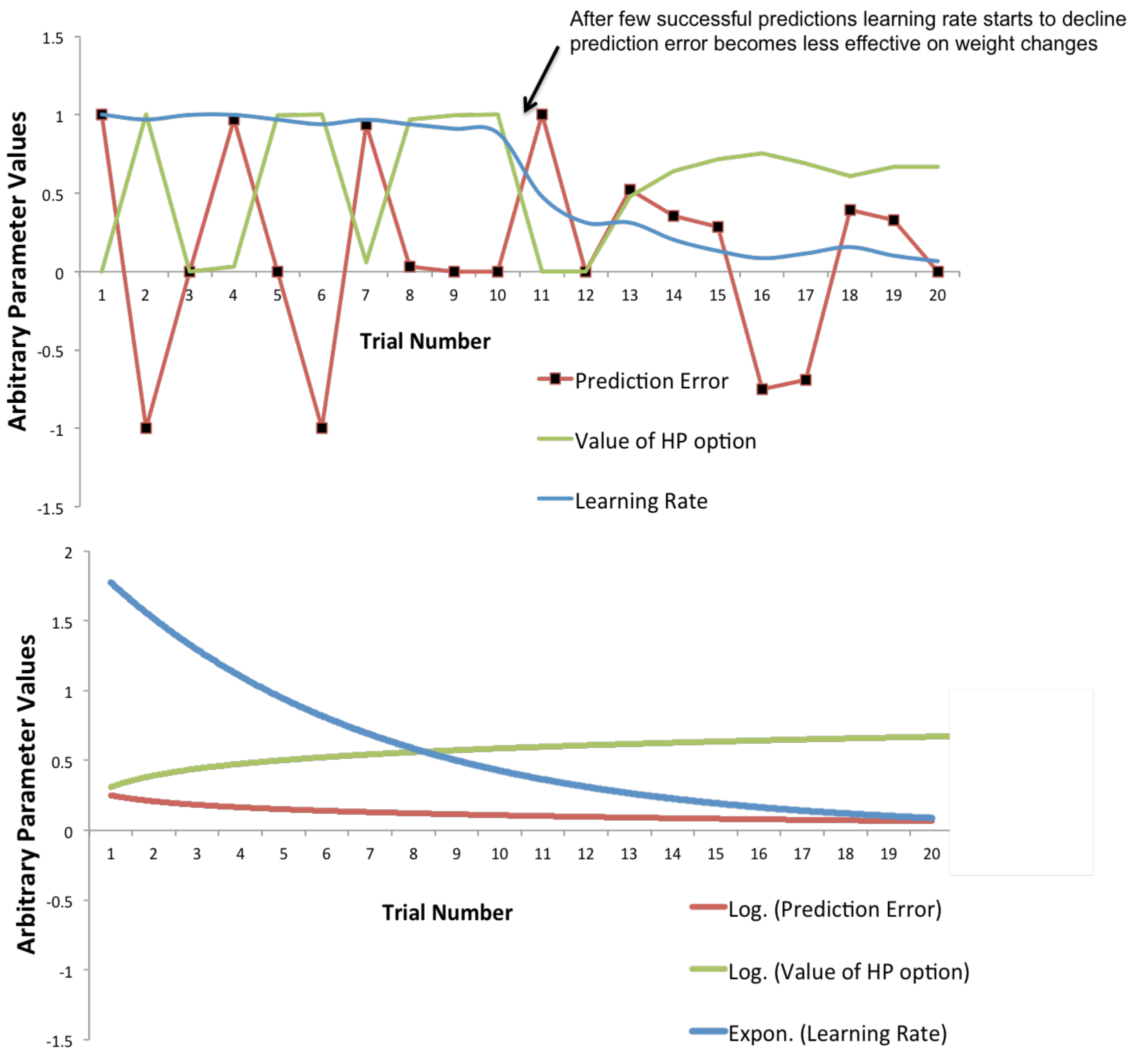


Figure 4.3 The results showed that over the course of the learning, the learning rate decreases over consecutive trials and in the later trials the value of high probability option didn't effected from the prediction error as significant as in the early learning trials. During this simulation θ was equal to 0.4. Please note that all the values relate to prediction error, value and learning rate are discrete variables although it was presented as a continuing line to make it visually comprehensible.

4.5.2 BOLD Correlates of Learning Rate

Numerous imaging studies utilized fixed learning rates in the model fitting procedures (O'Doherty et al., 2007). Fixed learning rate models have been successfully used to find neural correlates of certain hidden variables like the prediction error responses (O'Doherty et al., 2007) and have various advantages when comparing the differences in learning rate of different populations when the research question for example is to understand the learning deficits of a particular group or subpopulation (see for). However, for the reasons mentioned in the previous sections fixed learning rates don't capture certain behavioural and physiogical findings and so some researchers have utilized adaptive learning rate models in model-based imaging studies (see, **Chapter 5**)

Recent neuroimaging studies found correlations between a measure of learning rate and BOLD activation in the anterior cingulate cortex, frontal cortex, and basal ganglia (Haruno et al., 2004; Behrens et al., 2007; Brown & Braver, 2008, Krugel et al., 2009).

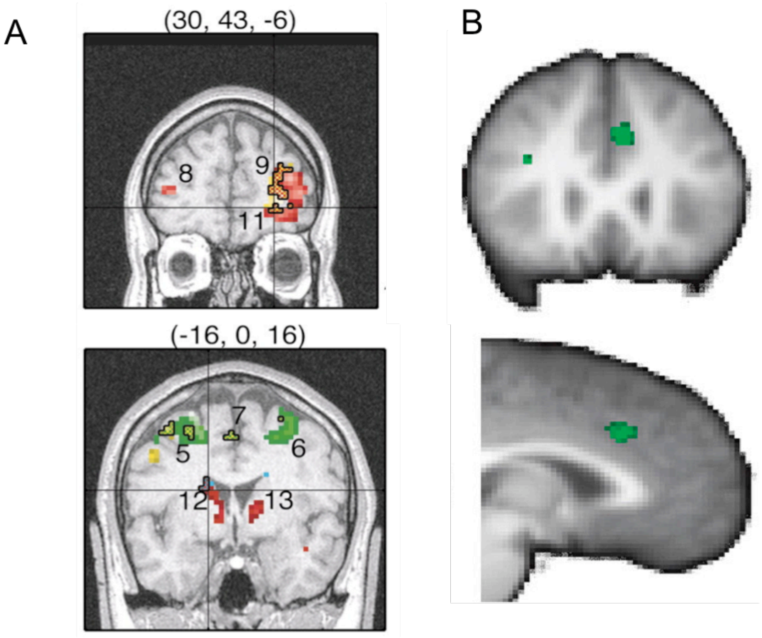


Figure 4.4 a) The brain regions shown in red show the neural correlates of dynamic learning. These brain regions include the dorsa-lateral prefrontal cortex, and the basal-ganglia. The figure taken from Haruno et al., (2004). b) The brain regions shown in green shows the neural correlates of the volatility signal that is calculated from a Bayesian dynamic-learning rate formulation. The regions include the posterior cingulate cortex. The figure is taken from Behrens et al., (2007).

In addition, it has been shown that patients with Parkinson's disease are impaired in inhibiting previously learned stimulus values in a probabilistic reversal-learning task, due to a decreased learning rate caused by an impaired dopaminergic system (Rutledge et al., 2009). Moreover, several researchers suggested that dopamine neurons are directly involved in coding the learning rate (Friston, 2009).

Chapter 5

5 Methods: Functional Magnetic Resonance Imaging

5.1 fMRI and Physics of The BOLD Signal

5.1.1 Physics of the BOLD Signal

fMRI measures the neural activity in brain regions indirectly (Logothetis, 2008). When action potentials occur in neurons they consume energy in the form of ATP, which is created from ADP by oxidative phosphorylation (Mitchell and Moyle, 1967). However, in order to do oxidative phosphorylation the cells need oxygen, which is delivered through blood by a protein called haemoglobin. In the center of the haemoglobin there is an iron atom and when oxygen binds to haemoglobin it becomes oxyhaemoglobin or called oxygenated haemoglobin a diamagnetic material. But when the haemoglobin is de-oxygenated it becomes paramagnetic (Pauling and Coryell, 1936). This paramagnetic effect of de-oxygenated haemoglobin has %20 greater magnetic susceptibility than oxygenated haemoglobin and decreases the transverse magnetization of T2* (MRI pulse sequence for functional imaging) weighted images. Therefore, increased oxygen in the blood reduces the MR field inhomogeneity by reducing the concentration of magnetized materials (de-oxyhaemoglobin), but in turn causes a 2-4% increase in the intensity of the T2* weighted functional images (Ogawa et al., 1990;

Turner et al., 1991). The observed changes in the magnetic field of oxygenated-deoxygenated haemoglobin ratio is called the blood oxygenated level dependent signal (BOLD) and is represented by the haemodynamic response function (neurophysiological model of ideal BOLD signal). However, later studies showed that the BOLD signal observed in imaging studies is correlated with local field potentials (LFP) rather the single neuron activity (Logothetis et al., 2001; Logothetis and Pfeuffer, 2004). The BOLD signal is slow in comparison to the electrical signal and usually takes 4-10 seconds to detect (Berens et al., 2010). In order to increase the signal to noise ratio fMRI experiments are usually designed to maximize the detection of the small signal variations of different regions resulting from different variation of local field potentials (Dale, 1999; Wager and Nichols; 2003; Henson, 2006).

5.1.2 Dopamine and the BOLD Signal

Several studies have suggested a direct relation between the observed BOLD signal and the quantity of dopamine neurotransmission in the prefrontal cortex (Krimer et al., 1998). However, Duzel et al. (2009) argued that the change in BOLD signal might be caused by various physiological sources of dopamine activity. For example, they argued that glutamatergic inputs can activate tonically active or silent DA neurons which can change the local field potentials and might influence the BOLD response. Also it is possible that the BOLD signal can be changed by phasic burst firing DA neurons. However, the actual source of dopamine activity in BOLD signal remains unknown and needs further research.

5.2 Analysis of fMRI data using Statistical Parametric Mapping

5.2.1 Design principles

In block designs, participants are asked to perform a condition, for example task A, which engages the targeted brain regions for about a block of time varying from 10 sec to several minutes. The task condition sometimes involves multiple presentations of the stimuli or multiple responses from the subject in a single block, so that sustained activity can be captured. After task A is over, the subject is asked to perform another task, B, in a similar blocked fashion which presumably disengages the targeted brain region for about the same amount of time. During the design of these two tasks, perceptual differences between tasks are minimized such that the different cognitive functions required for performing tasks A and B are reduced to the targeted function relevant to he research question. The period during which the engaging task is performed is called the ON period, and the period of the disengaging task is called the OFF period. ON-OFF periods are repeated back to back for N cycles. Multiple 3D fMR images are collected for each period, and in the final analysis, the "active" voxels in the brain are determined by looking at the statistical significance of the difference between the intensitiesobserved in the ON periods versus the OFF periods.

In the literature, a number of studies used block designs to show reward related activity. For example, in one study investigators used a block design and looked for the differences between romantic love and maternal love by showing pictures of the participants' partners or family members (Bartels & Zeki, 2004). Although this study did not use a learning paradigm, experiments that use block design fail to differentiate between the activities induced by conditional and unconditional stimuli or between a decision and its outcome since both events are presented together in the same block.

On the other hand a growing number of event-related fMRI studies on reward learning have been conducted in the last decade, allowing investigators to study a range of hypotheses related to reward detection and prediction. These findings importantly extend prior findings by allowing investigators to dissociate different phases of the reward process, which is not possible with block designs that cannot differentiate within-task differences. In the simplest event-related fMRI studies, a single trial consists of the delivery of a CS followed shortly afterwards by a US, and then the subject's response time in the case of instrumental learning.

In event related designs several 3D fMR images are collected for each single trial so as to allow for the observation of the transient change in the haemodynamic (BOLD) response, as well as the observation of timing differences in the multiple brain regions that are recruited to perform the given task. Also between trials there is a random inter-trial interval (jitter) allow separation of different type of events. However, in this paradigm, because the observed activity is transient rather than sustained, the magnitude of the signal is much smaller (0.5-1.5%) and fMRI images need to be tightly synchronized with the brief stimulus delivery.

Historically, block design paradigms were used in the early fMRI studies, because the subtraction of sustained activities measured during two opposing types of task enhanced the signal contrast and allowed overcoming many technical limitations. Upon the introduction of higher-strength scanners and improved synchronization between the task and scanner data acquisition, event-related studies emerged and became widely accepted in the field. Nowadays, most recent reinforcement learning experiments use an event-related design, which gives greater flexibility to the investigator for manipulating the independent variables. Therefore, event-related designs make much better estimations than block designs. Therefore, block designs are not very practical for decision-making and reinforcement learning experiments if the aim is to investigate complex decision variables within a reward-related learning paradigm rather than try to identify the effect of changes in a single independent variable.

5.3 Pre-Processing

5.3.1 Spatial Re-Alignment

Although in most fMRI studies head movements of participants are restraint, displacements of head motion occurs in each scan for about few milimeters. Even though this is very small, it can cause significant changes in the observed BOLD signal. Realignment of fMRI images involves correcting the functional images for head motion by using a 6 parameter (3 translation in x,y,z coordinates and 3 transformation) "rigid-body" transformation. In rigid-body transformation displacements in successive scans are calculate by minimizing the sum of squared differences between the reference scan (mean of all scans or the first scan) and successive scans. Then transformations arethan applied to each re-sampled image by using tri-linear interpolation (or spline) (Friston et al., 1995). However, this re-alignment cannot fix the movement related signal changes in functional images. For this reason, these movement parameters are later used as a covariate in the general linear model (Friston et al., 2006) to evaluate whether signal changes result from head movement per se.

5.3.2 Co-Registration and Spatial Normalization

After realignment coregisteration is applied to the anatomical image, which refers to the process where the anatomical images are registered onto functional images. Then spatial normalization applied to the data. Spatial normalization is the process where each participant's brain is normalized to a standard anatomical space by using a template image (e.g, Montreal Neurological Institute, MNI Template).

5.3.3 Spatial Smoothing

Spatial smoothing refers to applying a Gaussian kernel to haemodynamic responses for each voxel. Although there are various reasons for applying spatial smoothing the most important reason is perhaps inter-subject averaging which is due anatomical difference between individuals brain (Friston et al., 2006). Also the size of spatial smoothing applied to data may vary depending on the size of the effect the researchers expects (Friston et al., 2006).

5.3.4 Statistical-Analysis: General Linear Model

After pre-processing a statistical analysis is carried out to identify active voxels (a 3D volume of stored T2* weighted images together written as time series information) that respond to stimulus. The most standard way is to analyse voxel time series in a univariate way (treating each voxel separately) is general linear modelling (GLM). The simple formulation of the general linear model is shown by the following equation when there are two stimulus types x_1 and x_2:

$$y(t) = \beta_1 * x_1(t) + \beta_2 * x_2(t) + c + e(t) \qquad [5.1]$$

In the above equation $y(t)$ is the BOLD response in the observed data for a single voxel. $x_1(t)$ and$x_2(t)$ are the stimulus functions that are used in the design matrix. For example if $x_1(t)$ refers to the appearance of a stimulus on the screen it can be represented by 1 as stimulus "on" condition and 0 as stimulus "off" condition. Then for the entire time series of for example 56 seconds (each digit represents a typical 3 second TR time, which is the collection of a single brain volume) will look like the following 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 where the stimulus appears after the first 18 seconds. Than in order to get a good fit the stimulus function is convolved with the haemodynamic response function. β_1 and β_1 are the parameters to estimate for the model fit. If a particular voxel is active for stimulus type x_1 it's model-fitting will find a high β_1 and if a particular voxel is active forstimulustype 2 then model fitting will find high β_2 value. The term 'c' in the equation is the constant such as the baseline and the term 'e' is the residual error between the fitted model and the data. After finding the best fitted parameter estimates for β's separately for each voxel they are converted in to T-values by dividing the parameter estimates with its' standard error (derived from the variations of β across whole time series) to use in statistical tests. Then the results of each participant the results are combined to provide a second-level analysis for group comparison.

5.3.5 The Multiple Comparison Problem and Significance Thresholding

The multiple comparison problem refers to the situation when the standard significance results (e.g., $p<0.05$) are not acceptable. This problem is due to the total number of voxels in the brain. For example, if there are in total 20000 voxels in a brain and a confidence interval of 95% is used, 1000 of these voxels might show significant activity by chance. In order to decrease this several researchers suggested using Bonferroni correction (see for a review Bennett et al., 2009), but this is also problematic given that it is too conservative (P value is divided by the total number of voxels, 0.05/20000). Recent studies based on the Gaussian random field theory suggested considering the size of the cluster of activation using the false discovery rate (FDR) in which the probability of type 1 error is matched with the type 2 error.

5.3.6 Parametric Modulation of the Stimulus Function

In the previous section when the stimulus function was introduced it was defined as vectors consisting of ones and zeros. However, it is possible to assign different values other than ones and zeros to different individual trials. These different values modulate the weights assigned to the heamodynamic response function in a given trial and reffered as parametric modulation of the stimulus function (see for deails Friston et al., 2006). Parametric modulation can be used in many areas, for example it can be used to model the effect of a linear-nonlinear increase or decrease in stimulus intensity or it can be used to add the effect of participants reaction time in each trial to modulate the weight of the heamodynamic response function (Friston et al., 2006). But more importantly as was explained in **Section 5.3**, it can be used in model-based fMRI in order to modulate the haemodynamic response function according to estimates of the hidden model variables that are derived from fitting participant's behavioural data to a computational model.

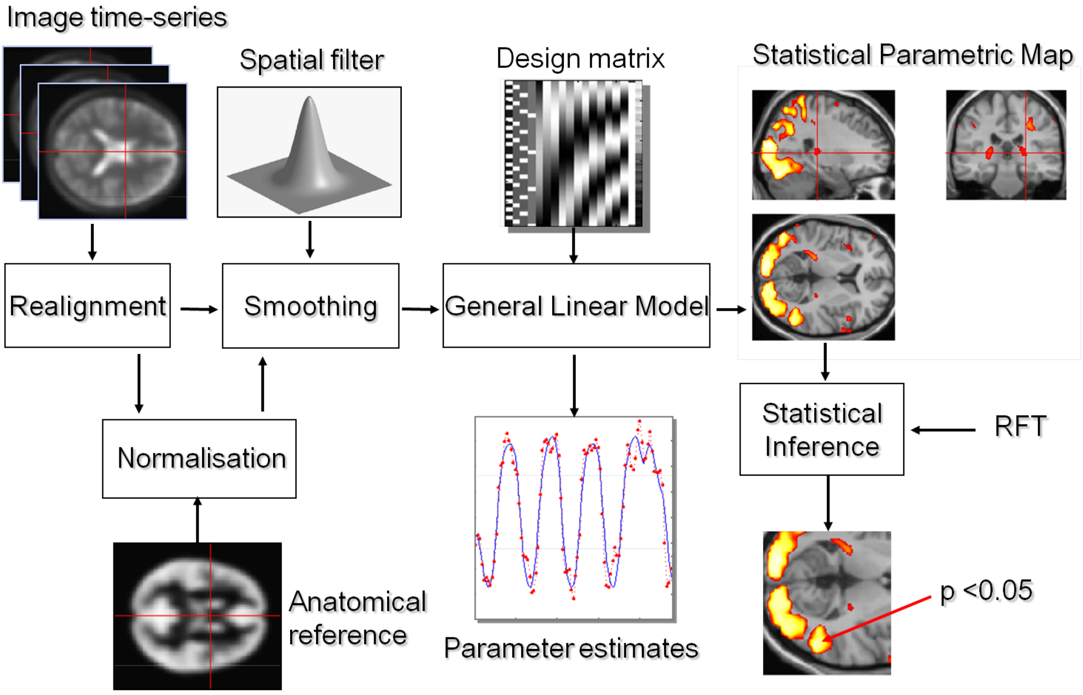


Figure 5.1 Pre-processing and statistical analysis steps of fMRI datausing statistical parametric mapping. Taken from Friston et al., (2006).

5.4 Model-Based Analysis

The classical correlative paradigm in fMRI simply refers to the manipulation of the independent variables of interest and observing the changes in BOLD response. Even though model independent paradigms (e.g., epoch analysis) have been useful and are still used by many researchers (e.g., Bartel and Zeki, 2000; 2004), it is not good enough to understand value based decision-making. According to O'Doherty et al., (2007) most human decisions are usually guided by subjective variables that are not directly observable or controllable by the experimenter. These type of variables might depend on a variety of factors such as the subjects' choice history or reward experience and computational models of cognitive processes compute such hidden variables (Corrado & Doya, 2007; O'Doherty et al., 2007). Examples of model-based analysis can also be found for electrophysiological recording studies from behaving monkeys (Samejima et al., 2005). For example, at the electrophysiological level the variables that affect the valuation processes considered as transient neural firings where the average spiking rate of neurons change from trial to trial depending on the value of that variable (Samejima et al., 2005). Also these subjective variables differ proportionately to individual differences (e.g., the learning rate). These types of questions have guided researchers to use solutions like model-based techniques. The essential point of the model-based analysis is not whether the brain uses that particular model or not, but most importantly it provides a framework for interpretation and therefore study hidden decision variables and their neural correlates that are critical for learning (Corrado & Doya, 2007).

The central approach in model-based fMRI is to use the behavioural responses of a participant to estimate the values of the hidden variables of a model over time. In the model-based analysis subjects' behavioural responses were entered into a computational model, and the computational model calculates the proxy subjective decision variables such as the prediction error response (see, **Figure 5.2**).

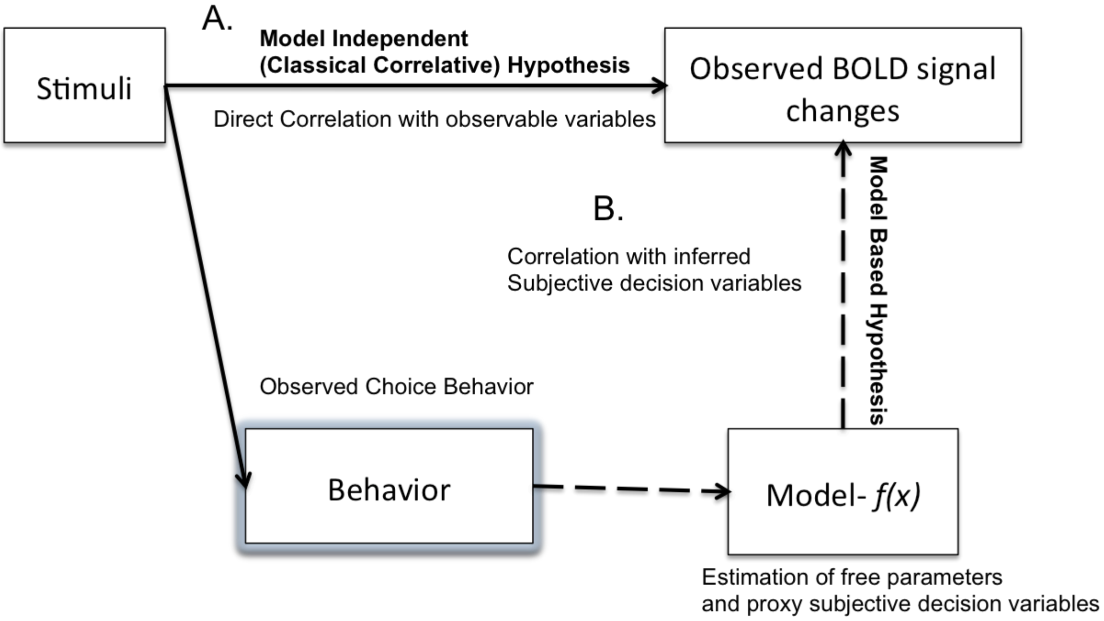

A.
Model Independent
(Classical Correlative) Hypothesis
Stimuli
Observed BOLD signal changes
Direct Correlation with observable variables
B.
Correlation with inferred
Subjective decision variables
Model Based Hypothesis
Observed Choice Behavior
Behavior
Model- *f(x)*
Estimation of free parameters
and proxy subjective decision variables

Figure 5.2 A) In the model Independent classic correlative paradigm, the observable variables are directly correlated against fMRI data. B) In the model based analysis the hidden (proxy) variables were calculated from the behavioural responses of the participants and then convolved against fMRI data.

However, in every model there are free parameters such as the learning rate or exploration rate in the case of reinforcement learning algorithms that need to be calculated with model fitting techniques (e.g., maximum likelihood, or mean least-squares procedure). After all the decision variables are identified and trial to trial values are estimated then they have to be convolved with the hemodynamic function by using parametric modulation (see, **section 5.2.3**) and regressed against the observed bold signal. The hidden variables are then correlated with fMRI data (see, **Figure 5.2**).

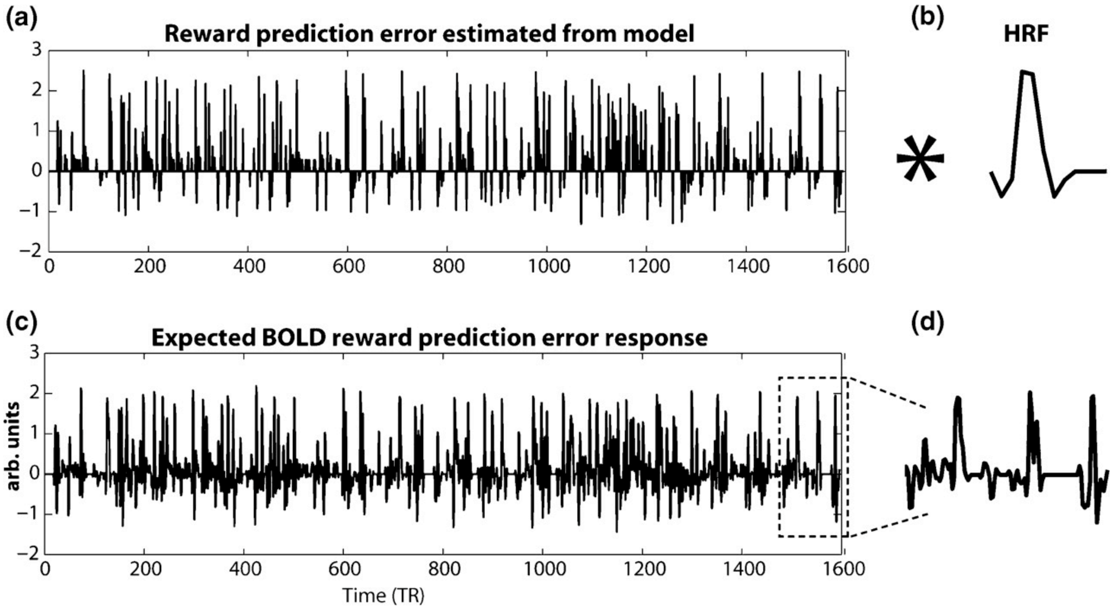


Figure 5.3 Parametric modulation for the prediction error signal. A) The stimulus function parametrically modulated by the prediction error signal. B) The prediction error is convolved with the heamodynamic response function. C) Expected BOLD response after convolution. D) Taken from Cohen (2007).

In spite of differences in their choice of model, most fMRI studies use similar analysis procedures (Friston et al., 2007). In the standard data analysis procedure, images are realigned, spatially normalized to a standard template (for instance MNI or Talairach) and spatially smoothed with a Gaussian kernel. Later, the time series in each session is high-pass filtered to remove potential slow scanner drift or low frequency noise such as heart beat (Friston et al., 2007). After this standard analysis procedure, a statistical linear regression model is fitted to the data. At this point, each trial is represented in the design matrix and the prediction error signal is treated as a parametric modulator to the design matrix. However, the prediction error signal has to be computed separately and most probably by a second party program (e.g. Matlab). One of the most important issues in calculating the prediction error signal is the process of finding the best choice for free model parameters, such as the learning rate a (see **Chapter 4**). There are a number of methods being used in the literature, likelihood estimation (Pessiglione et al., 2006) and particle filtering (Samejima et al., 2004: Samejima et al., 2005) being the most popular.

One last important issue in model based fMRI analysis is to compare different models (Pitt & Myung, 2002). As one can imagine there could be many mathematical or computational models that can explain the behavioural data. The responsibility of the researcher is either to test the significance of a particular model or if he/she is not sure about which model is more biologically plausible he/she might want to compare different models by using model comparison techniques such the Akaike's information criterion (AIC) or Bayesian information criterion (BIC) (Pitt & Myung, 2002).

Chapter 6

Experiment 1: Learning Actions from Rewards and Punishments

6.1. Analysis for Gains and Losses during the Outcome and Expectation Periods

6.1.1 Introduction

Recent imaging studies in humans and neural-recording studies in primates and rodents revealed the neural correlates of reward and punishment related processes in various parts of the brain mostly reporting activity in the basal ganglia and frontal cortex emphasizing the involvement of dopamine in learning rewards and punishments (see, **Chapter 2**). However, controversial suggestions have been made concerning the neural correlates of anticipation, and outcome related activity for monetary rewards and punishments (also see **Chapter 2** for a detailed discussion). For example, according to some studies, outcomes of monetary rewards and punishments activate similar fronto-subcortical networks including ventral striatum (Liu et al., 2011) and some other studies it was suggested that gains and losses are represented in different neural systems including bilateral amygdale (Yacubian et al., 2006), insula (Pessiglione et al., 2006) and antero-medial cingulated cortex (Shakman et al., 2011) which were described in more detail in **Chapter 2.** Therefore, the main research questions that are tested in this section are as follows: Does feedback involving monetary rewards and losses activate a similar or separate fronto-subcortical network during the outcome phase? Secondly, does the activity for expecting potential rewards and losses overlap the same brain regions and therefore share common neural systems ?

Based on the literature described in **Chapter 2** it was hypothesized that during the anticipation phase rewards and punishments share activity in similar striatal regions due to common dopaminergic striatal representation of motivational saliency. In addition it was also hypothesized that some distinct brain regions should also get activated during the outcome phase because of the emotional feelings that occur due to negative or positive valence of the feedbackwhich is also proposed as stimulus specific representations of unconditional stimuli in the Konorskian opponency model (for details see **Chapter 2**). Therefore partial overlapping and segregated regions were expected for rewards and punishments during the outcome phase.

It is also important to note that this is an exploratory experiment where different statistical thresholds were used for different contrasts in order to see the extent of activation for a particular contrast. This is due to different number of trials in each contrast (e.g., reward trials compromise 80% of all gain trials whereas punished trials compromise almost 20% of all loss trials). Therefore one should be cautious in comparing different contrasts. In the following analysis the thresholds were set to p<0.005 and p<0.001 (uncorrected). The rationale for using those particular statistical thesholds was based on Lieberman and Cunningham (2009) which suggested that using a more liberal statistical theshold of p<0.005 is also acceptable when there are not many participants (n>20) and in fact if the cluster theshold (k) is set to k>20 voxels it is equal to the corrected FDR p-value of 0.05.

6.1.2 Materials and Methods

6.1.2.1 Participants

Fifteen right-handed healthy normal volunteers (8 male, 7 female; mean age 25, range: 22-28) were recruited to the experiment but only 12 participants (6 male, 6 female) included in the analysis. Three of the participants were excluded from the analysis due to excessive movement inside the scanner (movement greater than 6 mm) in one case, and the other two participants were excluded due the loss of behavioural data. The participants were pre-assessed to exclude those with a prior history of neurological and psychiatric illness. All participants filled a written informed consent form before fMRI measurements and all the participants were invited by both written and verbal requests which outlined the purpose and nature of the study, before the fMRI session and they were debriefed after experimental session and paid according to their performance in the task. The study was approved by the Bedfordshire NHS Ethics committee board and Local Ethics committee.

6.1.2.2 Task

The whole experiment consisted of 3 sessions each separated from the other with an average of ~2 min. In each session the colour of the stimuli indicated the trial type where the colour and type of the stimuli differs for each fMRI session for reward and punishment trials except for the neutral trials that remained the same for all three sessions (see, **Figure 6.1)**. Hence in each session the participants have to learn from the scratch which colour indicates which trial type. Within the sessions, each trial is an instrumental learning task involving monetary feedback. Each trial began with simultaneous presentation of one of three pairs of stimuli (all symbols were letters taken from Agathodaimon font) and each pair of symbols signified the onset of three trial types: Reward, Avoidance, and Neutral whose occurrence was fully randomized throughout the experiment. The participant's task was in each trial to choose one of the two symbols by selecting the right or the left key button from the response box. For each pair of stimuli the position of the symbols (right or left) is also counter balanced within the session. When the trials started, a fixation cross (null event) was shown at the centre of the screen indicating the start of the trial. The fixation cross-stayed on the screen for 0.5 s. This was replaced by the conditional stimulus (two symbols) presented on the screen to the left and right of where the cross-had previously been for 4s. The participants had to choose which of these two symbols would be rewarded in this 4s time period. Once the symbol was selected, the chosen symbol was shown by an arrow for 0.5 s and it was followed by the outcome. Between the outcome and selected symbol screen there was a random inter stimulus interval (ISI) of about average ~2s for the scanner trigger. The outcome for the participants' choice reward (£1), punishment (£-1) and neutral was shown on the screen for 3s. When the participants failed to press either button they were instructed at the outcome feedback that they will receive a neutral outcome for the gain pair or (£-1) for the loss pair. All three trials types were pseudo randomly intermixed throughout the three sessions. In the reward trials, when the participants choose the high probability of symbol they received monetary reward with 0.8 probability and received neutral feedback with a probability of 0.2. On the other hand, following the choice of a low probability symbol, participants received a reward with a probability of 0.2 and neutral outcome with a probability of 0.8. Similarly on the loss trials, if participants choose the high probability symbol, they received neutral outcome in with 0.8 probability, whereas the choice of the low probability symbol led to a loss of with a probability of 0.2, while the low probability symbol gave a loss of (£-1) with probability 0.8 and a neutral outcome with probability 0.2. On neutral trials participants always received a neutral outcome independent of the symbol choice. Allparticipants underwent three ~13 min scanning sessions, each consisting of 60 trials

(20 trials per condition). Prior to the experiment participants were instructed that they would be presented with three pairs of stimuli where the colour of the stimuli would indicate whether it was a gain trial, loss trial or a neutral trial. They were also instructed that depending on their choices they would win money, lose money or get a neutral outcome. They were not told which coloured pair of stimuli was associated with a particular type of outcome. All participants were instructed to win as much money as they could. Before the experiment they were told that they could earn a maximum of £30 if they chose the correct response in all trials otherwise they were told that their earnings would depend on their performance in the experiment.

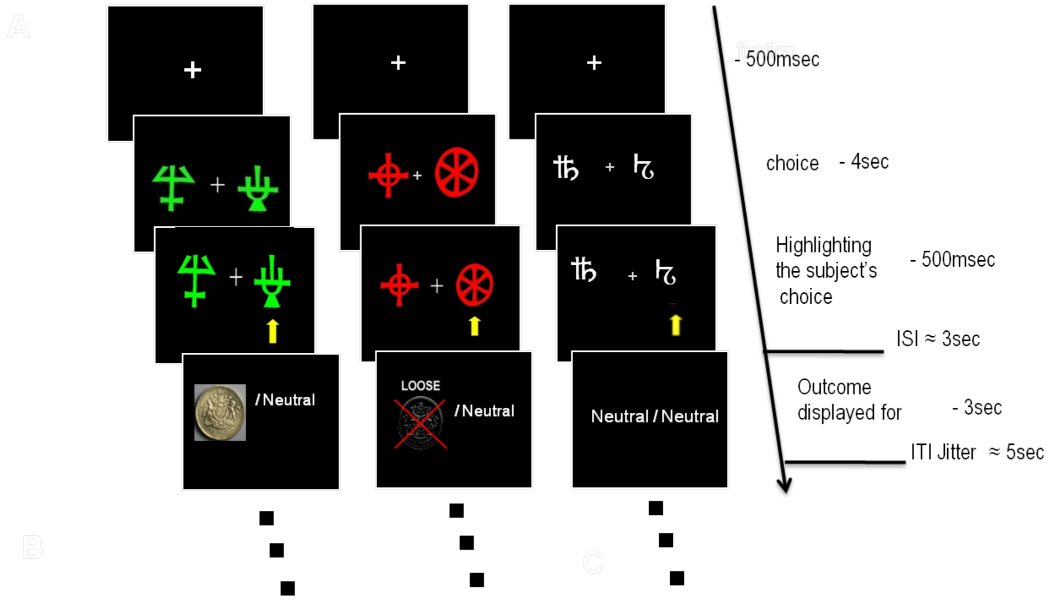


Figure 6.1 Schematic of the experimental design. Three conditions, reward trials (green), avoidance trials (red) and neutral trials (white) were represented by different colours and symbols where they are randomly intermixed during an fMRI session.

6.1.2.3 Functional Magnetic Resonance Image Acquisition

The functional imaging was conducted by using 3-Tesla Siemens MRI scanner to acquire gradient echo T2* weighted echo-planar (EPI) images with BOLD (Blood Oxygenation Level Dependent Signal) contrast. Each slice was collected parallel to the anterior-posterior commissure line. Each volume compromised 36 axial slices of 3–mm thickness and 3-mm in plane resolution with a TR time (repetition time) of 3s. The flip angle was 90 degrees. T1 weighted structural images (1x1x1-mm voxel size) also acquired for each participant. Head movement was minimized with padding participants' head.

6.1.2.4 Functional Magnetic Image Analysis

Image analysis was performed using statistical parametric mapping SPM8 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, United Kingdom) software. For all participants the images were realigned according to the first volume in order to correct for motion in the scanner. For all participants anatomical images were co-registered to functional EPI images and were normalized to a standard EPI template. Spatial smoothing was applied using a Gaussian kernel with full width half-maximum (FWHM) of 8 mm for each participant's data.

6.1.3 RESULTS

6.1.3.1 Behavioural Results

Over the course of the experiment participants showed significant preference for the higher probability rewarding option rather than non-rewarding option, $t_{(11)}$= 21.06, p<0.001, two tailed (**Figure 6.2a**). The high probability reward option chosen more than the neutral option in the neutral trials, $t_{(11)}$= 11.13, p<0.001, two tailed. Participants also avoid choosing the high probability punishing option $t_{(11)}$= 5.48, p<0.001 and show successful avoidance of monetary losses. Probability of choosing the punishing option was significantly lower than choosing the neutral option in the neutral condition, $t_{(11)}$= 4.69 $p<0.05$ two tailed, indicating that they showsuccessful avoidance from monetary losses. As expected, participants preferencefor choosing options in the neutral condition was not significantly different to chance, significance for the least frequently chosen option and most frequently chosen option were $t_{(11)}$ = -1.19, $p>0.05$, two-tailed; and $t_{(11)}$ = 1.19,$p>0.05$, two-tailed respectively.Analysis of the mean reaction time (RT) taken for participants tomake a choice in the avoidance and reward conditionsrevealed that participants had significantly shorter RTs forreward trials than avoidance trials $t_{(11)}$ = 3.45, $p<0.05$, two-tailed; and significantly shorter RTs for reward trials than neutral trials $t_{(11)}$ = 5.46, $p<0.001$, two-tailed. Also comparison of mean reaction times between the avoidance trials and neutral trials revealed that participants responded to avoidance trials significantly quicker than neutral trials $t_{(11)}$ = 2.19, $p<0.05$, two tailed.

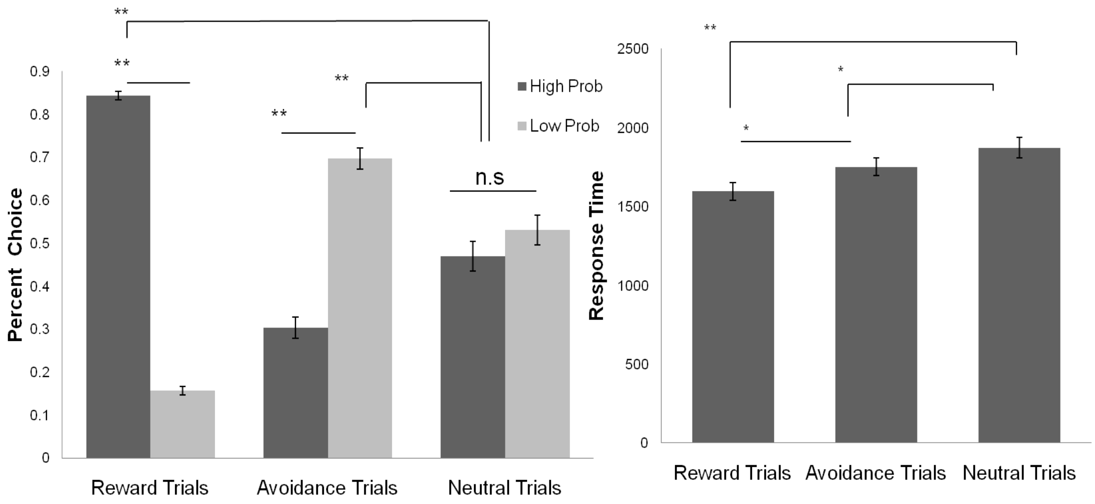


Figure 6.2 a) Behavioural data averaged across all 12 participants showing the percent of responses allocated to high (0.8 probability of getting or losing money) and low (0.2 probability of getting or losing money) probability options for the gain, loss and neutral conditions. Participants choose the high probability rewarding option significantly more in reward trials than the neutral option in the neutral trials and they choose the low punishing option significantly more than the neutral option in the neutral trials (** indicates significance $p<$ 0.001, two-tailed, * indicates significance $p<$ 0.05, two-tailed). Comparison of high and low probability options that provide neutral feedback in the neutral condition is not significantly different than the chance level (n.s indicates significance $p>0.05$, two tailed), which suggests that participants choose randomly in the neutral condition. b) Plot of the reaction times for the three conditions regardless of the outcome and probability of winning. Participants were significantly faster in the reward trials and punishment trials than neutral trials.

6.1.3.2 Functional Magnetic Resonance Image Results

Individual time series data were analysed using a general linear model (Friston et al., 1996). There were eight orthogonal regressors. Regressors of interest were: expectation in gain trials, a stick function at the time of stimulus presentation (con 1); feedback for a correct response on gain trials (ie 1£), or an incorrect response in gain trials ie neutral outcome were modelled with two separate stick functions at the time of outcome (con 2 and con 3 respectively) , expectation in the avoidance trials (con 4), correct feedback as neutral outcomes in the loss trials (con 5), incorrect feedback as -1£ loss in the loss trials (con 6), expectation for neutral outcomes (con 7), and (con 8) neutral outcomes in neutral trials were modelled in a similar way to their counterparts in the gain trials. The motion parameters calculated for the realignment procedure were also included to account for the residual effect of movement (covariates of no interest). All three sessions were included in the analysis of individual results. A random effects analysis for all 12 participants wascomputed for the group analysis (level 2) and the peak coordinates of the significant activations were reported in MNI (Montreal Neurological Institute) coordinates.

6.1.3.3 Brain Regions Involved in Expectations of Rewards and Losses

During the expectation phase in gain trials, results demonstrate a robust activation in the basal ganglia and cingulated cortex. Specifically, foci in the midbrain (T= 4.76; x=0, y=15, z=11, p<0.001, uncorrected), bilateral NAcc (right, T=11.83, 11, 11, 3; left, T=4.09, x=8, y=11,z=1p<0.001, uncorrected), bilateral putamen (right, T=11.26, x=21, y=14, z=-11; left, T=9.8; left x=-24, y=8, z=-11 p<0.001, uncorrected), and right caudate (T=7.29; x=18, y=14, z=4 p<0.001, uncorrected), showed activation correlated with reward expectation. Additionally activation was found in the subcortical clusters specifically in the thalamus (T=5.23;z=7, y=15, z=16 p<0.001, uncorrected), and cerebellum (T=4.68; x=3, y=52, z=-35p<0.001, uncorrected) showed. Cortical clusters in the right (T=4.76; x=-30, y=53, z=4p<0.001, uncorrected), and left (Z=7.15; x=36, y=53, z=4 p<0.001, uncorrected) medial frontal gyrus, and left and right dorsolateral prefrontal cortex (T=6.33; x=30, y=44, z=40 p<0.001, uncorrected), left anterior cingulate cortex (T=5.45; x=3, y=26, z=37 p<0.001, uncorrected) also showed activation.

A similar analysis was performed for anticipated monetary losses. The results showed a similar pattern of activity which included the head of the caudate nucleus bilaterally (T=7.55; right x=15, y=20, z=10; left T=7.75 x=-16, y=14, z=10 p<0.001, uncorrected), bilateral putamen (T=10.04; left x=-16, y=14, z=-2; right T= 7.30 x=18, y=14, z=-5 p<0.001, uncorrected) and cingulated cortex (T=8.33; x=-12, y=14, z=37) show significant activity for anticipated losses at p-value < 0.001 (uncorrected). Activity in these regions increased when the participants saw the conditional stimulus that predicts future losses (**Figure 6.3b**).

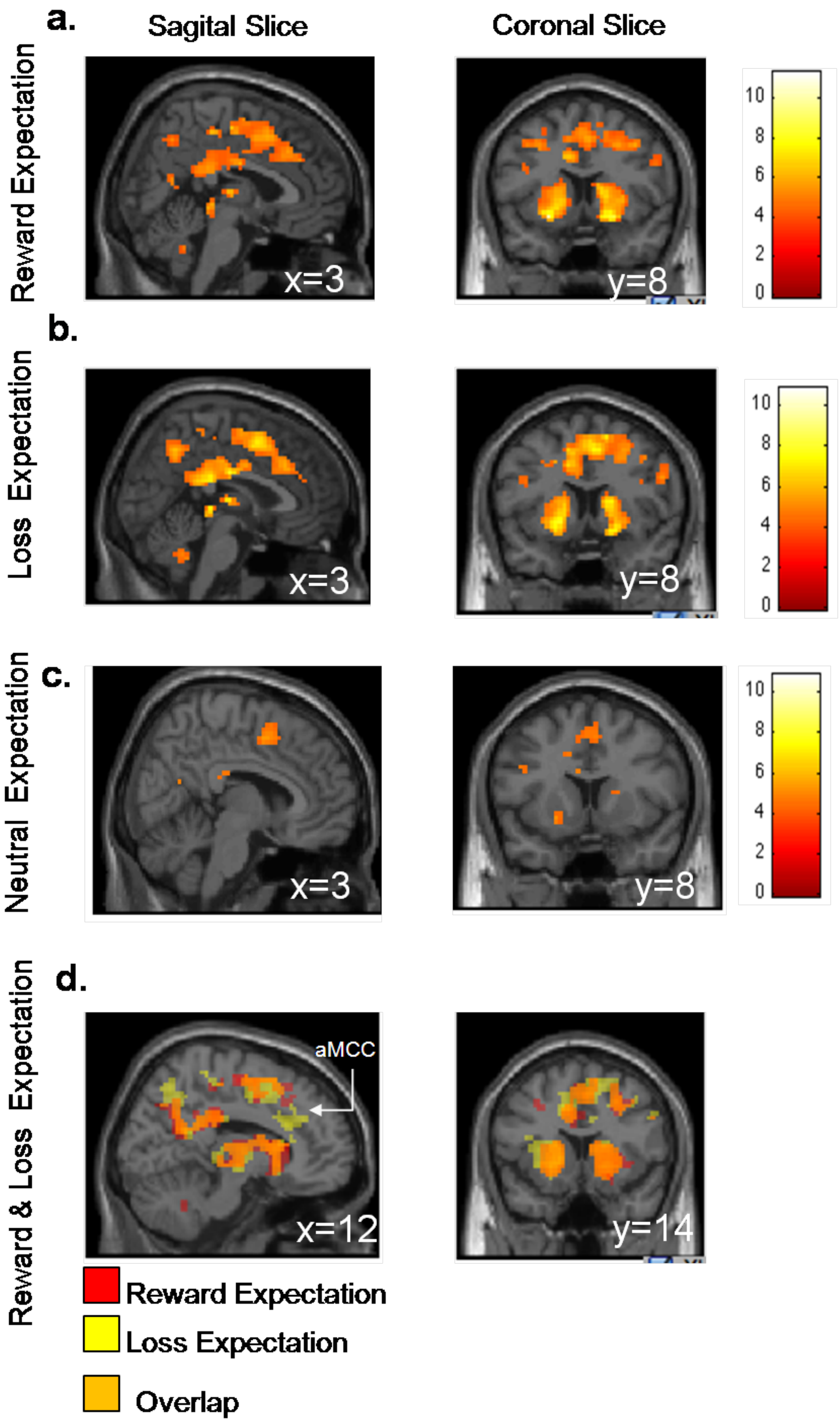

a.
Sagital Slice
Coronal Slice
Reward Expectation
x=3
y=8
10
8
6
4
2
0
b.
Loss Expectation
x=3
y=8
10
8
6
4
2
0
c.
Neutral Expectation
x=3
y=8
10
8
6
4
2
0
d.
Reward & Loss Expectation
aMCC
x=12
y=14
Reward Expectation
Loss Expectation
Overlap

Figure 6.3 Group maps of regions whose activation correlates during expectation of reward, loss and neutral expectation with BOLD responses are threshold at p< 0.001 (uncorrected). All group level activations were overlaid on single subject T1 weighted images. Yellow color signifies high activation (greater z-value), whereas red colour signifies less activation. a) Showing a significant change in activity for expectation of future rewards. b) Showing a significant change in activity for expectation of future losses. c) Showing a significant change in activity for expectation of neutral outcomes. Sagital slice on the left show activity in cingulated cortex and left figure show bilateral striatal**.** d) Figure 6.3a and Figure 6.3b together.Overlay shows activity for reward expectation (red areas) and loss expectation (yellow areas) and the overlapping regions are shown orange.

6.1.3.4 Brain Regions Involved in Expectations of Rewards and Losses but not in Expecting Neutral Outcomes

In order to test for the significance of reward and loss expectation compared to expectation of neutral outcomes, we performed a subtraction analysis of the reward and loss expectation with the baseline of neutral expectation (for the subtraction analysis p value was set to p<0.005 uncorrected). This analysis revealed that regions in the bilateral putamen (right T=3.75; x=27, y=11, z=-2, left T=3.23; x=-21, y=2, z=-2), lateral prefrontal cortex (right T=3.93; x=39, y=53, z=4, left T=4.47; x=-42, y=36, z=-5) and medial frontal cortex (right T=3.83; x=6, y=56, z=-8, left T=4.84; x=-12, y=53, z=-8) show greater response to reward expectation that neutral expectation. Regions with increased activity during trials in which the participant expected to receive losses compared to trials where they expected to receive neutral outcome showed activity in ventro-lateral OFC (T= 3.6 right x=27, y=56, z=-8 T= 3.95 left x=27, y=56, z=-8 T), right striatum (T=3.77, x=27, y=11, z=-5), and the midbrain (T= 3.60 x=9, y=-10, z=-8). Group random effect results with activation maps in the PFC and striatum with percent signal change plots for the peak voxels are shown in **Figure 6.4**.

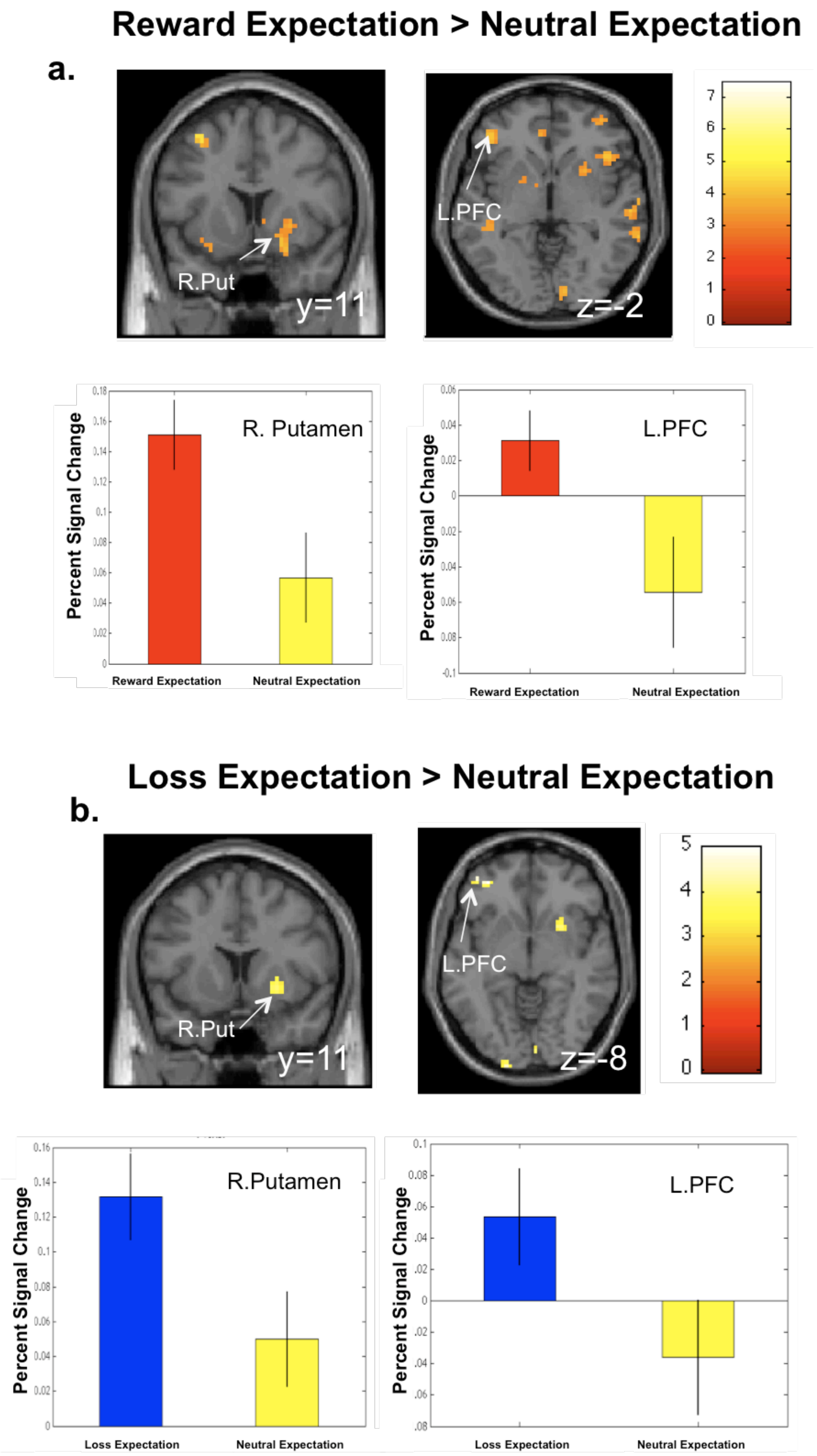
Reward Expectation > Neutral Expectation
a.
R.Put
y=11
L.PFC
z=-2
7
6
5
4
3
2
1
0
Percent Signal Change
R. Putamen
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
Reward Expectation
Neutral Expectation
Percent Signal Change
L.PFC
0.06
0.04
0.02
0
0.02
0.04
0.06
0.08
-0.1
Reward Expectation
Neutral Expectation
Loss Expectation > Neutral Expectation
b.
R.Put
y=11
L.PFC
z=-8
5
4
3
2
1
0
Percent Signal Change
R.Putamen
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
Loss Expectation
Neutral Expectation
Percent Signal Change
L.PFC
0.1
0.08
.06
.04
.02
0
.02
.04
.06
.08
Loss Expectation
Neutral Expectation

Figure 6.4 Blood oxygenation level-dependent (BOLD) responses for reward-expectation > neutral-expectation and loss-expectation > neutral-expectation.The contrasts are thresholded at p< 0.005 (uncorrected) a) Regions of striatum on the coronal slice (left) and regions of frontal cortex shown on the axial slice (left) show greater activity for reward expectation than expecting neural outcomes. The graph below shows the percentage signal changes for reward and neutral expectations. b) Regions of striatum on the coronal slice (left) and regions of frontal cortex shown on the axial slice (left) show greater activity for loss expectation than expecting neural outcomes. Down below the percentage signal change graph on the left for the peak voxels shows differences in percent signal change For both figures group level activation maps overlaid on the single subject structural anatomy. Bars represent means SEM (n12).

6.1.3.5 Brain Regions Involved in Expecting Future Rewards but not in Expecting Future Losses

We performed a subtraction analysis in order to identify the regions that respond more to the cues that predicted future rewards (nb the pair of coloured symbols) versus the cues that predicted future losses. We observedsignificantly activity for reward cues in the leftcaudate nucleus (T=3.85 x=-9 y=5, z=4 p<0.005, uncorrected) right ventral orbitofrontal cortex (T=5.01 x=36 y=44 =-14 p <0.005, uncorrected), bilateral midbrain (T=4.91 left x=-7 y=-20 z-19and right T=3.75 x=12 y=-20 z-22 p<0.005, uncorrected) and left middle frontal gyrus (T=4.24 x=-12 y=29 z40; p<0.005, uncorrected). We couldn't find any significant activity for the opposite contrast loss expectation > reward expectation at the level p<0.005 (uncorrected).

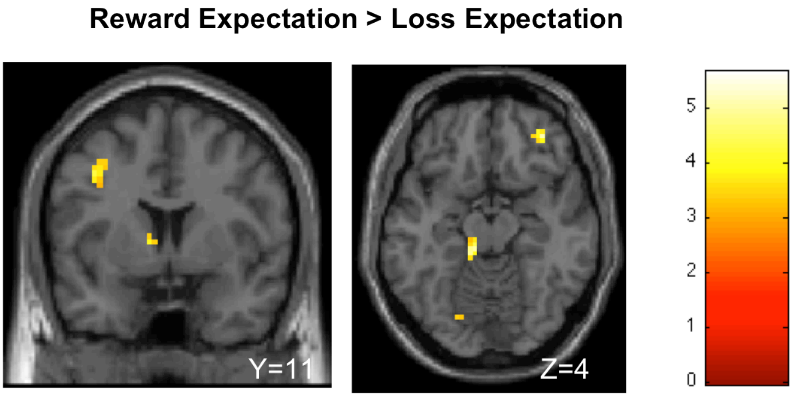


Figure 6.5 Activation of brain regions in the right ventral striatum showing greater activity for reward expectation than loss expectation. Group random effects results are shown superimposed on coronal and axial slices overlaid on a single subject structural MRI image. Significant effects are shown at p<0.005 (uncorrected).

6.1.3.6 Brain Regions Involved in ReceivingRewarding and Punishing Outcomes

In this section, we examined the influence of outcome to investigate whether there is a specific type of neural adaptation depending on the valence of the outcome (i.e., reward or punishment). The outcome phase was defined as a BOLD response evoked by neuronal activity at the moment when the result of a choice was revealed to participants. As a reminder, for the current study, we hypothesized that both the receipt of rewards and losses will activate basal ganglia because of common limbic system representations of motivational salience but they will also activate a differential circuitry because of the subjective emotional component that are created by negative and positive valence of the feedback.

6.1.3.6.1*Neural Responses to Monetary Rewards and Punishments*

First we looked at the contrast for just reward receipt leaving out all the missed reward trials. We found significant activations mainly in the left ventral striatum (T=4.8 x=-15, y=2, z=-2; p<0.005, uncorrected), medial orbito-frontal cortex (T=3.74 x=3, y=56, z=-14; p<0.005, uncorrected) and precuneus (T=5.2 x=12, y=-46, z=37; p<0.005, uncorrected) (see **Figure 6.6**). We also found significant activation in the midbrain for the reward outcome contrast (T=4.4 x=3, y=-19, z=-20;p<0.005, uncorrected) (see **Figure 6.6**). For the loss condition when the participants received a loss outcome we found activation in similar regions of the ventral striatum (T=3.83 x=-12, y=2, z=-8;p<0.005, uncorrected) and in the midbrain (T=6.99 x=6, y=-7, z=-17;p<0.005, uncorrected).

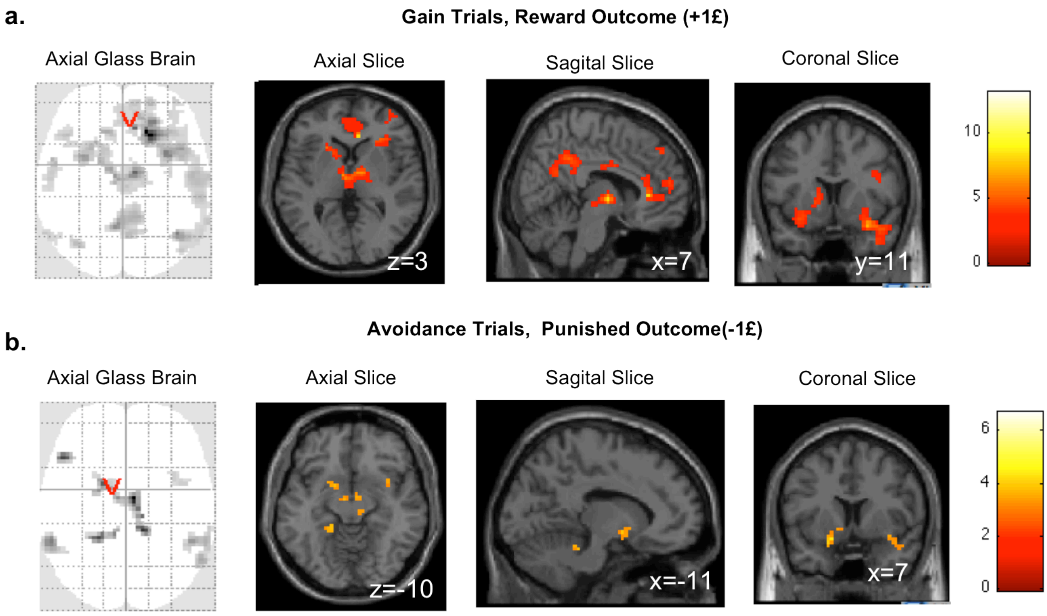


Figure 6.6 Group random effects results are shown superimposed on the axial, sagittal and coronal slices overlaid on a single subject structural MRI image.a)Activation of brain regions in the striatum, midbrain, medial-orbito-frontal cortex and precuneus showing significant activity for reward receipt (+1£).Significant effects are shown at p<0.005 (uncorrected). b) Activation of brain regions in the striatum and midbrain showing significant activity for punishment receipt (-1£). Significant effects are shown at p<0.005 (uncorrected). The image on the left most side of the upper and lower figures shows the activity for those contrasts on the glass brain.

Moreover, to confirm that these results were specific to reward/loss outcomes and not caused by a general feedback effect, we conducted a comparison with both therewarded outcomes andloss-punished outcomes compare with their neutral counterparts (e.g., reward received (1£) > reward not-received (0£)). Surprisingly the results of this contrast revealed activation in different brain regions (see **Figure 6.7**). When we looked at the contrast for the reward outcome that is greater than neutral outcome,we found activity in right hemisphere sub-gyral (T=5.36 x=27, y=29, z=16; p<0.005, uncorrected) (see **Figure 6.7**) and when we looked at the punishment activity that is greater than the neural outcome activity, we found significant changes in the bilateral insula (T=3.9 right x=-43, y=-3, z=7; left T=3.89 x=39, y=-1, z=-13; p<0.005, uncorrected) (see **Figure 6.7**).

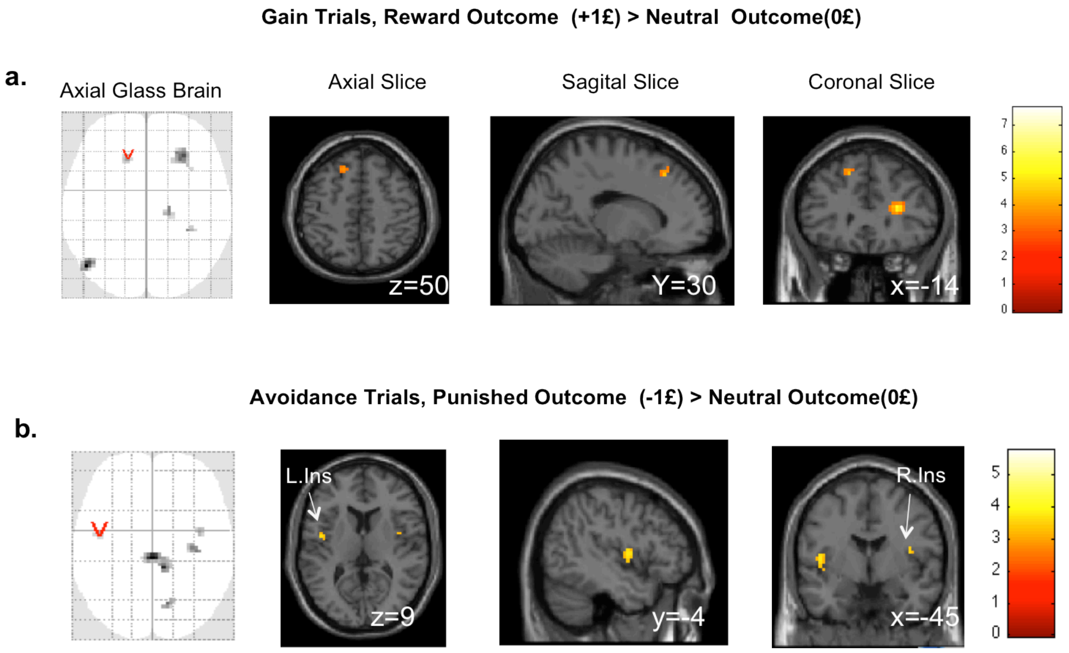

Gain Trials, Reward Outcome (+1£) > Neutral Outcome(0£)
a.
Axial Glass Brain
Axial Slice
Sagital Slice
Coronal Slice
z=50
Y=30
x=-14
7
6
5
4
3
2
1
0
Avoidance Trials, Punished Outcome (-1£) > Neutral Outcome(0£)
b.
L.Ins
z=9
y=-4
R.Ins
x=-45
5
4
3
2
1
0

Figure 6.7 Group random effects results are shown superimposed on the axial, sagittal and coronal slices overlaid on a single subject structural MRI image. a) Significant activity in the bilateral middle frontal gyrus, for reward receipt greater that neutral outcome (P < 0.005, uncorrected). b) Significant activity in the bilateral insula shows regions that are involved for punished trials greater than neural trials (P < 0.005, uncorrected). The image on the left most side of the upper and lower figures shows the activity for those contrasts on the glass brain.

Furthermore, a subtraction analysis was calculatebetween the reward-received and loss-punishment and the opposite contrast between the regions that respond stronger to the punished outcomes than to the reward-receipt was performed. The regions that respond stronger to a received reward contrasted against a received loss are mainly superior frontal gyrus (T=8.8 left x=-18, y=32, z=46; right T=6.26 x=15, y=35, z=43; p<0.005, uncorrected) and broadmann area 6 (T=5.34 left x=6, y=-4, z=64). Whereas the regions that responded more strongly to a received loss contrasted against the reward outcomes are mainly bilateral amygdala (left T=2.3 x=24 y= 0 z=-18, left T=2.6 x=-21, y=-1, z=-23)(see **Figure 6.8**)

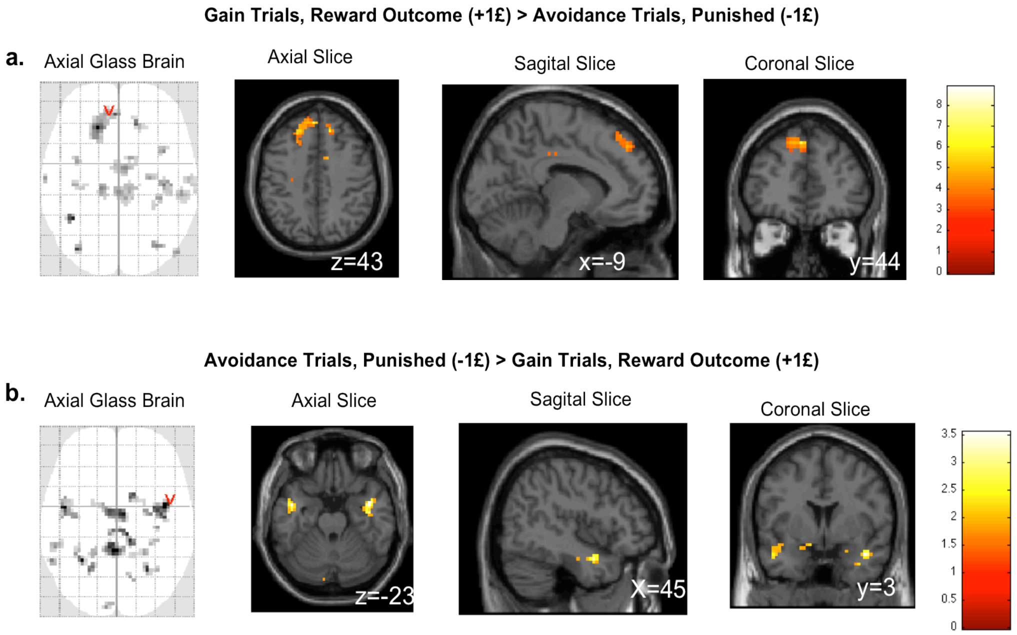


Figure 6.8 Group random effects results are shown superimposed on the axial, sagittal and coronal slices overlaid on a single subject structural MRI image. a) Significant activity shown greater activity in bilateralBroadmann area 9. b) Significant activation left insula, and bilateral occipital activations as revealed by contrast, exemplarily displayed for a coronal slice (P < 0.005, uncorrected). The image on the left most side of the upper and lower figures shows the activity for those contrasts on the glass brain.

6.1.3.7 Discussion

Learning of stimulus-outcome relations critically depends on processing of positive and negative information at various stages of a reinforcement-learning task such as the anticipation phase or the outcome-monitoring phase. As reviewed in **Chapter 2,**the behavioural and neuropsychological evidence suggests that processing of positive (e.g., gain) and negative (e.g., loss) reward information is hard to dissociate within the basal ganglia (Liu et al., 2011) and other parts of the brain (Camara et al., 2009b). In the results of analysis reported above, we tested how much of the activations specific to loss and reward outcomeswere coded in separate regionsor in similar regions during the expectation and outcome phase.

6.1.3.7.1Activity During the Expectation Phase

In the literature, tasks that are similar to the current experiment have demonstrated that the ventral striatum, particularly the nucleus accumbens, showed increased activation for anticipated rewards (Knutson et al., 2000; Ernst et al., 2005; Adcock et al., 2006; Knutson and Gibbs, 2007; Dillon et al., 2008) perhaps given its central role in reward prediction (Schultz, 2002). Similar to the previous studies our study examined brain activation during anticipation of monetary outcomes that varied in their valence (i.e., gain vs. loss). We found activation in these putative reward-related regions, namely the basal ganglia (both ventral and dorsal striatum) increased during both reward and loss anticipation. More over we showed that both expected rewards and expected punishments evoked increased activity compared to control neutral stimuli in the putamen as shown by the **Figure 6.4**. We think that the activation of ventral striatum might reflect the involvement of the dopaminergic system, being a key structure for motivational saliency for both potential rewards and potential losses (Bromberg-Martin et al., 2010).

Moreover during the anticipation period we found activity in the cingulate cortex. Previous studies showed that anterior cingulated cortex is involved in cue evaluation, response selection and conflict resolution (see for a review, Botvinick et al., 2004). Together, these results indicate that anticipatory activation in these regions reflects the motivational properties of the potential outcomes, not their valence because these regions are activated both by theexpected rewards and by the expected losses. Having said that, it is important to mention that the anticipatory activity for potential losses show a more robust activity in the aMCC, which is previously associated with negative emotions and potential future punishments (see **Chapter 2** for a discussion).

6.1.3.7.2Activity During the Outcome Phase

In a recent meta-analysis both ventral and dorsal striatum showed significant activity for loss during the outcome received (Liu et al., 2011). Our results also showed significant overlap in the regions activated in the meta-analysis. Hence, both gain and loss outcome activate basal ganglia and the midbrain (see **Figure 6.5**). Activation in the NAcc has been reported to correlate with the salience of the stimulus presented (Zink et al., 2003) However, there is also evidence that NAcc responses positively correlate with aversive stimuli (Delgado et al., 2004; Jensen et al., 2007; Salamone et al., 2007; Levita et al., 2009). As we discussed earlier it is possible that activation in this regions is modulates by the reward valence.

Furthermore, neurobiological evidence is started to confirm that the underlying motivational processes in financial loss share strong similarities with physical pain with the activity most commonly seen in the insular cortex (Delgado et al., 2006; Knutson et al., 2007; Wrase et al., 2007; Seymour et al., 2007). In our study we also found insular cortex activity when the punished outcomes were subtracted from the neural outcomes. For example Knutson et al., (2007) and Pessiglione et al., (2006) showed that financial loss activates insular cortex, where the activity in this regions previously shown to be correlated with expected pain (Seymour et al., 2004). On the other hand, the activity in the insular cortex can be interpreted as a response inhibition failure in our study because participants were trying the avoid from losses and they could have thought that they received negative feedback due to an inability of choosing the aversive option. Previous studies, which showed activity in bilateral insula, suggested that it plays a role in processing the significance of inhibitory failure (Preuschoff et al., 2006). Finally we found that amygdala selectively responsed to loss outcomes that are punished (-£1) compared to rewarded outcomes (+£1). Previous studies showed that amygdala is involved in monetary losses (Yacubian et al., 2006) and it is possible that this region might control the fight or flee response in a gambling task. Also it is important to keep in mind that punishment trials in Figure 6.6, Figure 6.7 and Figure 6.8 are rare events compromising roughly 20% of all trials where as reward events roughly compromise 80% percent of the events. Therefore the influence of punishment trials on learning may be greater than rewarding outcomes.

6.2 Analysis for Testing the Opponent Process Theory

6.2.1 Introduction

Based on opponent processes theory explained in **Chapter 2,** Kim et al., (2006) suggested that successful avoidance of an aversive outcome acts like a rewarding outcome and activates similar brain regions as financial gains in the medial orbito-frontal cortex. Additional evidence also supports this hypothesis, which showed that medial orbito-frontal cortex is involved in termination of painful events (Seymour et al., 2005). Based on this evidence we hypothesized that if avoiding an aversive outcome is itself rewarding, missing a rewarding outcome might be equally punishing. Amsel (1958, 1992) in his frustration theory argued that omission of an expected reward is a form of abstract punishment. Neural correlates of frustration due to missing of rewarding outcomes have been shown to increase activity in the insular cortex (Abler et al., 2005), as well as medial frontal cortex (Siegrist et al., 2005) and lateral prefrontal cortex (Ng and Blair, 2011). In this section we looked at the difference in the neural correlates of reward receipt and successful avoidance of punishment and compared the activity for the punishment receipt contrast with reward missed.

6.2.2 Regions Involving Receipt of Reward and Loss Omission

The experimental design allowed us to look at brain regions involved in loss omissions and reward outcomes. Loss omission is the contrast in which the participants get a neutral outcome in loss trials (con 5), and receipt of reward outcome is the contrast in which participants get a 1£ outcome in the gain trials (con 2). We first looked at the regions involved in reward receipt but not omission of losses (con 2 > con 5). Direct comparison of these contrasts revealed activity in the right medial frontal gyrus (T=6.06 x=30, y=5, z=52) a region previously shown to be involved in coding reward gains (Koch et al., 2008) and left sub-gyral (T=4.46 x=-18, y= 26, z=46) with $p < 0.001$ uncorrected (no other areas showed significant activity at this *p*-value). We also looked at the opposite contrast, which showed significant activity for omission of losses but not for reward receipt (con 5 > con 2) where no brain area showed significant effect at the level of $p < 0.001$ (uncorrected). Furthermore, we performed conjunction analysis (con 5 & con 2) in order to look for the regions involved in both reward outcomes and omission of losses. Consistent with a previous study (Kim et al., 2006), we found activation in (x=-6, y=35, z=4) with peak in anterior cingulate cortex and medial frontal cortex activity at (x=6, y=44, z=-2). This region showed increased BOLD response not only to reward receipt but also omission of losses at $p < 0.001$ (uncorrected). For the same contrast additional activation was also found mainly in posterior cingulate gyrus (x=-3, y=-10, z=34), bilateral ventral striatum (right x=26, y=17, z=-5.5) and (Left x=-20, y=17, z=-5.5), bilateral orbito-frontal cortex (BA 47 right x=30, y=14, z=17) and (BA 47 left x=-36, y=17, z=20) midbrain (x=6, y=-28, z=5) and ventral precuneus with a peak voxel activity in (x=-3, y=49, z=46). Moreover, activity in medial frontal cortex increases for monetary gains just like it increases for omission of losses (see **Figure 6.9a** and **Figure 6.9b**). Thus at the group random effect level, we provide evidence that medial frontal cortex responds both to monetary gains and omission of losses.

Additionally, in order to depict the areas responding more to receipt of rewards and avoidance of losses but not for omitted rewards and actual losses, we looked at the contrast (con 2 + con 4 – con3 – con 5). We found activity in the right putamen (T=3.24 x=21, y=6, z=13), bilateral pulvinar (left right T=5.52 x=27, y=-28, z=1, T=5 right x=21, y=-31, z=-2), Brodmann area 6 (pre-motor cortex) (T=3.89 x=-3, y=-1, z=53) and left Brodmann area 11 in the ventro-lateral orbito frontal cortex (T=4.04 x=-24 y=47 z=-11) at a more liberal threshold of $p< 0.005$ (uncorrected) (see Figure 6.10).

6.2.3 Regions Involving Receipt of Loss and Omission of Gain Outcome

We also tested for areas showing significant effects for receipt of loss outcomes (con 6) and omission of gain outcomes (con 3). At the group random effects level we first looked at the regions involved only in receipt of loss outcomes but not omission of gain outcomes (con 6 > con 3) and vice a versa (con 3 > con 6). Neither of these contrasts showedsignificant activity at the significance level of $p< 0.001$ (uncorrected). The results of conjunction analysis depict the regions involved both in receipt of loss outcomes and omission of gain outcomes (con 3 & con 6). Significant activity was found in the medial frontal cortex (BA10, x=6, y=50, z=-2), bilateral ventral striatum (right x=18, y=5, z=-5, left x=-15, y=5, z=-6.3) midbrain (x=-6, y=-28, z=-50), posterior cingulated cortex (x=-6, y=-19, z=46) and precuneus (x=-6, y=-52, z=-40) (see **Figure 6.9b**).

In order to depict the areas responding more to aversive loss outcomes and frustrative neutral outcomes compared to rewarded gain and loss omission trials we looked at the contrast (con 3 + con 6 – con 2 – con 5). Only two regions showed significant activity with a more liberal statistical threshold at $p< 0.005$ (uncorrected), in the left insula (T=3.26x=-42, y=29, =z=10) and the brainstem (T=4.33 x=12, y=-31, z=-20) (see **Figure 6.10**). This indicates that left insula is specific to negative outcomes. No other brain regions showed significant activity at $p< 0.005$ (uncorrected).

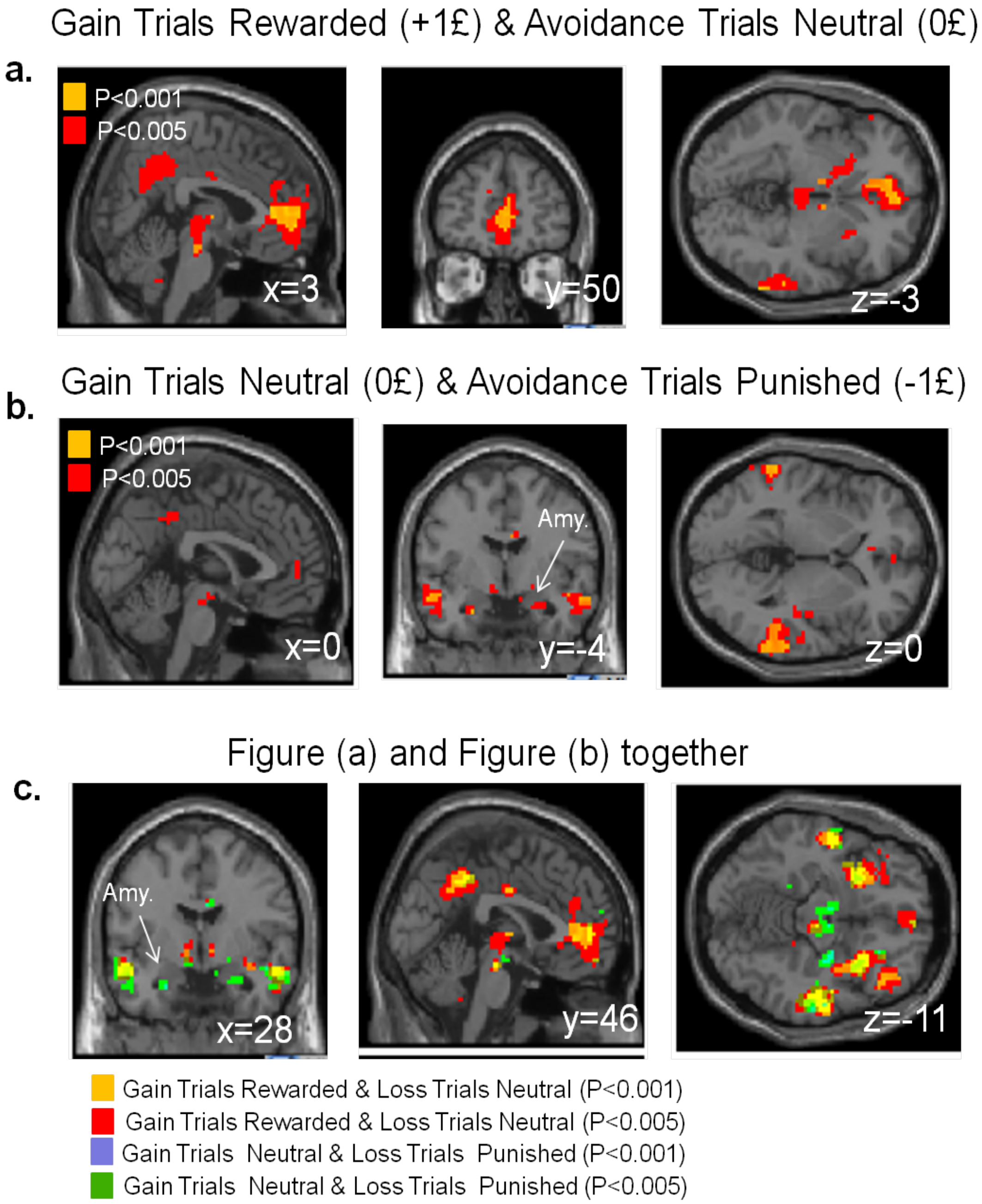

Gain Trials Rewarded (+1£) & Avoidance Trials Neutral (0£)
a.
🟧 P<0.001
🟥 P<0.005
x=3
y=50
z=-3
Gain Trials Neutral (0£) & Avoidance Trials Punished (-1£)
b.
🟧 P<0.001
🟥 P<0.005
x=0
Amy.
y=-4
z=0
Figure (a) and Figure (b) together
c.
Amy.
x=28
y=46
z=-11
🟧 Gain Trials Rewarded & Loss Trials Neutral (P<0.001)
🟥 Gain Trials Rewarded & Loss Trials Neutral (P<0.005)
🟦 Gain Trials Neutral & Loss Trials Punished (P<0.001)
🟩 Gain Trials Neutral & Loss Trials Punished (P<0.005)

Figure 6.9 Areas of whole brain showing significant activity during the outcome period for the probabilistic learning task. a) Group random effects results are shown superimposed on coronal, sagital and axial slices from on the single subject T1 weighted images (at the MNI coordinate indicted in the top right corner of image) for the conjunction contrast for gain outcome received and loss omission. Significant effects are shown at p < 0.001 in orange and p<0.01 in red (to show the full extent of activation). b) Group random effects results are shown for the conjunction contrast for gain omission and loss outcome punished. c) A plot of effect sizes in medial frontal cortex for the peak voxel was shown for gain outcome received (red) and loss omission (blue).

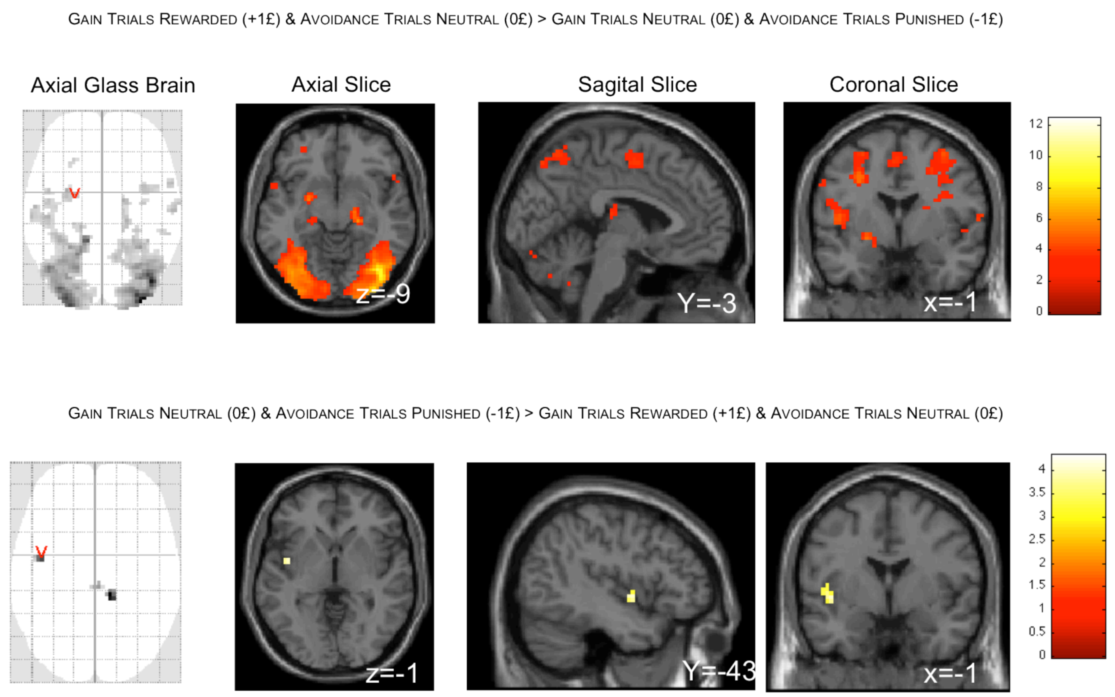

Gain Trials Rewarded (+1£) & Avoidance Trials Neutral (0£) > Gain Trials Neutral (0£) & Avoidance Trials Punished (-1£)
Axial Glass Brain
Axial Slice
Sagital Slice
Coronal Slice
z=-9
Y=-3
x=-1
12
10
8
6
4
2
0
Gain Trials Neutral (0£) & Avoidance Trials Punished (-1£) > Gain Trials Rewarded (+1£) & Avoidance Trials Neutral (0£)
z=-1
Y=-43
x=-1
4
3.5
3
2.5
2
1.5
1
0.5
0

Figure 6.10 a)Areas activated in conjunction of gain trials rewarded (+1£) + Avoidance trials neutral outcome (0£) compared to conjunction of gain trials neutral (0£)+ avoidance trials punished (-1£). Group random effects results are shown superimposed on axial, sagittal and coronal slices at the MNI coordinates indicated (below right corner of each image). Significant effects are shown at p<0.001 (uncorrected for multiple comparison) which are lateral OFC; occipital lope, pre-SMA. b) Areas activated in conjunction of gain trials rewarded (+1£) + Avoidance trials neutral outcome (0£) compared to conjunction of gain trials neutral (0£)+ avoidance trials punished (-1£).

6.2.4 Discussion

The present study provides evidence for overlapping opponent responses in medial OFC. These results suggest that during the outcome, the areas responding to reward receipt and avoidance of losses show overlapping activations with receipt of punishments and missed reward outcomes in the medial OFC. Previous studies showed that lesions to ventromedial frontal cortex disrupt reversal learning that requires a shift in behaviour in response to unexpected negative feedback as well as disrupting learning from negative feedback in a probability learning task that is similar to ours (Bechera et al., 1997; Fellows and Farah, 2005; Wheeler and Fellows, 2008). However, surprisingly we were unable to identify a difference between the ventro-medial-frontal region for reward outcomes and ventro-lateral frontal region for punishing outcomes as reviewed in **Chapter 2**. It is possible that this ventro-medial OFC is involved in evaluating feedback regardless of the type of feedback. In addition to that, we found a distinguishing activity in the amygdala for the trials, where participants received neutral outcome (0£) in gain trials and punishing outcome in avoidance trials (-1£). Previous studies also showed amygdala activity when they compared the neural correlates of monetary losses withneutral outcomes (Yacubian et al., 2006). Hence, the amygdala is well recognized to involve aversive learning. A recent gambling study involving mixed gains and losses of money, at differing amounts and probabilities, identified loss prediction errors in the amygdale (Yacubian et al., 2006) reported similar findings. Although it is difficult to place too much emphasis on the respective findings of the Yacubian et al., (2006) andour study, it is noteworthy that in both studies the amydgala showed greater response for aversive outcomes.

6.3 Model Based Analysis for Prediction Error and Expected Value

6.3.1 Introduction

It is well known that both animals and humans are capable of distinguishing conditional stimuli based on their positive and negativeoutcomes, but what is not well known is how the brain encodes, represents, and uses signals that indicate potential rewards and punishments. This was a challenging question for neuroscientists over several decades and discussed in more detail in **Chapter 2**. As a summary prediction error for financial gain and primary rewards has been found in striatum (please refer to **Chapter 2**). Similarly the prediction error for financial loss and aversive conditioning of painful shocks activates parts of the ventral striatum (Delgado et al., 2000; 2011). However, it is unclear, whether other parts of the brain that are involved in learning can also be involvedin processing loss prediction error necessary for aversive learning. Therefore in this section we decided to investigate the neural correlates of reward and loss prediction-errors. Moreover we decided to look at the CS related activations that are neural correlates of the expected value signal, which is calculated by the following equation:

Expected Value = (Predicted Value of CS_x *Probability of CS_x)+ (Predicted Value of CS_y * Probability of CS_y)

In the above equation HP and LP refers to high probability and low probability rewarding options respectively. Finally. In order to determine whether certain brain regions calculate specific signals for the reward and loss expected value and prediction error, we analysed our data in a model–based functional MRI (fMRI) fashion (O'Doherty et al., 2007, Corroda & Doya, 2007; Corrado et al., 2009, Mars et al., 2010). We used a reinforcement-learning model to estimate each subject's predicted-value of choosing an option and then calculated the expected value and prediction error profiles during learning and these two internal representations (hidden variables) were used in the fMRI correlation analysis.

6.3.2 Reinforcement Learning Model

The amount of information that an individual obtained from choosing a particular option (right or left symbol) for the gain trials were estimated by a simple Rescorla-Wagner reinforcement learningmodel (Rescorla& Wagner, 1972). This model keeps the estimate predited values for both types of options (high probability winning option and low probability winning option). The predicted-values are updated (only the chosen action value is updated) when the outcome turns up after execution of an action, based on the difference between the outcome and the estimated value of choosing that action. We used the Softmax action selection rule for updating the probability of selecting the options (high probability, HP or Low Probability; LP). For example if the participant chose the high probability-option the probability of choosing that option is calculated by the following equation.

$$p(hp) = \frac{e^{\beta Q(hp)}}{e^{\beta Q(hp)} + e^{\beta Q(lp)}} = \frac{1}{1 + e^{-\beta(Q(hp) - Q(lp))}}$$

β is the inverse temperature, which inversely relates to the randomness in action selection. For example, high β means higher probability of random action selection (β> 0). The prediction error was calculated by the difference between the actual reward received minus the value of choosing left or right symbol. We set the value of the reward *r* to 1 for positive feedback, and -1 for neutral feedback. The action values Q (left, right) were also set to 0 at the beginning of each learning session. When the outcome for the particular symbol was presented, the value of choosing that symbol was updated by the following equation

$$Q(hp) = Q(hp) + \alpha_t \delta_t$$

To determine the parameters with which the model best fit with the behavioural data of participants' actual choices, we calculated the likelihood function *l*(Q|z)for each set of parameters (Q= α,β) with participants actual choices (z). The model fitting procedure is as follows: we first calculate the action values with using all possible combinations of parameter values (incremental search). Then we estimate the probabilities for all possible parameter values for each trial. Then from the probabilities that a participant can select the symbol *a* in trial *i* was inserted in the likelihood function. The following equation shows the likelihood function, which is the product of the probabilities in all trials, included in the parameter space, z.

$$l(Q \mid z) = \prod_t p(a, t \mid Q)$$

The Matlab algorithm can be found from the **Appendix B.**

6.3.3 Behavioural Results

During model fitting we estimated each individual's learning rate, α and exploration parameter, β. When we performed a group statistical analysis for the differences in learning rate and exploration parameter, we found that there are no differences in learning rate between gain and loss condition ($p>0.05$, two tailed), but there is significant difference in the amount of exploration of the other option (see **Figure6.11**). Based on the higher exploration parameter for loss condition we can conclude that participants explore more when they were faced with the option that indicate potential losses ($t_{(11)}=4.3$ $p<0.05$, two tailed).

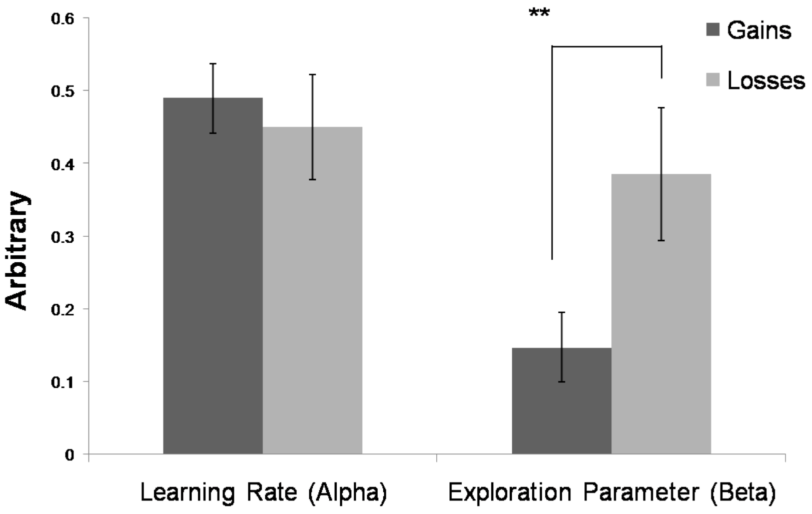


Figure 6.11 Differences of a group of participants for the parameter estimates for gain and loss condition. The figure shows that there were no differences in the learning rate of participants between the gain and the loss condition, but there is significant difference in the amount of exploration they perform.

After estimating each participant's learning rate and exploration parameter we inserted them into the reinforcement-learning model that was summarized above and calculated the prediction-error and predicted value of choosing a particular option. Also in order to validate how well the reinforcement-learning model fitted with actual choices of participants we looked at the model estimated probabilities of the selected options and actual choices of the participants as can be seen from **Figure 6.12.**

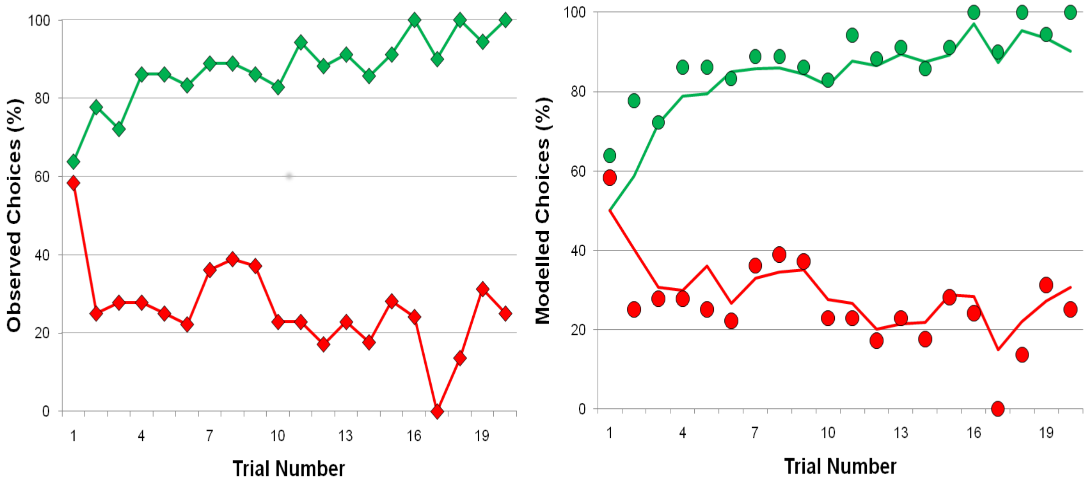


Figure 6.12 Behavioural model fitting results. Left: observed behavioural choices for reward trials (green) and avoidance trials (red). The learning curves depict, trial by trial, the proportion of participants that chose the high-probability option (symbol associated with a probability of 0.8 of winning £1) for the reward trials (green circles), and the high-probability option (symbol associated with a probability of 0.8 of losing £1) in the avoidance trials (red circles). Right figure: modelled behavioural choices for gain and loss condition. The learning curves represent the probabilities predicted by the computational model. Circles representing observed choices have been left for the purpose of comparison.

6.3.4 fMRI Results

As a part of our model based hypothesis, we tested regions correlating with the la-Wagner prediction error (PE) and expected value signals derived from our model.We inserted the estimated expected-value and prediction error in each trial and entered those values in to general linear model with respect to the time of the stimulus presentationand at the time of outcome presentation. Then each individual's results were carried to second-level group analysis. We wanted to see the difference between gain and loss prediction error during the outcome, and expected value activity during the cue presentation. The results revealed that expected value for reward outcomes are coded mainly in the nucleus accumbens (right x=12, y=11, z=-8; left x=-6, y=8, z=-5) (p<0.05

small volume corrected) and posterior putamen (right x=27, y=-7, z=1; left x=27, y=-10, z=-5)(p<0.05 small volume corrected)(see **Figure 6.13**).

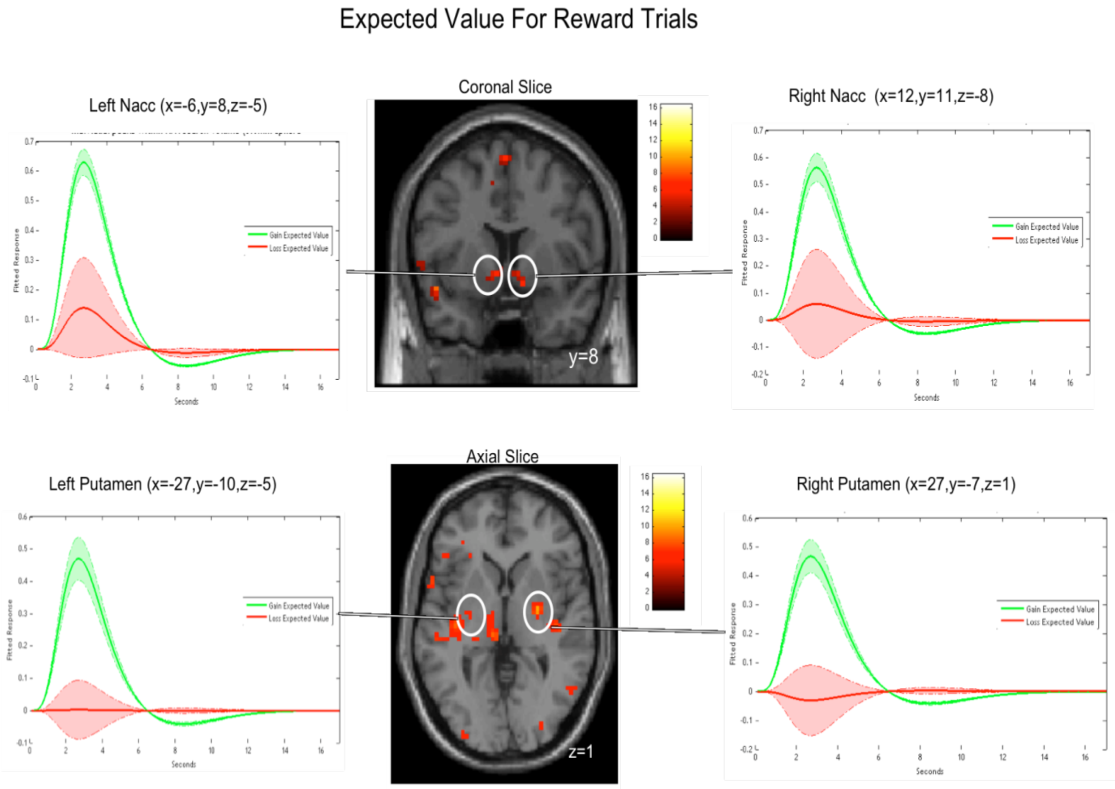

Expected Value For Reward Trials
Left Nacc (x=-6,y=8,z=-5)
Coronal Slice
Right Nacc (x=12,y=11,z=-8)
y=8
Left Putamen (x=-27,y=-10,z=-5)
Axial Slice
Right Putamen (x=27,y=-7,z=1)
z=1
Gain Expected Value
Loss Expected Value
Fitted Response
Seconds

Figure 6.13 Activation for the expected value signal. Upper middle figure shows the activation in the NAcc from a coronal view. Lower middle figure shows the activation in the NAcc from an axial view. Functional images are thresholded to a statistical significance level of p<0.005 (uncorrected). Group level BOLD activation overlaid on to a single subject T1 weighted anatomical image. The event related responses extracted from 5mm ROIs as represented with white circles (significance, p<0.05 small volume corrected). On the sides expected value for reward (green) and expected value for loss trials (red) seen as event-related responses calculated from the activations in ROIs (significance, p<0.05 small volume corrected).

Consistent with previous literature, we found significant PE activity in the ventral striatum (x= 6, y = 11, z = 8). Another important region showed significant correlation with reward prediction error signal include left medial frontal gyrus (x=-12, y =35, z=40).

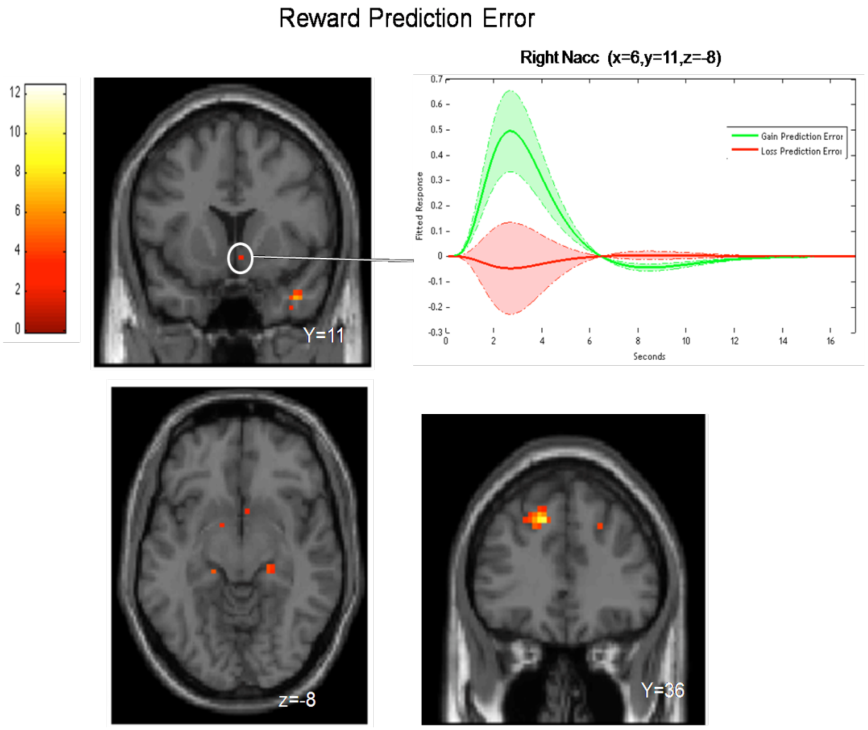
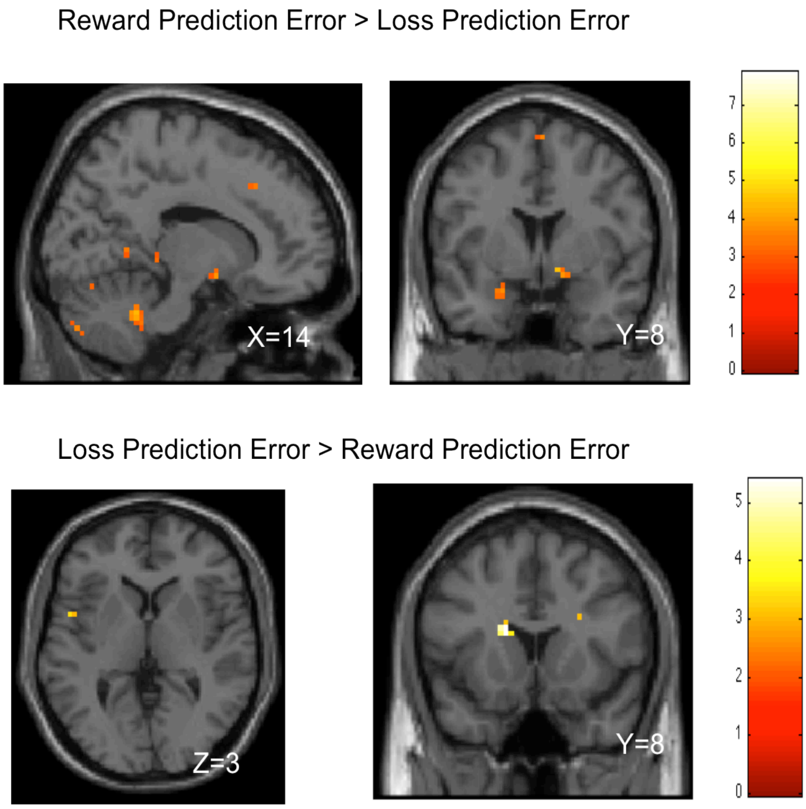


Figure 6.14 Brain areas activated by reward prediction errors. a) Upper figure axial slices show areas of Nacc activation by the main effect reward prediction error (P < 0.005, uncorrected for multiple comparisons). Most significantly active areas are bilateral ventral striatum (lower left figure) and middle frontal gyrus in the prefrontal cortex (lower right and left figure) where both show significant activity at the level of p<0.05 small volume corrected, 5 mm ROI. Upper figure on the right shows event-related responses for reward prediction error (green) and loss prediction error (red line) shows that Nacc shows a greater response for reward-prediction error.

We also tested prediction error signals for the loss trials. We looked for all areas that show a loss prediction error signal, that is, increasing activity when a loss outcome was received when unexpected, and decreasing activity when a loss outcome was not received when expected. Unfortunately we couldn't identify any significant brain region for this loss prediction error (at p<0.005, uncorrected) and decided to perform a subtraction analysis between loss and reward prediction errors. The results of the subtraction analysis showed greater activation for loss prediction error in left caudate (T=5.4 x =-15, y=14, z=16 small volume corrected with 5mm ROI p<0.05) and left insula (T=3.4 x=-46, y=5, z=0.7 p<0.01, uncorrected), whereas the opposite contrast that is the regions which showed higher activity for reward prediction error than loss prediction error showed significantly greater activity in the right ventral striatum (T=4.36, x=12, y=2, z=-8 small volume corrected with 5mm ROI p<0.05), left amygdala (T=4.28 x=-23, y=1, z=-20 small volume corrected with 5mm ROI p<0.05).


Reward Prediction Error > Loss Prediction Error

Loss Prediction Error > Reward Prediction Error

Figure 6.15 Brain areas activated by reward and loss prediction errors. a) Upper figure axial and coronal slices show those areas activated by the main effect reward prediction error that is greater than the loss prediction error. The statistical significance threshold was set to p<0.01 uncorrected and small volume correction applied to test further region of interest. Significantly active areas for reward PE>loss PE contrast is ventral striatum and the amygdala. b) Lower figure shows those areas respond more strongly to loss-prediction error than reward prediction-error. Areas that show significant activity for that contrast are left insular cortex (axial slice) and caudate nucleus (sagittal slice) (p< 0.01, uncorrected) but only caudate nucleus is significant at p<0.05 (small volume corrected).

6.3.5 Discussion

We reported activity in areas in which we had identified in our experimental hypotheses, which had been based on the review of all relevant literature. These regions have established roles in both aversive and appetitive predictive learning and involve sub-compartments of basal ganglia. We found evidence that inside the basal ganglia different sub-compartments might be differentially active for gain and loss related learning signals.

Prediction errors for gain trials generated by the model correlated with activity in the nucleus accumbens. This reward prediction error activity in the nucleus accumbens also confirms those reported in previous studies given that this the brain region which receives a large number of axonal projections from the midbrain DA neurons is very likely to be involved in prediction error calculation (Schultz et al., 1997; Waelti et al., 2001; Daw et al., 2002; Holroyd and Coles, 2002; O'Doherty et al., 2003; Schultz, 2004; Seymour et al., 2004; Rodriguez et al., 2006; Abler et al., 2006). However, we couldn't identify any significant activity for loss prediction errors. The results of the subtraction contrast revealed that the loss prediction errors influence caudate nucleus activity more than reward prediction errors.

6.3.6 Interim Summary

The research questions that Experiment 1 answered in this chapter include:

1. Is there a similar organization within the brain for processing the *anticipation of gain and loss* in reinforcement learning?
2. Is there a similar organization within the brain for processing the *outcomes of gains and losses* in reinforcement learning?
3. Which parts of the brain are involved in the computation of prediction error and expected value for gains and losses?

For the first question our results indicate that reward and loss related anticipation activate a similar fronto-striatal network including mainly striatum, cingulate cortex, medial and dorsa-lateral prefrontal cortex. We found an additional activation in the antero-medial cingulate cortex during anticipation of loss outcome but this activity is absent for reward expectation.

In general we also found that reward expectation generates a greater extent of activity than losses. This difference can be seen in caudate and ventro-lateral prefrontal cortex (see **Figure 6.5**). When we looked at the contrasts of reward expectation > neutral expectation and loss expectation > neutral expectation both contrasts showed ventro-lateral OFC activity but medial OFC activity was only found for the reward expectation > neutral expectation contrast. This particular result in fact contradicts previously reported findings summarized in **Chapter 2**, which suggest a distinction for reward (ventro-medial OFC) and punishment processing (ventro-lateral-OFC) in the ventral orbito-frontal cortex. The findings of the current study can be interpreted as a ventro-lateral OFC system involved in both reward and punishment expectation but medial OFC involved only in reward expectation. It is also possible that the ventro-medial reward and ventro-lateral punishment distinction might be more robust for the feedback related activity rather that expectation related activity which might be related to goal-values of outcomes rather that their predicted values.

Furthermore, we found differences in brain activity between gains and losses during the outcome period. During the outcome period receipt of monetary losses and gains both activate ventral striatum but additionally receipt of losses activates bilateral amygdale and insular cortex as revealed by the subtraction analysis. In general it is hard to distinguish the activity for motivational saliency from action initiation, because the CS presentation is also correlated with action initiation and might involve regions involving motor preparation.

Further model based analysis on gains and losses revealed that reward prediction error is coded in the ventral striatum and loss prediction error is coded in caudate nucleus. These results confirmed the previous studies that showed reward prediction error in the ventral striatum but in addition to that we showed the involvement of caudate nucleus for loss prediction error. Overall the results are compatible with the electrophysiological findings (Haber et al., 2000) and theoretical predictions (Haruno &Kawato, 2006).

Chapter 7

7.1. Analysis for Early versus Late Learning Trials for Prediction Error and Predicted-values

7.1.1 Introduction

In Chapter 3 a review was provided of human brain imaging studies and non-human primate singleunit-recording studies which have shown that the basal ganglia is crucial for reward and motor processing (fora review see, Packard & Knowlton, 2002; Montague, King-Casas, Cohen, 2006; Yin & Knowlton, 2006; Graybiel, 2008; Rangel, Camerer, Montague, 2008; Doyon et al., 2009). Over the years, one of the most important findings is that some regions in the striatum which are thought to carry motor signals are highly influenced by the probability of upcoming rewards, in other words those group of neurons are both responding to actions (motor movements) and reward expectations (Apicella et al., 1991; Kawagoe et al., 1998; Tremblay et al., 1998; Lauwereyns et al., 2002; Takikawa et al.,2002; Miyachi et al., 2002; Pasupathy and Miller, 2005; Watanabe and Hikosaka, 2005; Hollermanet al., 1998; Posquire et al., 2007; Hori et al., 2009). During the course of learning many of these neurons adapt theirfiring rate for the motor actions most likely to result in a reward. We can refer to these as the action with highest predicted value (for a review see, Schultz, 2003).

In Chapter 3 we also reviewed studies, which showed that during instrumental learning of outcomes medial orbito-frontal cortex is involved in coding goal-directed values of instrumental actions whereas dorsal striatum is involved in coding the well-learned action values in the form of habits.

Based on that evidence we hypothesized that during the early learning trials the predicted values of chosen options will be coded by a goal directed system and should activate frontal cortex but the during the late trials the predicted values should be coded by a habit system and activate dorsal striatum. In order to prove that hypothesis, we re-analysed the data that was presented in **Chapter 6**. According to this new analysis, we separated the behavioural data in to two parts as the early learning trials (first ten trials) and late learning trials (last ten trials) and looked for the neural correlates of prediction-errors and predicted-values of selected options.

7.1.2 Materials and Methods

The following analysis is based on thefMRI data of the 12 participants that was presented in **Chapter 6**, Therefore the materials and methods are exactly the same as the ones that were used in the previous chapter, therefore for details please refer to **Chapter 6**.

7.1.3 RESULTS

7.1.3.1 Behavioural Analysis

In order to understand whether there is a performance difference between the early and late trials we looked at the differences in response times for selecting an option for the reward, avoidance and neutral conditions. Over the course of the experiment participants showed significant reduction in the response times for all three conditions when the average response times for all three sessions (60 in total) were compared. The comparison between the response time for early and late learning trials revealed a statistically significant difference between all conditions (reward condition $t_{(58)}=3.105$, $p<0.01$, two tailed, avoidance condition $t_{(58)}=4.35$, $p<0.01$, two tailed, neutral condition $t_{(58)}= 5.503$, $p<0.01$, two tailed). These results indicate that during learning a participant's response becomes quicker which is an indication of shift to habit or rather automaticity in action selection (**Figure 7.1**).

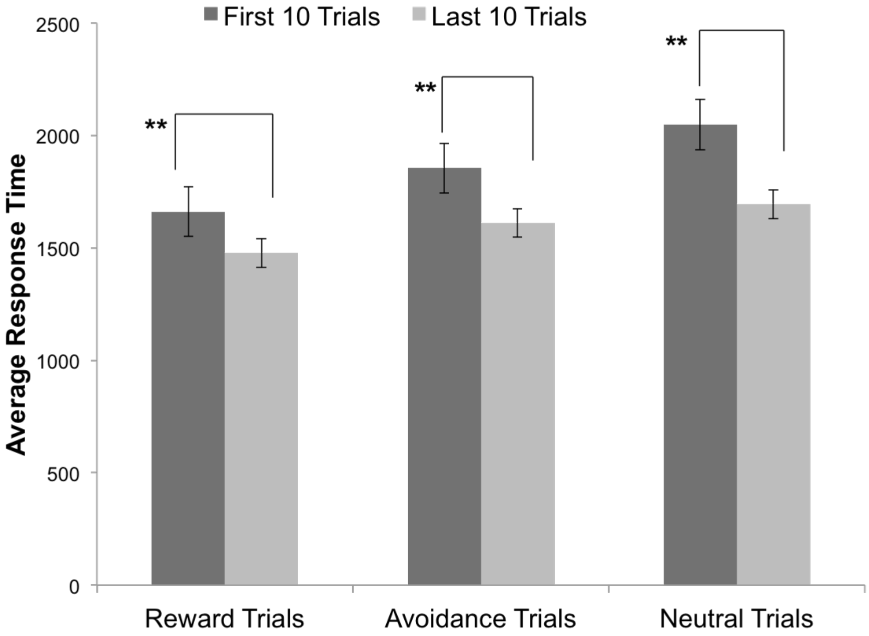


Figure 7.1 Plot of the reaction times for the three conditions regardless of the outcome received. Participants were significantly slower in the early trials than later trials for all conditions. Bars represent standard errors. (**) Represents significance ($p<0.05$, two tailed). The data represented above belongs to the average of 12 participants.

7.1.3.2 *Reinforcement-Learning Model*

The model used in this chapter is the same as the one used in the previous chapter where the predicted values are updated only for the symbol that is chosen. The updating of the predicted value of the chosen symbol is based on the prediction error. As described in **Chapter 6**, a softmax action selection rule was used for updating the probability of selected actions. The predicted-values or so called Q values, (high probability, hp and low probability, lp) were set to 0 at the beginning of each learning session. When the outcome for the particular symbol was presented, the value of the choosen option was updated by the following equation:

$$Q(hp) = Q(hp) + \alpha_t \delta_t$$

In the above equation alpha and delta refers to learning rate and prediction error respectively and defined earlier in **Chapter 6 section 6.3.** Given that we are using the same behavioural data and same computational model that was used in **Chapter 6**, we didn't repeat the model fitting procedure (see **Chapter 6 section 6.3**) and assumed the same parameter values for the learning rate (alpha) and exploration parameter (beta) but we just separated early versus late predicted values for reward and avoidance conditions. The statistical analysis comparing early and late predicted values showed a significant difference for the reward condition ($t_{(28)}$=2.9 $p<0.05$, two tailed) (see **Figure 7.2**). The average predicted value of the chosen option (Q-value chosen) in the late trials was significantly greater than the average in the early trials. However the early versus late predicted values for the loss trials were not significantly different from each other ($t_{(28)}$=-1.056 $p>0.05$, two tailed).

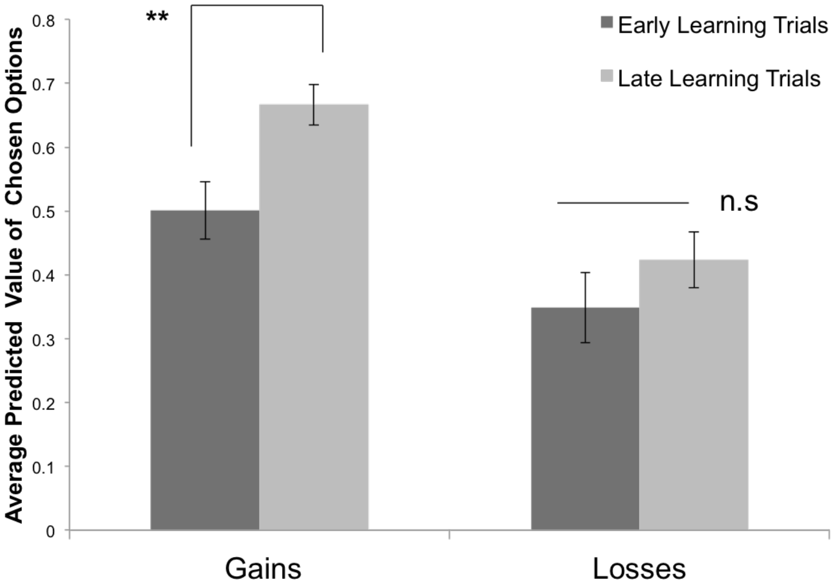


Figure 7.2 Average changes in the predicted value of chosen option for the early versus late trials for the reward and avoidance condition.

It is important to keep in mind that the reason for the statistical comparison of the early versus late predicted values are exploratory. Although it is possible that there might be a statistically significant difference between the early and late predicted values without any neural correlates, the opposite is plausible too. That is to say there might be no difference in the predicted values between the early versus late trials but there might be difference in their neural correlates. The former might be true for processing reward value and the latter true for processing loss.

7.1.3.3 fMRI Results

Using the model parameters described above, we took the trial by-trial predictions of our computational model for the predicted-values and entered these into a regression analysis against the fMRI data at the time of cue (CS) presentation. Early and late prediction values and prediction errors are entered in to the design matrix as separate contrasts.

In order to compare the difference in activity for early versus late contrasts separateROI (region of interest) analyses were performed for each anatomical sub-region. In the anatomical ROI analysis medial frontal cortex and several sub-compartments of the basal ganglia were used. The specific selection of those ROI regions was based on the previous studies, which showed significant change in BOLD signal for coding reward prediction errors and predicted values of reward outcomes (Haruno and Kawato, 2006). The ROI for the sub-regions of basal ganglia was taken from the BGHAT template (Prodoehl et al., 2008) and the ROI for the medial orbito frontal cortex is taken from the AAL atlas (Tzourio-Mazoyer et al., 2002). However, there were no previously defined ROI's in the MNI (Montreal neurological institute) space for nucleus accumbens (Nacc) therefore we have to define Nacc by drawing by hand using MRIcron (http://www.mccauslandcenter.sc.edu/ mricro/mricron). The hand drawn Nacc ROI is smoothed with a 3 mm Gaussian kernel and normalized to the Montreal neurological institute (MNI) template. There are 12 regions of interest in total (6 in each hemisphere) and each of these regions was tested separately for 8 contrasts, namely early reward predicted value, late reward predicted-value, early loss predicted value, late loss predicted value, early reward prediction error, late reward prediction error, early loss prediction error, late loss prediction error. This makes a total of 96 test altogether.

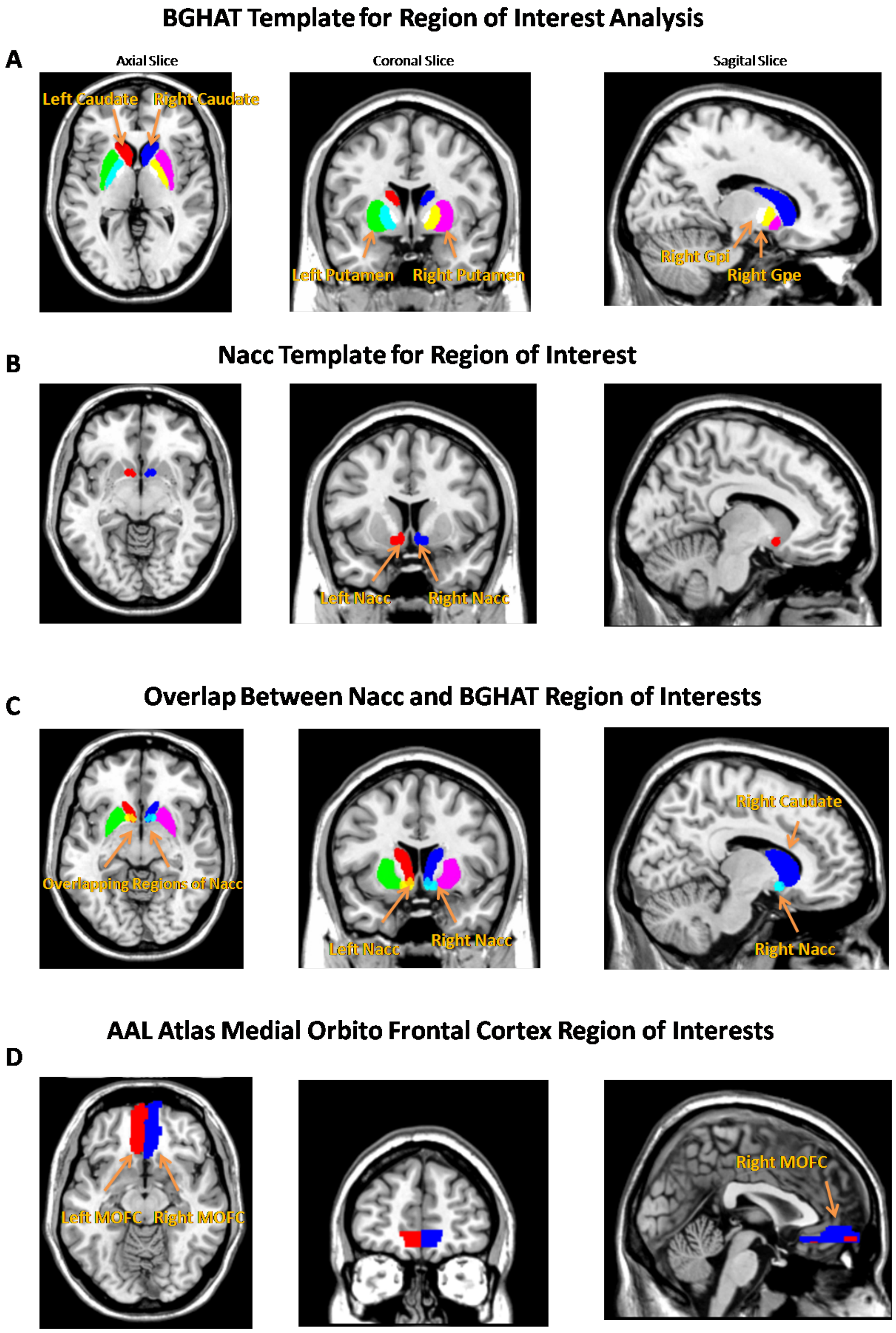

BGHAT Template for Region of Interest Analysis
A
Axial Slice
Coronal Slice
Sagital Slice
Left Caudate
Right Caudate
Left Putamen
Right Putamen
Right Gpi
Right Gpe
B
Nacc Template for Region of Interest
Left Nacc
Right Nacc
C
Overlap Between Nacc and BGHAT Region of Interests
Overlapping Regions of Nacc
Left Nacc
Right Nacc
Right Caudate
Right Nacc
D
AAL Atlas Medial Orbito Frontal Cortex Region of Interests
Left MOFC
Right MOFC
Right MOFC

Figure 7.3 ROI's used in the fMRI analysis. a) BGHAT ROI template for the basal ganglia sub regions. b) Hand drawn ROI for the Nacc. c) Overlapping regions between the ROIs for Nacc, caudate and putamen. d) Medial orbito-frontal cortex ROI based on AAL atlas.

The first contrasts that we looked at were the predicted-values of chosen options for reward contrast, where we tested each anatomical ROI (including left and right hemispheric regions) separately (see **Table 7.1** for the t-values). The results of the early reward predicted values showed significant positive BOLD changein the medial-orbito-frontal cortex (see **Figure 7.4a**) only and late reward predicted value showed significant change in the bilateral putamen only (see **Figure 7.4b**).

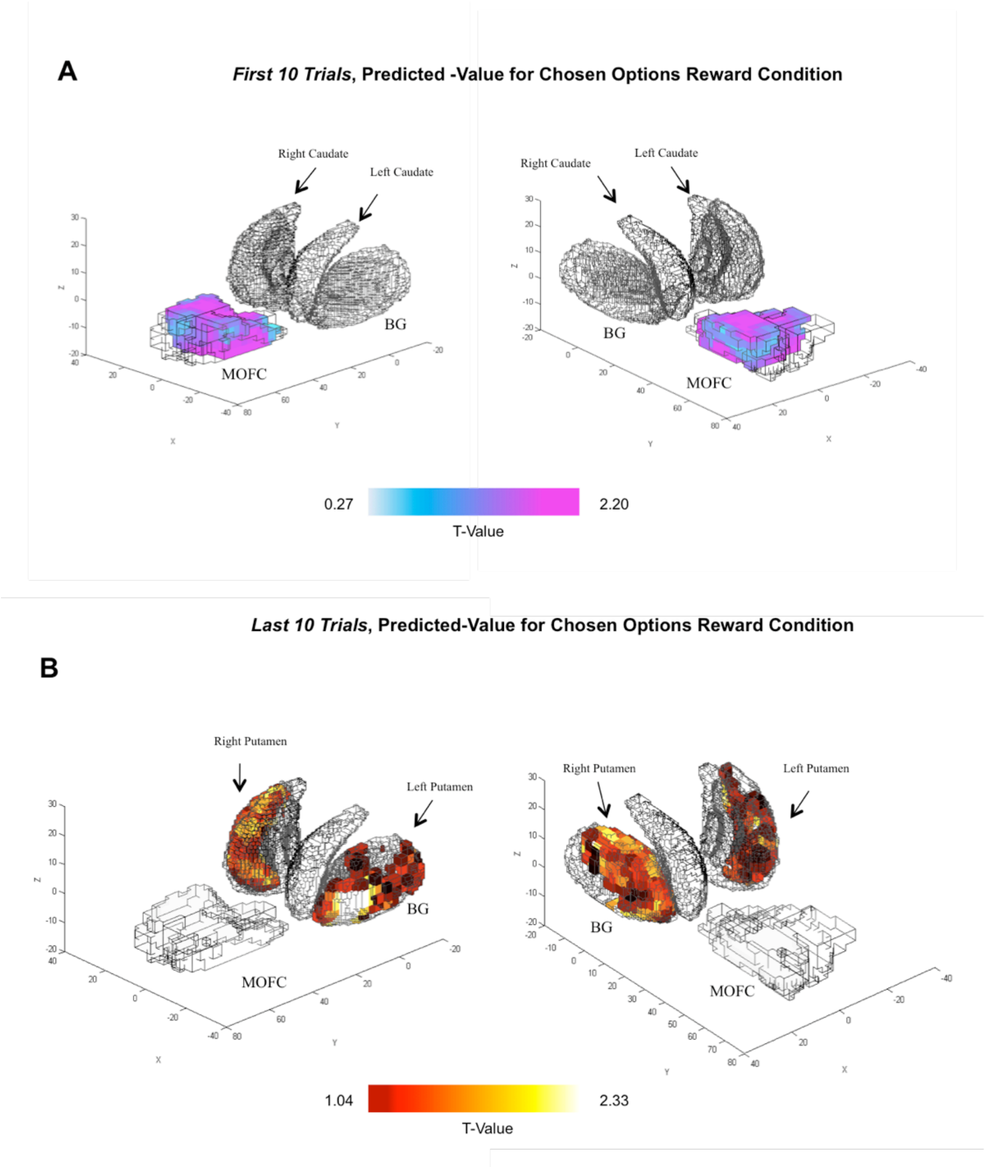

A
First 10 Trials, Predicted -Value for Chosen Options Reward Condition
Right Caudate
Left Caudate
BG
MOFC
Z
Y
X
0.27
2.20
T-Value
Last 10 Trials, Predicted-Value for Chosen Options Reward Condition
B
Right Putamen
Left Putamen
BG
MOFC
Z
Y
X
1.04
2.33
T-Value

Figure 7.4 Predicted-values of chosen options during early and late reward trials. a) Activity in the medial frontal cortex correlated with the reward predicted-value in the early learning trials. b) Activity in the right and left putamen is correlated with the reward predicted-value in the late learning trials. The gray mesh frame includes the medial prefrontal cortex ROI (AAL template),

the entire basal-ganglia (BGHAT template) and the Nacc ROI. The activations in each voxels have arbitrary dimensions based on multicolor software (*www.cns.atr.jp/multi_color_download*). Eachvoxel is associated with T-values that are represented by the brightness of the colors as shown in color bars.

We secondly looked at percent signal changes for the predicted value of rewarding stimuli (see **Figure 7.5**). We found that the medial-orbito-frontal cortex is sensitive to reward predicted values early in learning but putamen is sensitive to later in learning.

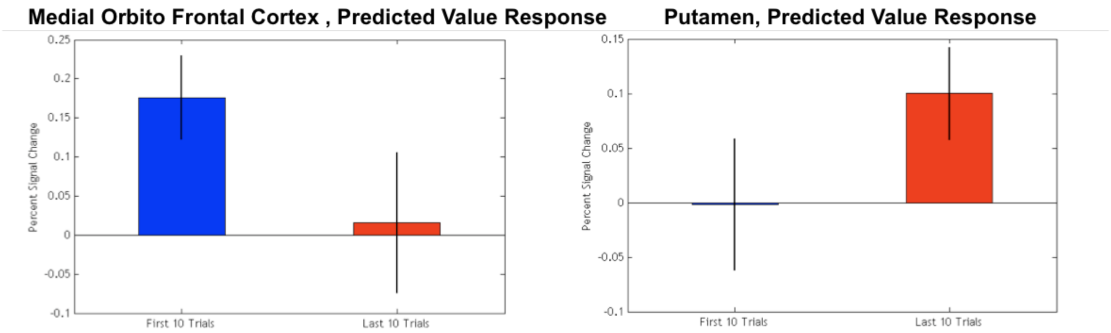


Figure 7.5 Percent signal changes for predicted-value in the medial-orbito-frontal cortex and putamen. Percent signal changes calculate using the whole ROI region.

Secondly we tested for statistically significant changes in signal in these ROIs for early and late loss predicted-values. None of the ROI's showed significant signal change for the early loss predicted-value (p<0.05, Uncorrected). For the late loss predicted-value we found significant signal change (p<0.05, Uncorrected) in the left globus pallidus internal segment (Gpi). The comparison of this region with the late reward predicted-value showed that this region was only sensitive to loss predicted values.

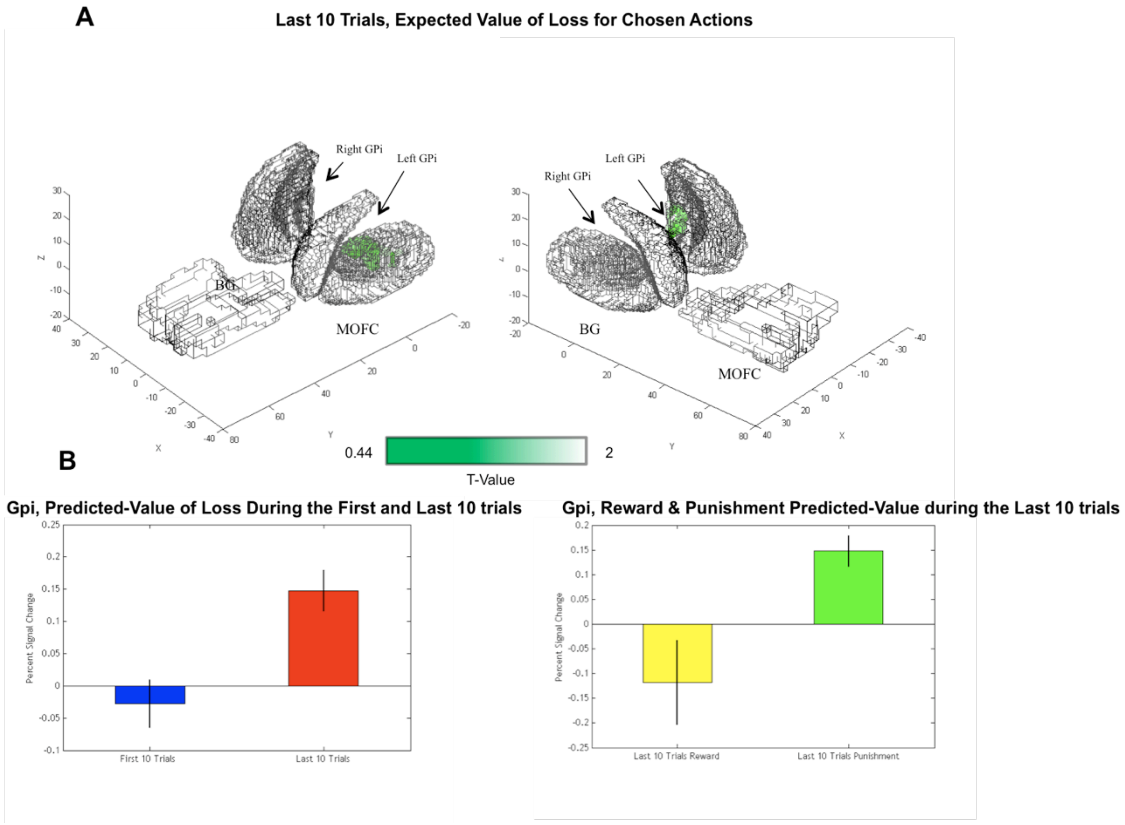


Figure 7.6 a) BOLD activation for loss predicted-value in the left globus pallidus internal segment. The activity in the ROI overlaid on the mesh frame (gray) that is created by the multicolor software (*www.cns.atr.jp/multi_color_download*). b) Percent signal change in the left Gpi for the first and the last ten trials. c) Percent signal change for the last ten trials of reward predicted-value (yellow bar) and loss (green) predicted value.

Finally we carried out the same analysis for loss and reward prediction errors. We found that both reward and loss prediction errors produce significant effects only during the first 10 trials and no significant effects were found in any of the ROIs' during late learning trials. For the reward prediction error during early trials, significant activity was found in the bilateral Nacc and for the loss prediction error we found significant activity in the bilateral caudate nucleus (see **Figure 7.7**). We also looked for the percent signal change in the NAcc for the loss prediction errors and vice versa for the caudate nucleus for reward prediction error in order to examine the possibility that these regions are specific for loss and reward prediction errors (see **Figure 7.7b** and **Figure 7.7d**). We found that the caudate showed negative BOLD signal for the reward prediction error and Nacc showed no percent signal change.

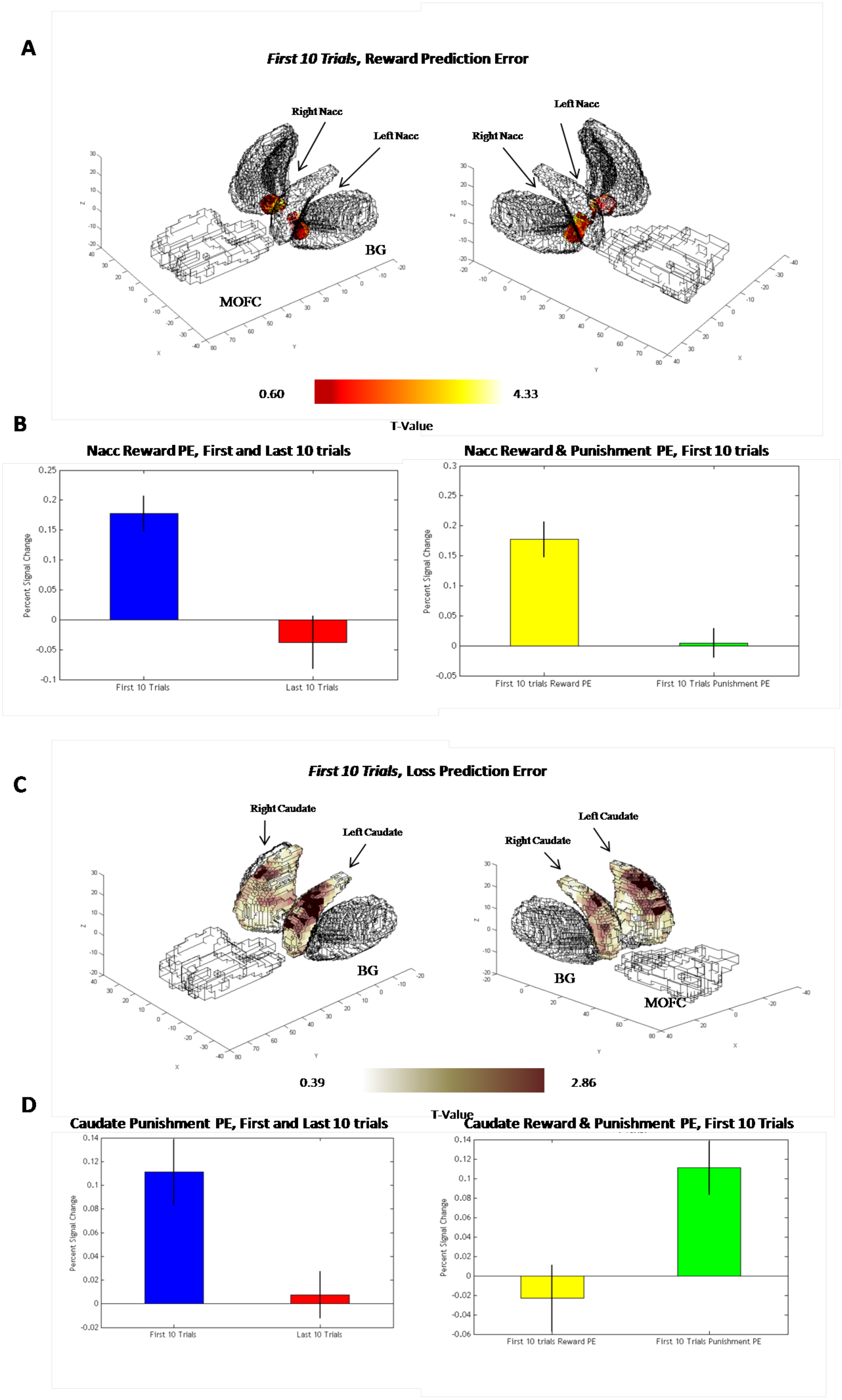

A
First 10 Trials, Reward Prediction Error
Right Nacc
Left Nacc
BG
MOFC
Left Nacc
Right Nacc
0.60
4.33
T-Value
B
Nacc Reward PE, First and Last 10 trials
Percent Signal Change
First 10 Trials
Last 10 Trials
Nacc Reward & Punishment PE, First 10 trials
First 10 trials Reward PE
First 10 Trials Punishment PE
C
First 10 Trials, Loss Prediction Error
Right Caudate
Left Caudate
BG
Left Caudate
Right Caudate
BG
MOFC
0.39
2.86
T-Value
D
Caudate Punishment PE, First and Last 10 trials
Percent Signal Change
First 10 Trials
Last 10 Trials
Caudate Reward & Punishment PE, First 10 Trials
First 10 trials Reward PE
First 10 Trials Punishment PE

Figure 7.7a) Activity in the bilateral Nacc for the reward prediction error during early learning. b) Percent signal change for the first and the last ten trials for the reward prediction error (figure on the left). Percent signal change for the reward and punishment prediction error for the first ten trials (figure on the right) c) Activity in the bilateral caudate nucleus for the loss prediction error during early learning. d) Percent signal change for the loss prediction error in the caudate nucleus for the first and last ten trials of learning.

Table 7.1 Results of the ROI analysis.

Early Gain Action Value			
	Laterality	t-statistic	Uncorrected P-Value
Frontal_Med_ORB_L	L	2.2	0.029
Frontal_Med_ORB_R	R	2.67	0.014

Late Gain Action Value			
		t-statistic	Uncorrected P-Value
Putamen	L	2.33	0.019
Putamen	R	1.85	0.045

Late Loss Action Value			
		t-statistic	Uncorrected P-Value
Gpi	L	2	0.035

Early Gain PE			
		t-statistic	Uncorrected P-Value
NAcc	L	2	0.024
NAcc	R	4.43	0.002

Early Loss PE			
		t-statistic	Uncorrected P-Value
Caudate	L	2.86	0.01
Caudate	R	2.78	0.011

Early Loss PE			
		t-statistic	Uncorrected P-Value
Caudate	R	2.33	0.02

7.2 Discussion

We were interested in identifying brain areas responding to changes in predicted-values and prediction errors for early versus late learning trials. We found that early in learning the reward predicted value correlates with activity in medial-orbito-frontal cortex but later in learning this activity shifts to putamen. On the other hand, we found left globus pallidus external segment activity for loss predicted-values in late learning trials only. In addition to that, we also replicated the well-established findings that showed prediction error signal in the ventral striatum for early learning trials only, whereas loss prediction error activated caudate nucleus and showed no prediction-error related activity for late learning trials.

Recent studies have suggested that there might be more than one type of value signal and predicted-value signals are only one of them (Rangel and Hare, 2010). It was suggested that the predicted-value signals are involved in the processes of evaluating the anticipated outcome (O'Doherty, 2011). Rangel and Hare (2010) argued that predicted values are anticipatory value signals, which reflect the anticipated outcome of each possible decision during selecting of an option in a decision making task. Also it is important to keep in mind that in instrumental conditioning task outcomes might be associated with the actions (see for a discussion, O'Doherty, 2011). Given that, we found medial orbito-frontal cortex activity for the early-predicted values and putamen for late predicted values this anatomical and functional dissociation between medial-orbito-frontal cortex and putamen also lends support to the hypothesis that these separate regions of human brain are involved in goal directed and habitual learning respectively (Balleine & O'Doherty, 2010). Correspondingly, it has been suggested that the sensori-motor striatum is important in chunking motor patterns in the form of habits (Graybiel, 1998) and the associative-striatum is important in goal directed learning and sensitive to outcomes (Balleine & O'Doherty, 2010).

Studies of cortico-striatal anatomy showed that the rostral striatum has connections to orbito frontal cortex (limbic loop) and sensori-motor striatum involving putamen has connections to motor and supplementary motor areas (motor loop) (Haber et al., 2000; Haber & Knutson, 2010) suggesting there might be distinct pathways. Therefore, it is plausible that early in learning medial-orbito-frontal cortex might be engaged in exploringthe response options and putamen is fine-tuning motor movements (Kim et al., 2009; Thorn et al., 2010; Stalnaker et al., 2010).

Finally, although this study specifically focused on the neural correlates of predicted values it is highly related to computational models of reinforcement learning and raises the question whether the predicted-values of options are coded with actions or not. In the reinforcement learning literature, several possible implementations exist for coding action values such as Q-learning or Actor-Critic architecture (see **Chapter 4**). There is a long ongoing debate suggesting that the basal ganglia learns the outcomes of actions in a manner similar to that described in the Actor-Critic model (for a review, Joel, Niv, Ruppin, 2002). According to this model, dorsal striatum, mainly the putamen and caudate nucleus, learn the action selection policy and are therefore only responsible for the selection of motor actions whereas the ventral striatum acts as the critic and calculates the prediction error signal and sends this signal to actor. As a working hypothesis, depending on the results of the current study we propose that dorsal striatum is not only coding the action policyin the form for probability of action selection, or policy but it is also coding related Q-values such as the value of selecting an option let's say *a*, when the agent is in state S: Q(s,a). Crites and Barto (1995) suggested a modified Actor-Critic architecture that is equivalent to the Q-learning algorithm that encodes the Q-values in the policy. This suggestion by Crites and Barto (1995) seems highly plausible given that the putamen is active for predicted-values of selected options in later trials. In the current study however, we didn't fit our model to the actor-critic architecture (Sutton & Barto, 1998) or modified actor-critic architecture (Crites and Barto, 1995) because we initially intended to investigate action values. Comparison of those models is necessary as a part future work in order to prove the working hypothesis.

Chapter 8

Experiment 2: Neural Correlates of Reinforcement Learning with Novel and Familiar Stimuli

8.1 Analysis for Familiar and Novel Trials During the Anticipation and Outcome Periods

8.1.1 Introduction

Most of the time human decisions are performed automatically with little or no attention paid to stimuli that indicates available options. This is because most of the time the choices we faced are based on familiar stimuli with known outcomes that are learned through trials and errors. However, in order to adapt to new environments where people are confronted with novel stimuli, they have to pay more attention to the novel stimuli in order to learn what the stimuli indicates and decide as quickly as possible in accordance with the novel situations demands. Research into the brain areas involved in novelty has been rich in quantity (Ranganath & Rainer; 2003). For example, recent neuroscientific studies have shown that the occurrence of novel events might trigger activity in brain regions that are involved in various cognitive processes including attention, learning and memory (Yin & Knowlton, 2006, Graybiel, 2008, Seger & Spiering, 2011).

During reinforcement learning the effect of novelty can be seen as changes in behavioural and neural responses to conditional stimuli (Duzel et al., 2010) where it implies that the perceptual properties of the stimuli are completely unknown to the subject (Tulving et al., 1996). Previous studies showed that neural responses for novel conditional stimuli can predict whether the event is occurring with known (familiar) or unknown (novel) stimuli. Recent studies showed that familiarization of the novel stimuli happens fast, usually in tens of trials (Kobayashi and Schultz, 2010). This process is in fact much faster than the acquisition of other novel events like motor-skills, or perceptual category judgement which was previously shown to take extensive training ranging from few hours to few months (see for a review, Yarrow et al., 2009; Helie & Cousineau, 2011). Animal studies showed that during learning reductions in associated neural activity occurs in the dorso-lateral frontal cortex (Assad et al., 1998) and subcortical brain regions particularly in the midbrain dopaminergic neurons and opposite pattern of increased activation was shown in these regions when the conditional stimuli are novel (Ljungberg et al., 1992; Redgrave, Gurney, Reynolds, 2007). Effects that are similar to animal electrophysiological studies also reported in human functional magnetic resonance imaging (fMRI) studies (Bunzeck and Duzel, 2006, Wittmann et al., 2007). Detecting novel task-sets have been shown to involve brain regions mainly hippocampus (Knight, 1996; Lisman & Grace, 2005; Kamuran & Maguire, 2009), amygdala (Breiter et al., 1996; Wilson & Rolls, 1993; Schwartz et al., 2003; Wright et al., 2006), midbrain (Krebs et al., 2011) and lateral-prefrontal cortical structures (Alexander et al., 1995; Knight and Scabini, 1998; Daffner et al., 2000; Kishiyama et al., 2009). More specifically the effect of novelty during reinforcement learning has been observed in several studies as increased activity in dorso-lateral prefrontal cortex (Turner et al., 2004; Duzel et al., 2004), ventral-striatum (Wittmann et al., 2008) and several studies have shownan increase in BOLD signal in midbrain dopaminergic activity to novel stimuli compared to familiar stimuli (Bunzeck and Duzel, 2006; Krebs et al., 2011) (please refer to Chapter 3 for a detailed discussion). This effect of dopaminergic neurons to novel stimuli is called novelty bonus (Kakade & Dayan, 2002) as reviewed in more detail in **Chapter 4**.

Another important factor that influences stimulus novelty is the predictive relationship between the stimulus and the outcome (Kagan, 2009). In reinforcement learning when familiarity with a stimulus requires familiarization in two distinct aspects of stimulus processing. It can perceptually familiar ie all of its perceptual properties are known, or what it signifies in terms of consequences (what it predicts) can be familiar familiar . If a stimulus is perceptually familiar but the outcome is unpredictable, then exposure to the stimulus will still elicit a novelty orienting response, with attentional resources being automatically diverted towards that stimulus (Corbetta & Shulman, 2002). Several researchers formalized this type of novelty-predictability relationship as expected and unexpected uncertainty (Dayan and Yu, 2003). According Yu and Dayan, (2005) expected uncertainty arises from known predictive relationships within a familiar environment, and unexpected uncertainty rises from unknown predictive relationships within a novel environment. For example, mis-forcasting of weather for a familiar region like one's owntown is considered to be as expected uncertainty but when a mis-forcast of weather occurs for a region that we are not familiar with then it is considered as unexpected uncertainty (e.g., unexpected occurrence of a weather storm in Florida for a person who is living in London). Several neurocomputational studies proposed that the neuromodulators acetylcholine and norepinephrine play amajor role in the brain's implementation of the expected and unexpected uncertainty computations (Dayan and Yu, 2003). There is also a wealth of evidence from electrophysiological and human functional imaging studies which suggest that striatal regions also carry additional computations for coding uncertainty in the context of learning novel stimulus response contingencies (Volz et al., 2003; Preuschoff et al., 2006; Bunzeck et al., 2010).

In this study we used functional magnetic resonance imaging to see which brain areas are involved in coding stimulus novelty during instrumental learning. To be consistent with Expt 1 and be able to compare results we adapted the event-related design of a probabilistic reinforcement-learning used Experiment 1 (see **Chapter 6**). In Experiment 2 we pre-trained all participants with a set of stimulus pairs over many trials. This was done outside the scanner and before the main experiment in the scanner. As in Experiment 1 participants learn via trial and error to select one of the stimuli in the pair in preference to the other since it has a higher probability of reward than the other stimuli. We shall refer to this pre-training as the familiarization tasksince participants saw and responded to some two hundred forty trials using the same set of stimulus pairs. From pilot work this results in overlearning, or a high level of familiarity, given that participants reach the asymptotic levels of the learning curves after some 30 or so trials.

In a separate session (main experiment) each of these subjects completes a similar task under the scanner , except that now the subject is not only presented with the set of stimuli which have become highly familiar (overlearnt) but they are also presented with a set of novel stimuli which they have never encountered previously. This design should be more powerful than Expt 1 in distinguishing between novel stimuli (eliciting goal directed behaviour) and familiar stimuli (eliciting automatic behaviour) In both the familiarization task and the main experiment we only include reward-learning trials with loss trials removed altogether.

There are at least two important reasons why we are conducting this study. Firstly, only a few fMRI studies have taken into consideration both the effects of stimulus novelty and the effect of outcome predictability when comparing novelty and familiarity (Kagan, 2009). The interaction between the neural mechanisms of these two types of novelty (stimulus novelty & outcome predictability) is still not well understood. For example, some studies used either stimuli with known perceptual properties but unknown stimulus-outcome associations (Elliot et al., 2010), whereas others have used abstract stimuli with unknown perceptual properties and unknown stimulus-outcome associations (Pessiglione et al., 2006; Palminteri et al., 2009). In such experiments it is hard to localize the brain regions that are engaged with coding stimulus novelty and the brain regions that engage in coding stimulus-outcome uncertainty. In the current experiment we manipulate both the stimulus novelty and outcome predictably by making both the novel and the familiar stimuli highly predictable and unpredictable by manipulating the uncertainty of CS-US associations.

Secondly, we used functional connectivity analysis (PPI) in order to see which regions that are differentially activite for novel as opposed to familiar stimuli. We believe that the results from this approach might reveal underlying mechanisms by which the brain operates proactively in recognizing novel and familiar stimuli.

8.1.2 Materials and Methods

8.1.2.1 Participants

Nineteen right-handed healthy normal volunteers (12 male, 7 female; mean age 25, range: 24-32) were recruited but only 18 participants (12 male, 6 female) included in the analysis. One of the participants was excluded from the analysis due to insufficient behavioural performance (participant couldn't be able to show any learning performance, and informed the experimenter saying about his random choices in all conditions). All of the participants were pre-assessed to exclude those with a prior history of neurological and psychiatric illness. All participants were invited to take part with written and verbal instructions about the experiment before the fMRI session. Prior to the main experiment, which involved fMRI, participants filled a written informed consent form before entering the scanner. All participants were debriefed after experimental session and paid according to their performance in the task. The study was approved by the Bedfordshire NHS Ethics committee board and Local Ethics committee.

8.1.2.2 Experimental Design

The experiment is made up of two parts:a familiarization task followed in a different session by the main experiment. The familiarization task was performed outside the scanner for pre-training the participants and the main experiment performed inside the scanner. For both sessions e-prime software (www.psychologysoftware.com) was used to control the presentation of stimuli and collect response data.

8.1.2.2.1 Familiarization task

The familiarization set of stimulus comprised 3 stimulus pairs. Each pair of stimuli was made from either Chinese or Agathadaimon font, which were counter balanced across participants, so that each participant was familiarized with only one font type. Three types of stimulus pair were used which differed in terms of the uncertainty of reward. The reasoning behind this is that by varying the reward contingency it allows us to differentiate the effect of expected and unexpected uncertainty of outcome that was previously mentioned in the introductory section. For a high outcome uncertainty pair the probability of reward is 0.6 for one of the options and 0.4 for the other, with corresponding probability of no reward being) 0.4 and 0.6 respectively. For the mid outcome uncertainty pair the probability of reward for one of the option was 0.8 (0.2 for no reward) whereas the probability of reward for the other option was 0.2. Finally the last stimulus category is deterministic rather than probabilistic such that if the participant chooses the correct option then the outcome is certain ie he will be rewarded on every occasion (probability of 1), whereas if s/he chooses the other option then s/he will always receive no reward. The familiarization task included 240 trials (80 for each level of uncertainty). Each trial begins with a fixation cross of 500 milliseconds presented in the middle of a LCD screen, which is followed by the presentation of one pair of stimuli for approximately 3 seconds. In that 3 seconds participants have to respond by choosing one of two response keys to indicate which of the 2 stimuli they prefer. This sequence of events is explained to each subject prior to commencing the first trial. Subjects are not instructed about the stimulus-reward contingencies. All responses are made with either the index finger or middle finger of the right hand. Responding is made on a Lumina response box, which is MRI, compatible. The response box has 4 key pads of which only the two at one end were used so that the subject could position their index and mid finger over the key and minimize hand movements. If the stimulus on the left side of the screen is preferred then the participant presses their index finger, and if the right side stimulus is preferred then the participant presses the key below their middle finger. After the participants make their choice feedback was immediately presented on the screen for about 3 seconds. For each correct prediction the rewarding feedback was large image of a £1 coin and for neutral feedback a £1 coin with a well defend red cross centred on the centre of the coin. Subjects were informed that the amount of money they would receive at the end of the study would be contingent on the money accumulated after completion of the both the familiarization task and main experiment. However, in order to avoid discrimination among participants at the end of the experiment same mount money provided to all participants that is of 20£.

8.1.2.2.2 Main Experiment

There were four scanning sessions in which each session was separated from the next by 2 minutes. In each scanning session there were in total 3 sets of novel stimulus-pairs and 3 sets of familiar stimulus pairs presented. For both pairs the outcome uncertainty adopted in the familiarization task was maintained, that is each stimulus pair was either low/mid/certainty outcome uncertainty (ie corresponding to the 3 probabilistic categories used in the familiarization task) (see **Figure 8.1**). The novel and familiar stimulus set were presented as separate blocks in order to avoid activity related cto ontextual novelty of so called temporal un-expectancy (i.e. unsure which type of stimulus pair would come next). This type of novelty (contextual) was commonly seen in tasks that are similar to the the oddball task (Knight et al., 1984; Knight and Scabini, 1998) where a novel stimulus is presented after familiar stimuli have been presented on several previous consecutions of trials and the effects of contextual novelty was avoided in the current experiment. During the scanning session each stimulus-pair was presented for 10 times (for each probability condition in a single session) making a total of 120 trials for the novel stimuli and 120 trials for the familiar stimuli for the whole fMRI sessions (each scanning session 30 novel and 30 familiar stimuli is used). On the basis of the learning curves observed in Experiment 1 asymptotes were reached by about 10 trials, although this varied according to outcome uncertainty. In order to maintain stimulus novelty throughout the main experiment for the set of novel stimulus pairs we did not want to exceed 10 trials for any one pair. Each trial in the scanning session took for about 10 seconds and the overall scanning time for a single participant took for about 48 minutes. Furthermore, at the end of each trial in the main experiment there was a random inter trial interval for about 2-8 seconds (jitter) in order to separate the rewarding outcome and stimulus presentation in the next trial. This inter-trial interval wasn't used in the familiarization -procedure. The basis of our judgement on the timing of the stimulus presentationis based on the statistical efficacy of the BOLD signal (T-values) of other studies and in particular our experience from the previous fMRI experiments.

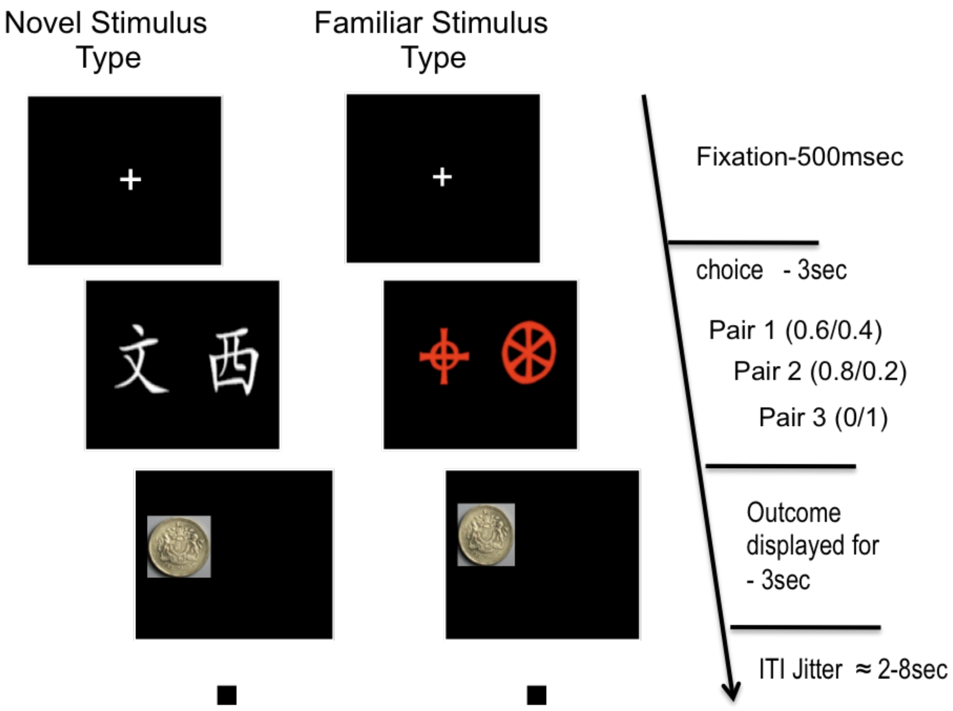


Figure 8.1 Illustration of task display for the probabilistic learning task during the scanning session.Subjects were presented with two abstract visual stimuli (either familiar or novel pair). During the choice subjects selected one of the stimulus, and their choice was followed by a feedback (a pound or nothing).

8.1.3 Functional Magnetic Resonance Image Acquisition

The functional imaging was conducted using 3-Tesla Siemens Magnetom MRI scanner to acquire gradient echo T2* weighted echo-planar (EPI) images with BOLD (Blood Oxygenation Level Dependent Signal) contrast (3x3x3-mm voxel size). Imaging parameters were optimized to minimize signal dropout in medial ventral prefrontal and anterior ventral striatum: we used a tilted acquisition sequence at 30° to the AC-PC line (Deichmann et al. 2003). Each volume compromised 36 axial slices of 3–mm thickness and 3-mm in plane resolution with a TR time (repetition time) of 3s. The flip angle was 90 degrees. T1 weighted structural images (1x1x1-mm voxel size) also acquired for each participant. Head movement was minimized with padding the participants' head.

8.1.4 Functional Magnetic Image Analysis

Image analysis was performed using statistical parametric mapping SPM8 software (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, United Kingdom). For all participants the images were realigned according to the first volume in order to correct for motion in the scanner. For all participants anatomical images were co-registered to functional EPI images and were normalized to a standard EPI template. Spatial smoothing was applied using a Gaussian kernel with full width half-maximum (FWHM) of 8 mm for each participant's data. Statistical analyses were performed on individual participant's data using the general linear model in SPM8. The fMRI time series data were modeled by a series of events convolved with a canonical haemodynamic response function (HRF). The presentation of the conditional stimulus and feedback screen was modeled as 1-second duration events. GLM thus included one regressor for each conditioned stimulus type (novel vs familiar) and each level of uncertainty of outcome (probability of reward for correct choice = 0.6, 0.8, 1.0) which we will refer to as novel_CS-06, novel_CS-08, novel_CS-1, familiar_CS-06, familiar_CS-08, familiar_CS-1). There were also six feedback types which shall be refered to asnovel_US-06, novel_US-08, novel_US-1, familiar_US-06, familiar_US-08, familiar_US-1). We haven't separated the feedback into further categories of rewarded and non-rewarded but coded them as 1 (reward delivered) and 0 (no reward delivered). The second level of the analysis consists of voxel-wise comparisons across subjects (one-sample t-tests and ANOVA) that were computed from the single subjects' contrast images treating each subject as a random effect. Coordinates of significant local maxima are reported in a standard stereotaxic reference space (MNI, Montreal Neurological Institute) and group functional overlays are displayed on the single subjects' anatomical scans.

8.1.5 Behavioural Results

To investigate the differences in learning performance for different stimulus pairs, we calculated the percentage of correct responses for each probability condition (0.4/0.6, 0.2/0.8, 0/1) and for each stimulus category (novel and familiar). Furthermore, in order to measure the learning performance between the novel and familiar stimulus sets, we compared these two conditions with the last 60 trials of the pre-training familiarity session. For the familiar stimuli, no significant difference is found in the performance during pre-training (familiarization session) and the main experiment (scanning session) for all three conditions. Also for the novel stimuli there was no significant difference (Novel_CS_0.6-Familiar_06, p>0.05, Novel_CS_0.8-Familiar_08 p>0.01, one tailed). However we found significant difference for Novel_CS_1 vs Pre_traning_1 (p<0.01, one tailed T=4.89), and Novel_CS_1 vs Familiar_CS_01 (p<0.01, one tailed T=5.189), (see **Figure 8.2**) was due to fewer choices made for the correct option for the novel pairs as would be expected in the early stages of learning.

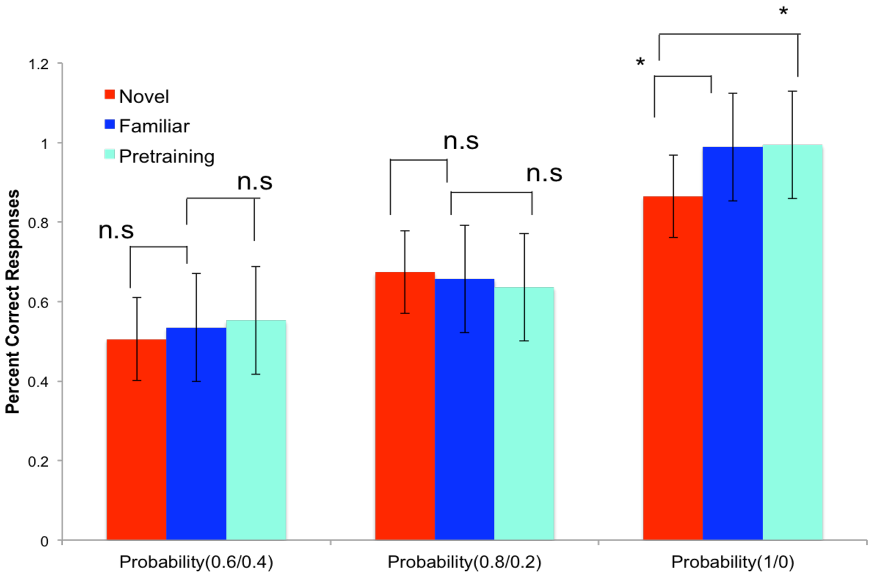


Figure 8.2 Behavioural data averaged across all 18 participants showing the percent of correct responses for familiar, novel during the scanning session and familiar stimuli during the last sixty trials of the pre-training session.

Response times (RT) significantly differ between novel and familiar (1069 ±22 ms and 819 ±99ms, respectively, see **Figure 8.3** for details). Responses were significantly slower for novel stimuli compared to familiar stimuli (T= 4.83, p<0.01, one tailed). But the RTs doesn't change between Familiar Stimuli during the scanning session and familiar stimuli during the end of familiarization session (Pre-Training) (p>0.01, one tailed).

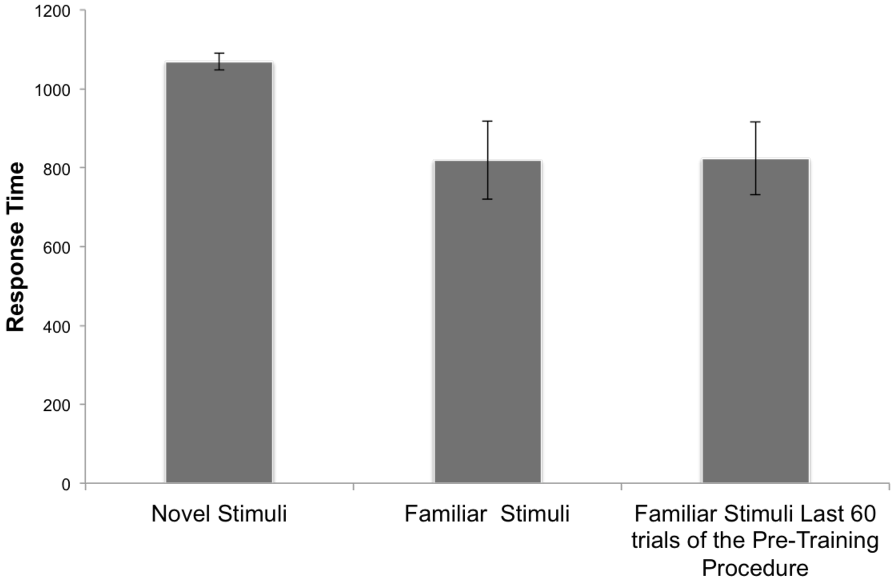


Figure 8.3 Plot of the reaction times for the novel, familiar stimuli and of familiar stimuli during the pre-training session regardless of the outcome probability. Participants were significantly faster in the familiar trials compared to novel trials.

8.1.6fMRI Results

First, we identified the neural correlates of conditional stimulus processing by comparing the (novel vs. familiar CS) contrast across all participants. Selected t-tests were conducted to compare conditions of interest. This analysis revealed BOLD responses for Novel CS > Familiar CS in several regions including the left and right DLPFC and parietal cortex (see **Figure 8.4**). The opposite contrast (Familiar CS > Novel CS) resulted in increased activation mainly in medial surface of the frontal cortex including medial frontal and anterior frontal cortex (see **Figure 8.5**). The coordinates for these comparisons (Novel and Familiar) are reported in **Table 8.1**. Also based on the above contrasts, we extracted the mean BOLD signal change from the peak voxels in order to investigate potential modulations by the respective probability factor. In particular, we were interested in how far regions involved in novelty would be influencedby the outcome uncertainty in the predictive relationship between CS and US (see **Figure 8.6**). The full extent of the activations is reported in **Table 8.1**.

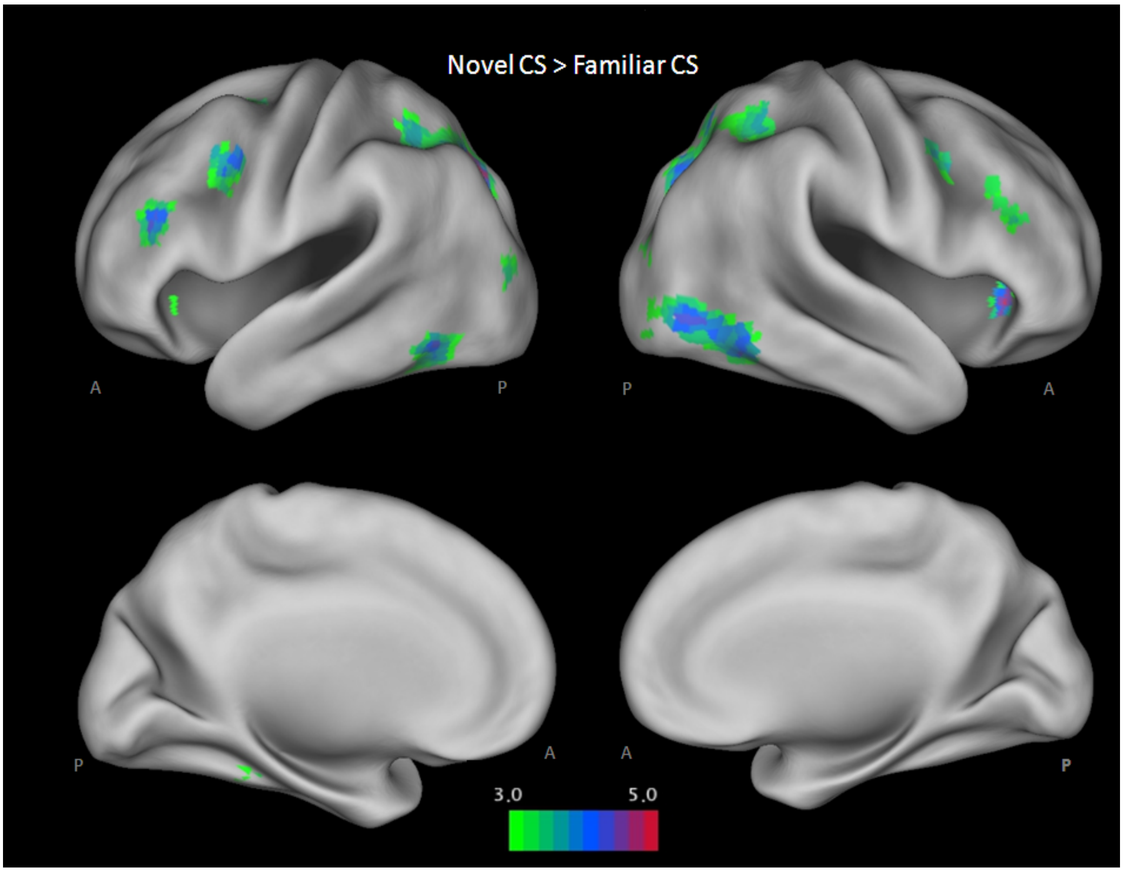


Figure 8.4 Activations shown for the (Novel CS > Familiar CS) contrast across all participants. For the 2 images at the top activation is overlaid on the lateral surfaces of the cortex separately for each hemisphere (images on the right side of the figure are right hemisphere whereas images on the left side of the figure are left hemisphere) and the images at the bottom show activation in the medial surface of the cortex. The activations were overlaid on the group average of inflated anatomical images using the CARET software. The color bar represents the T-Value where the green regions represent less significance and blue and red regions represent higher significant.

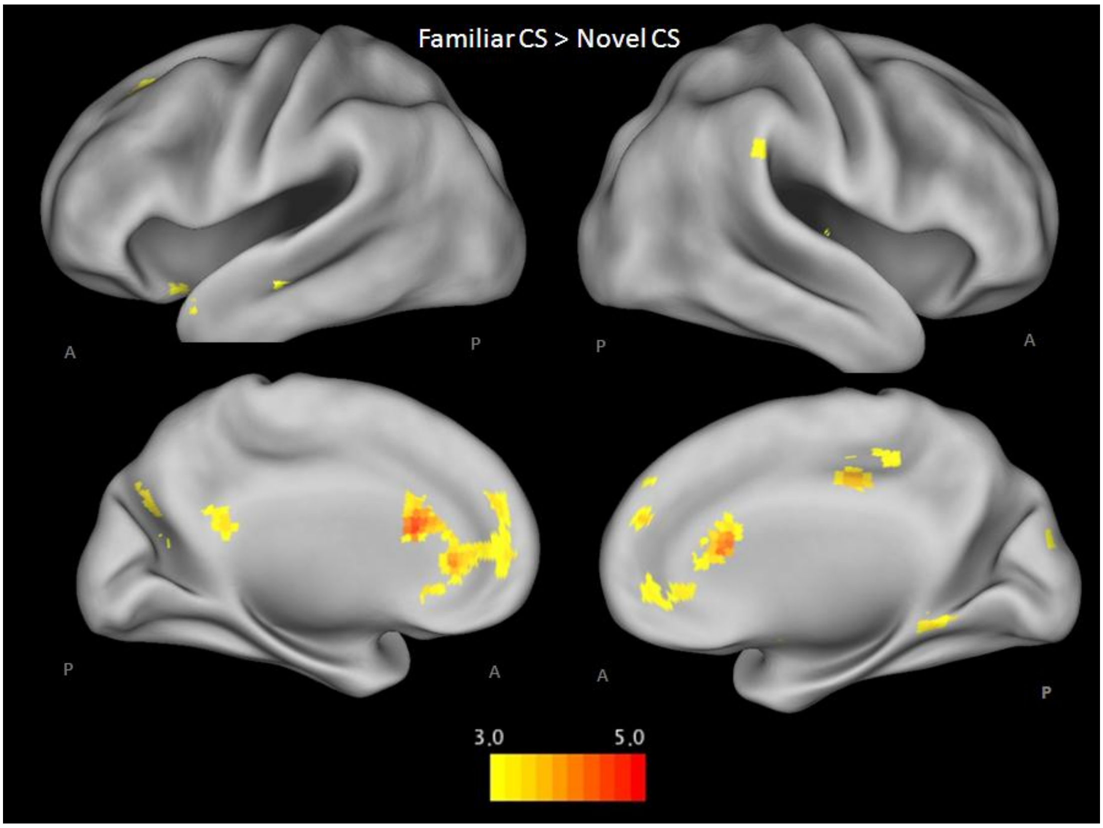


Figure 8.5 Activations shown for the (Familiar CS > Novel CS) contrast across all participants. For the 2 images at the top activation is overlaid on the lateral surfaces of the cortex separately for each hemisphere (images on the right side of the figure are right hemisphere whereas images on the left side of the figure are left hemisphere) and the images at the bottom show activation in the medial surface of the cortex. The activations were overlaid on the group average of inflated anatomical images using the CARET software. The color bar represents the T-Value where the yellow regions represent less significance and red regions represent higher significance.

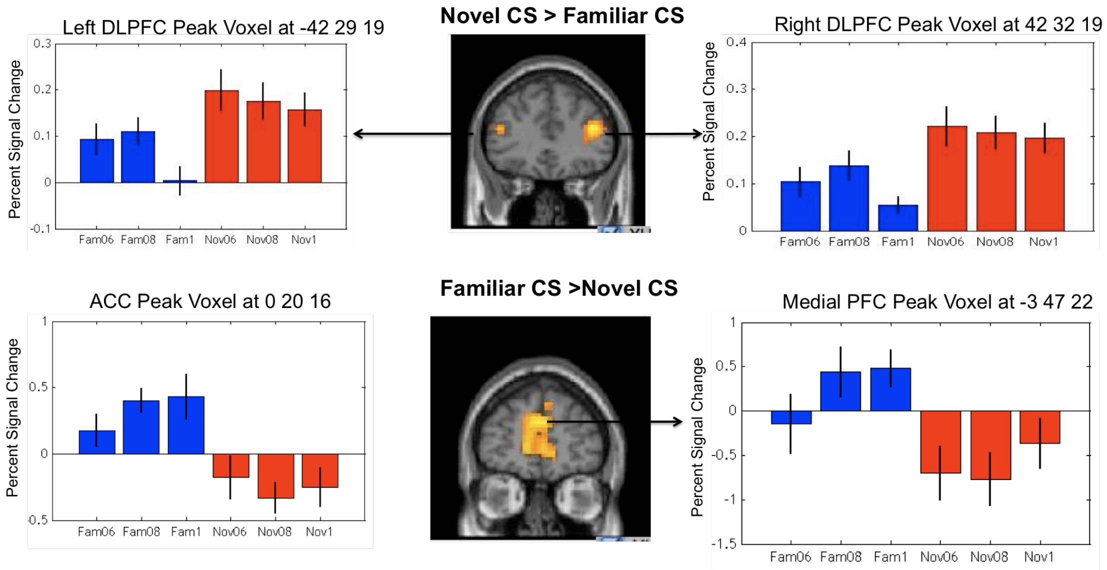
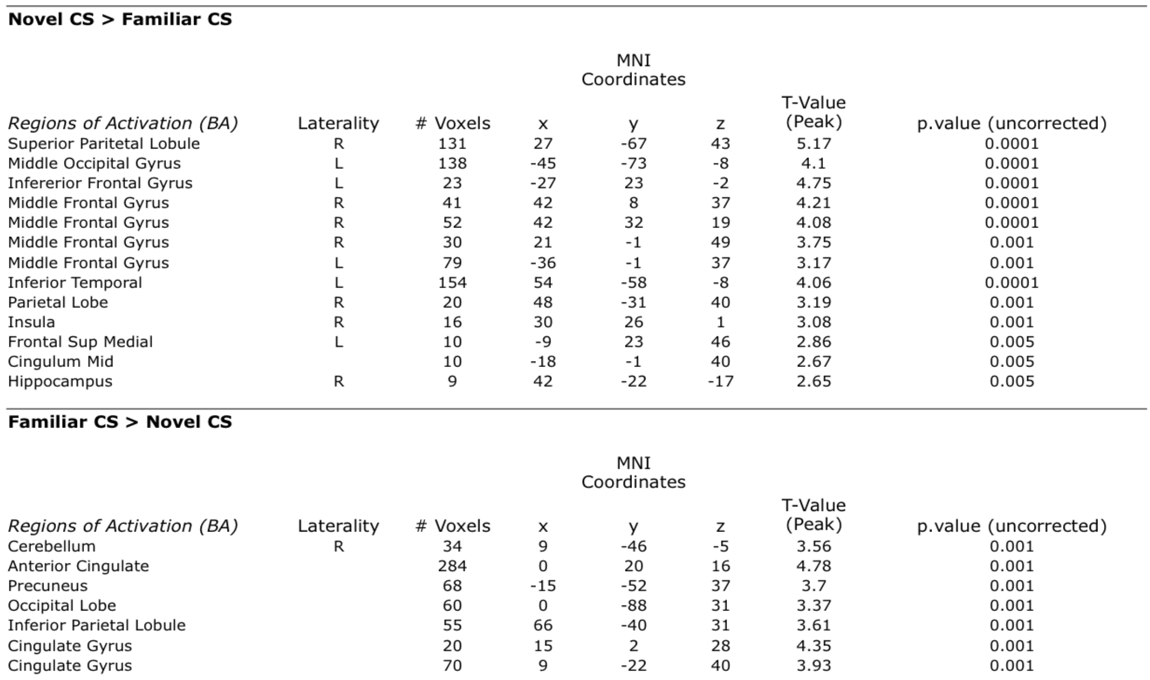


Figure 8.6 The coronal fMRI image at the top middle shows the Novel CS > Familiar CS contrastacross all participants. On the right and left side of the same figure percent signal changes and standard errors are shown for Novel CS (red) and Familiar CS (blue) for the right and left DLPFC respectively. At the bottom middle are the coronalfMRI image shows activation for the Familiar CS > Novel CS contrastacross all participants. On the right and left side percent signal changes and standard errors for Novel CS (red) and Familiar CS (blue) was shown for the ACC (on left side) and medial PFC (on the right side) respectively.

Table 8.1 The Subtraction Analysis Between Novel CS and Familiar CS



Novel CS > Familiar CS

				MNI Coordinates			
Regions of Activation (BA)	Laterality	# Voxels	x	y	z	T-Value (Peak)	p.value (uncorrected)
Superior Paritetal Lobule	R	131	27	-67	43	5.17	0.0001
Middle Occipital Gyrus	L	138	-45	-73	-8	4.1	0.0001
Infererior Frontal Gyrus	L	23	-27	23	-2	4.75	0.0001
Middle Frontal Gyrus	R	41	42	8	37	4.21	0.0001
Middle Frontal Gyrus	R	52	42	32	19	4.08	0.0001
Middle Frontal Gyrus	R	30	21	-1	49	3.75	0.001
Middle Frontal Gyrus	L	79	-36	-1	37	3.17	0.001
Inferior Temporal	L	154	54	-58	-8	4.06	0.0001
Parietal Lobe	R	20	48	-31	40	3.19	0.001
Insula	R	16	30	26	1	3.08	0.001
Frontal Sup Medial	L	10	-9	23	46	2.86	0.005
Cingulum Mid		10	-18	-1	40	2.67	0.005
Hippocampus	R	9	42	-22	-17	2.65	0.005

Familiar CS > Novel CS

				MNI Coordinates			
Regions of Activation (BA)	Laterality	# Voxels	x	y	z	T-Value (Peak)	p.value (uncorrected)
Cerebellum	R	34	9	-46	-5	3.56	0.001
Anterior Cingulate		284	0	20	16	4.78	0.001
Precuneus		68	-15	-52	37	3.7	0.001
Occipital Lobe		60	0	-88	31	3.37	0.001
Inferior Parietal Lobule		55	66	-40	31	3.61	0.001
Cingulate Gyrus		20	15	2	28	4.35	0.001
Cingulate Gyrus		70	9	-22	40	3.93	0.001

8.1.7 Conjunction and Interaction Analysis

The full network of regions activated across both novel and familiar event types was investigated using a standard analysis of variance (ANOVA) implemented in SPM8 where we look at the brain areas engaged by both CS categories (Novel CS and Familiar CS trials) and for all probabilistic categories (0.4/0.6, 0.2/0.8, 0/1). A widespread network of regions was activated for all CS types including basal ganglia, medial frontal regions, cingulate cortex, hippocampus, lateral parietal cortices, precuneus, and cerebellum ((P<0.000001, uncorrected) (see **Figure 8.7**).

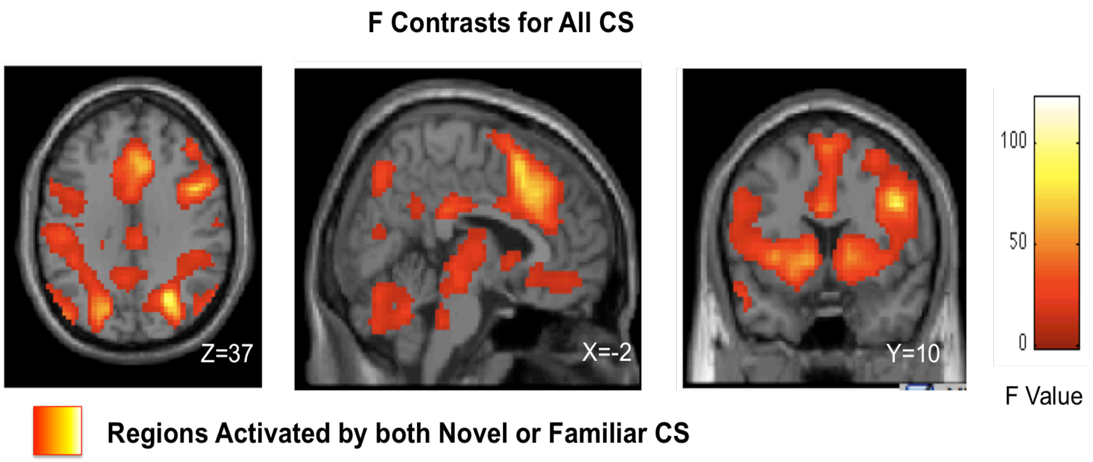

F Contrasts for All CS

Figure 8.7 Areas activated for all conditional stimulus types (P<0.000001, uncorrected). The color bar indicates the F score associated with each voxel.

We also looked at each CS category (Novel and Familiar) separately with an F-Contrast (see **Figure 8.8**). For all Novel CS regardless of the uncertainties in the outcome, the analysis revealed activity in various brain regions mainly in the basal ganglia, parietal cortex, cingulate cortex, bilateral DLPFC. A similar activation was also found for Familiar CS in the basal ganglia, parietal cortex and cingulate cortex. However, the activity in the DLPFC and posterior parietal cortex is more prominent for Novel CS compared to Familiar CS.

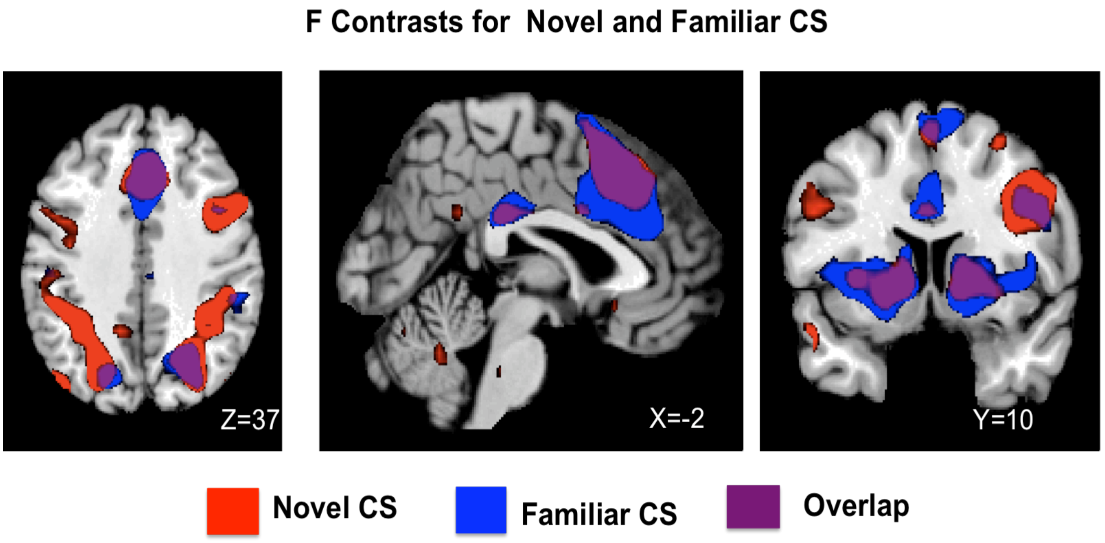

F Contrasts for Novel and Familiar CS

Figure 8.8 From left to right axial, sagittal and coronal slices show brain regions whose

activation are modulated by Novel and Familiar CS. The brain regions preferentially activated when Novel CS is presented (regions in red) compared to Familiar CS is presented (regions in blue) are overlaid on to single subject anatomical brain image. Regions that were active in both contrasts are shown in Purple. Both the Novel and Familiar contrasts were threshold with P<0.000001, uncorrected.

Furthermore, in order to verify the potential interaction between the stimulus novelty and uncertainty we applied a two-by-three repeated-measures ANOVA where stimulus novelty and uncertainty are the two factors. rANOVA was performed as implemented in SPM8 with stimulus novelty with two levels (novel vs. familiar), and uncertainty with three levels (0.4/0.6, 0.2/0.8, 0/1). The main purpose of the rANOVA was to investigate potential voxel-wise interactions rather than the overall main effects of stimulus novelty, since these were illustrated using the contrasts above. The main effect of uncertainty revealed activations in left insula and left putamen (see **Figure 8.9a**), whereas interaction between novelty and uncertainty was significant in inferior frontal gyrus (p<0.001, uncorrected) (see **Figure 8.9b**). The interaction analysis suggests that decrease in novelty result in increase in uncertainity. The coordinates for activation peaks in ANOVA analysis for the main effect of uncertainty and uncertainty novelty interaction reported in **Table 8.2.**

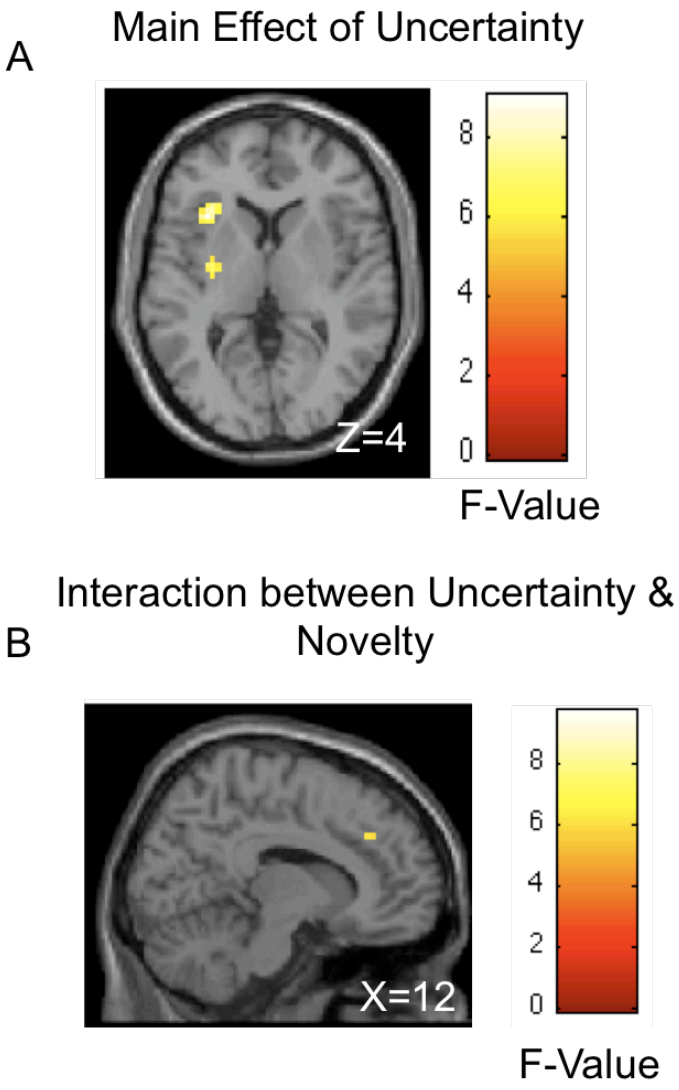

A
Main Effect of Uncertainty
8
6
4
2
0
Z=4
F-Value
B
Interaction between Uncertainty & Novelty
8
6
4
2
0
X=12
F-Value

Figure 8.9 A) Areas of left insular cortex and putamen showing uncertainty related activity for CS regardless of whether the stimulus is novel or familiar. B) Areas of inferior frontal gyrus show significant interaction between uncertainty and novelty.

Table 8.2 ANOVA The Effect of Uncertainty and Novelty & Uncertainty Interaction

Main Effect of Uncertainty

Regions of Activation (BA)	Laterality	# Voxels	MNI Coordinates x	 y	 z	F-Value (Peak)	p.value (uncorrected)
Insula	L	38	-33	20	4	9.04	0.001
Insula	R	8	33	23	-2	6.34	0.003
Putamen	L	22	-27	-7	1	7.07	0.001
Cingulate gyrus	L	10	-3	17	49	5.71	0.004

Novelty & Uncertainty Interaction

Regions of Activation (BA)	Laterality	# Voxels	MNI Coordinates x	 y	 z	F-Value (Peak)	p.value (uncorrected)
Parietal Lobe	L		-39	-40	31	9.69	0.001
Inferior frontal Gyrus	R		15	26	31	7.16	0.001

We also looked at areas involved in expected and unexpected uncertainty using a T-contrast. For the expected uncertainty, we looked at the contrast of Familiar CS with high uncertainty P=0.6 and Familiar CS with certainty P=1. For unexpected uncertainty we used the contrast of Novel CS P=0.6 and Novel CS P=1. The results revealed that the activity for expected uncertainty significantly increases in cingulate gyrus, bilateral insula and caudate nucleus. When we looked at the unexpected uncertainty with the contrast Novel CS P=0.6 > Novel CS P=1, we couldn't be able to identify any activation even with more liberal statistical threshold at p<0.005 (uncorrected).

Expected-Uncertainty: Familiar CS P=0.6 >Familiar CS-P=1

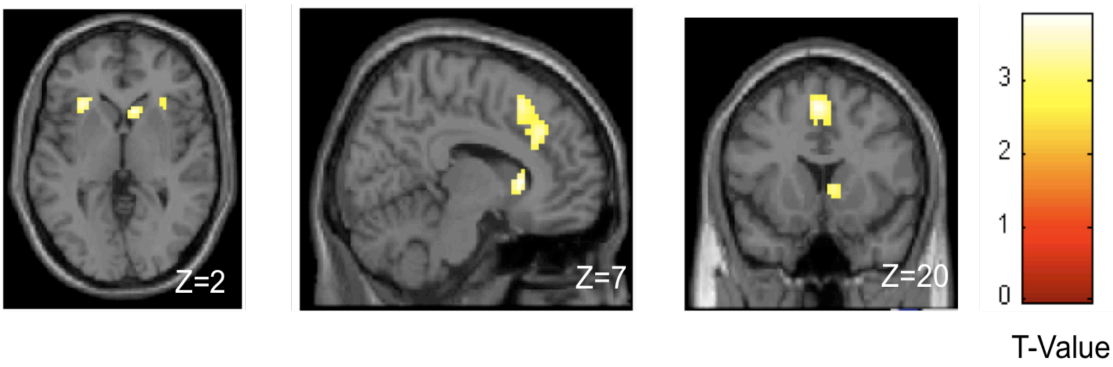


Figure 8.10 From left to right axial, sagittal and coronal slices show brain regions whose activation are modulated more by the high uncertainty Familiar CS P=0.6 than deterministic Familiar CS P=1. Areas of bilateral insular cortex, cingulate cortex and caudate showing significant activity at the level of P<0.001 (uncorrected).

8.1.9 Region of Interest Analysis for Stimulus Novelty in the Midbrain

Previous studies showed that midbrain dopaminergic neurons show increased activation for novel stimuli (Krebs et al., 2011). Based on this prior evidence, we decided to explore the relation between the effects of stimulus novelty by applying ROI analysis. To further characterize the activity in the midbrain for novelty processing Marsbar toolbox was used with SPM8 (http://marsbar.sourceforge.net, Brett et al., 2002) to perform Region of Interest (ROI) analyses. In order to be able to delineate spatially confined activity clusters within the midbrain, we used the coordinates of two anatomical ROIs that were localized in the center of left and right substania nigra of the AAL template (Automated Anatomical Labeling Template), which was shown in **Figure 8.10**. 5 mm spherical ROIs were created around the center of each region in order to include VTA. It is important to note that this midbrain ROI includes the substantia nigra as well as VTA since we were unable to separate VTA from as we were unable to obtain high enough resolution anatomical scans of individual participants and wereunable to identify a VTA-ROI template in MNI space. The beta values were then extracted for each participant in order to calculate group percent signal changes for Familiar and Novel CS contrasts.

The results showed that both left and right midbrain ROIs' produced greater activity for novel stimulus than for the familiar stimulus for the stimulus pair in which there is certainty (probability =1 ($p<0.05$, one tailed). However we couldn't identify any significant difference in Beta values for the pairs that have 0.6/04 and 0.2/08 ($p>0.05$, one tailed).

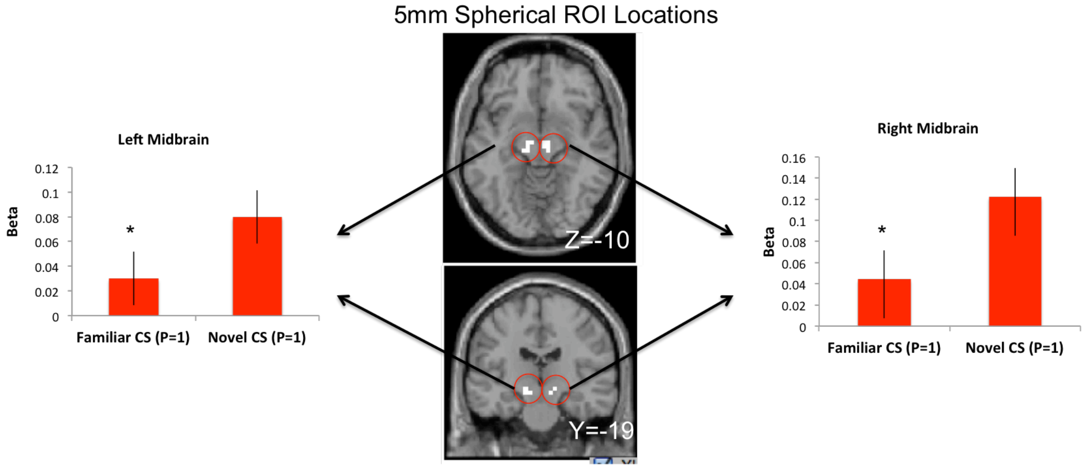


Figure 8.10 The results of ROI analysis above shows both right and left midbrain are shows significantly higher activation for Novel CS than for Familiar CS when the outcome is fully predictable p <0.05, (FWE).

8.1.10 Functional Connectivity Analysis

We investigated the functional connectivity for the novelty-related increase of the reward-anticipation signal within DLPFC. In order to do that, a psycho-physiological interaction analysis (PPI) is conducted as is implemented in SPM8. PPI analysis assesses how the activity within brain networks is modulated by varying task conditions in an fMRI experiment. Specifically, the individual DLPFC seed activity for the physiological signal was extracted for each participant using the contrast Novel CS >Familiar CS.The Novel CS >Familiar CS contrast reflects the additional enhancement of the anticipation response by novelty. The seed was defined as 8mm spherical ROI around each participants peak voxel in lateral prefrontal cortex (See, **Figure 8.11a**). In order to highlight the center of the seed region we applied 8mm Gaussian smoothing to the peak voxels as seen by **Figure 8.11b**. Note that this smoothing operation is nothing to do with PPI analysis and just for representational purposes for the center of mass of the seed region. The PPI term was created for each participant by multiplying the deconvolved and mean-corrected BOLD signal with the psychological vector. After convolution with the HRF, mean correction, and orthogonalization, the three regressors (PPI term, physiological vector, and psychological vector) were entered into the statistical analysis to determine condition- dependent changes of functional connectivity over and above any main effect of task or any main effect of activity in the corresponding brain areas. In the PPI contrasts, the PPI term was computed against implicit baseline. Random-effects analyses were performed on single-subject PPI contrast images (p < 0.001, uncorrected) and carried to second level group analysis.

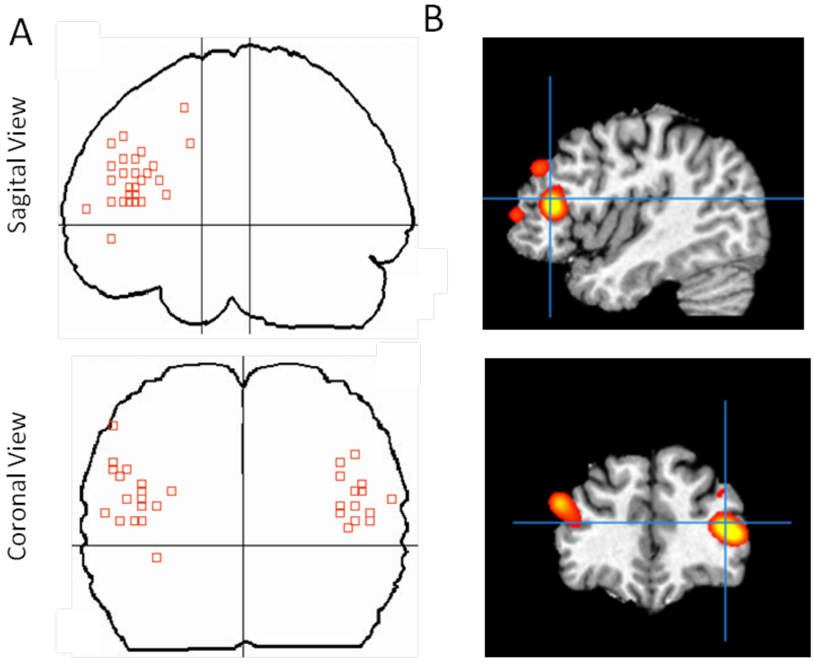


Figure 8.11a) On the upper and lower left side of the figure glass brains are shown from a sagittal and coronal view. Red squares represents individual peak voxels of the participants for the Novel CS> Familiar CS contrast. The spherical ROIs for seed-region activity was taken from individual participants' peak voxels. b) On the upper and lower right side of the figure the coloured areas show smoothed (8mm Gaussian filter) individual peak voxels for demonstrative purpose (smoothing was not applied during PPI analysis) to show the centre of
mass of the peak voxels.

The PPI that focused on novelty-related changes in the context of reward (seed contrast: Novel_CS vs. Familiar_CS) revealed co-variations between the DLPFC seed region and bilateral ventral striatum (NAcc), bilateral amygdala, insula, and several small clusters in lateral prefrontal cortex and cingulate gyrus (see **Figure 8.12**). The clusters in DLPFC were highly overlapping with the novelty-sensitive cluster observed in the Novel CS>Familiar CS contrast (compare Fig. 2A, right panel). Please see **Table 8.3** for a full list of activations.

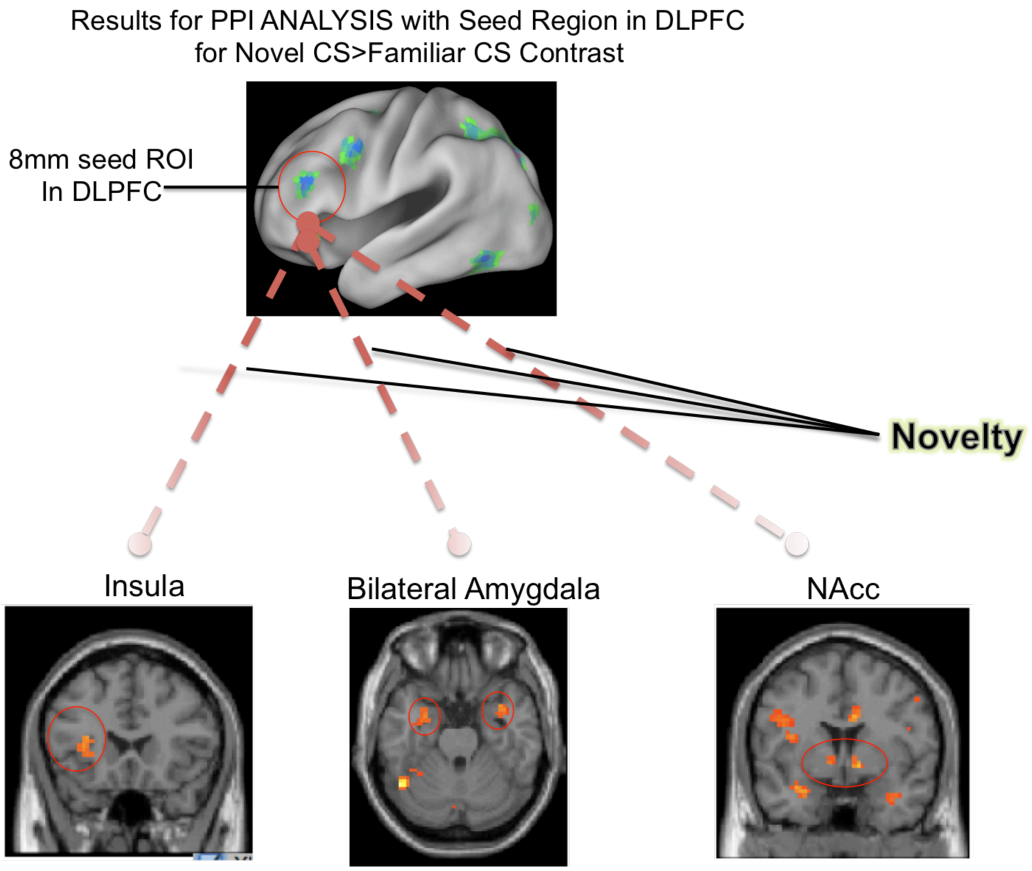
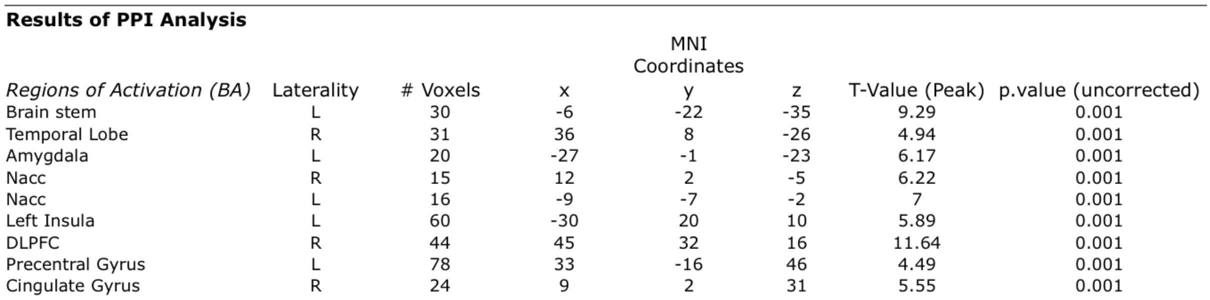

Results for PPI ANALYSIS with Seed Region in DLPFC
for Novel CS>Familiar CS Contrast
8mm seed ROI
In DLPFC
Novelty
Insula
Bilateral Amygdala
NAcc

Figure 8.12 Functional connectivity (PPI) with the DLPFC. Functional connectivity between the novelty sensitive right and left DLPFC and bilateral amygdala, insula, ventral striatum are show by the top and three bottom images respectively. Activities in these regions are increased for novel as compared to familiar reward-predictive stimuli.

Table 8.3 Results of the PPI analysis. Coordinates in MNI space.



Results of PPI Analysis

Regions of Activation (BA)	Laterality	# Voxels	x	MNI Coordinates y	z	T-Value (Peak)	p.value (uncorrected)
Brain stem	L	30	-6	-22	-35	9.29	0.001
Temporal Lobe	R	31	36	8	-26	4.94	0.001
Amygdala	L	20	-27	-1	-23	6.17	0.001
Nacc	R	15	12	2	-5	6.22	0.001
Nacc	L	16	-9	-7	-2	7	0.001
Left Insula	L	60	-30	20	10	5.89	0.001
DLPFC	R	44	45	32	16	11.64	0.001
Precentral Gyrus	L	78	33	-16	46	4.49	0.001
Cingulate Gyrus	R	24	9	2	31	5.55	0.001

8.1.11 Activity During the Outcome Period

Here we compare the outcome to novel stimulus pairs and familiar stimulus pairs outcomes (regardless of the feedback valence). The reasons for this are that the outcomes for the familiar stimulus pairs would be expected (predicted by the CS) whereas those for the novel pairs are not predicted. The contrast Novel US >Familiar US showed increased activity in bilateral striatum and right ventro-medial frontal cortex (see **Figure 8.13**). For the opposite contrast (Familiar US > Novel US) the areas showing greater responses to Familiar outcomes than novel outcomes were found in the border of medial OFC andanterior portion of the cingulate cortex and precuneus (see **Figure 8.13**).

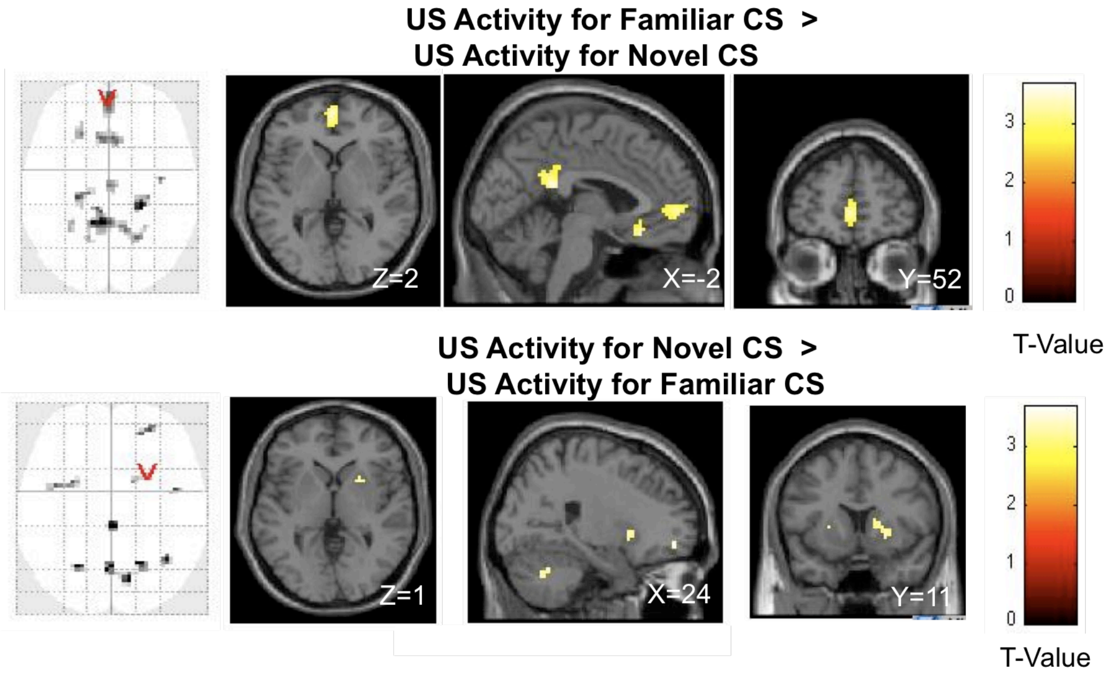

US Activity for Familiar CS >
US Activity for Novel CS
Z=2
X=-2
Y=52
3
2
1
0
T-Value
US Activity for Novel CS >
US Activity for Familiar CS
Z=1
X=24
Y=11
3
2
1
0
T-Value

Figure8.14 A) Areas of ventral PFC showing outcome related activity for Familiar US> Novel US contrast. The first image on the left side is the glass brain from the axial view shows group effects for the same contrast that enables one to appreciate activations in all locations and levels in the brain simultaneously. The next three figures are group random effects results that are superimposed on axial, sagittal and coronal slices overlaid on the single subject structural MRI image [at the Montreal Neurological Institute (MNI) coordinates indicated in the bottom right corner of each image]. **B)** Similar to the upper figure, the figure below shows group random effect results for the areas of striatum that show outcome related activity for the Novel US > Familiar US contrast. Significant effects are shown at p < 0.005, uncorrected.

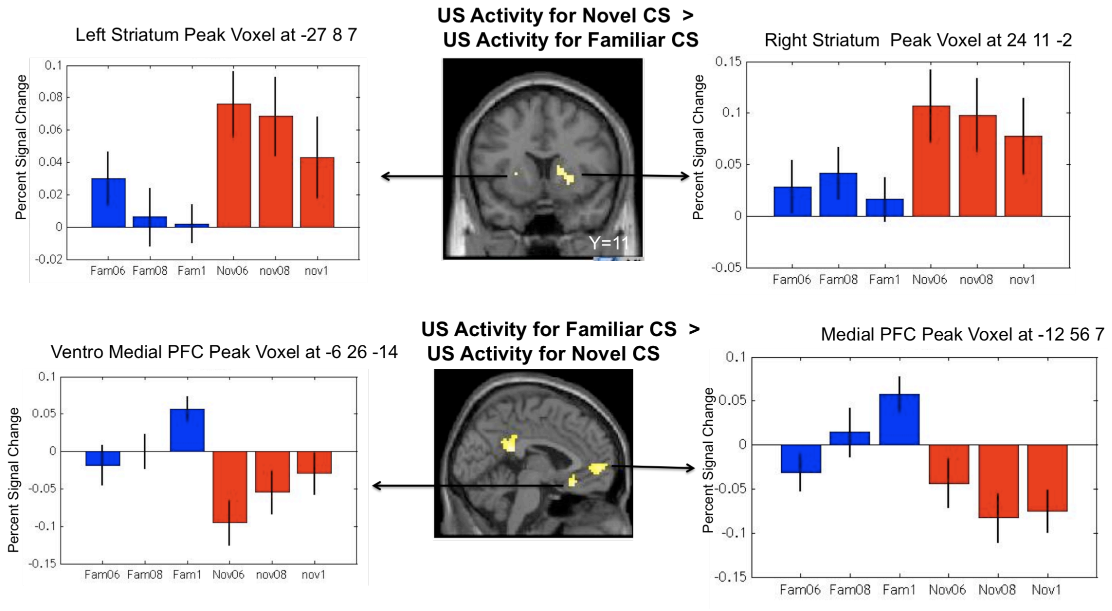


Figure8.15 On the above coronal fMRI image activation was shown for the (US Activity of Novel CS > US Activity of Familiar CS) contrast across all participants ($p<0.005$, uncorrected). Bar graphs on the left and right side of the upper figure show percent signal changes and standard errors for US Activity for Novel CS (red) and US Activity for Familiar CS (blue) for the left and right striatum respectively. b) The sagital fMRI image below activation was shown for the (US Activity for Familiar CS > US Activity for Novel CS) contrast across all participants. On the right and left side percent signal changes and standard errors for US Activity for Novel CS (red) and US Activity for Familiar CS (blue) was shown for ventro-medial PFC (on left side) and medial PFC (on the right side) respectively.

8.2 Discussion

8.2.1 Novelty Responses during The Anticipation Phase

Our results showed that novel conditional stimuli, compared to the pre-trained familiar stimuli elicited a greater activation mainly in the dorsolateral prefrontal cortex and posterior parietal cortex with additional activations found in the hippocampus and right insula. On the other hand when we performed a subtraction analysis to compare the regions that show higher activation for familiar stimuli than novel stimuli we found anterior cingulate gyrus activity. The regions activated for the novel stimuli involved a network that is frequently referred as dorsal fronto-parietal attention network (Quintana and Fuster, 1999; Corbetta and Shulman, 2002). Previous studies have shown that this networkusually gets activatedat the early stage of learning when attention is required (Duncan and Owen, 2000; Corbetta and Shulman, 2002). In addition dorsal-frontal executive functions are especially important in the early, more intentional phase of learning, as compared tothe later, more automatic phase (Jenkins et al., 1994; Jueptner, 1997, Antzoulatos and Miller, 2011).In fact in their meta-analytic review Duncan and Owen (2000) identified activation in the dorsa-lateral prefrontal cortex in a range of cognitive tasksand they suggest that this region is involved in processing stimulus novelty. Also evidence from ERP studies showed that dorsolateral PFC (DLPFC) is important for novelty processing (Alexander et al., 1995; Knight and Scabini, 1998; Daffner et al., 2000). In humans Kishiyama et al., (2009) showed that patients with lateral

PFC damage were impaired in recollection based recognition memory for novel items compared to non-novel items. Additional evidence from primate studies also showed thatthe effects of stimulus novelty were eliminated when the DLPFC was lesioned (Parker et al., 1998). It is possible that DLPFC might actually be involved in encoding novel items in to working-memory (van Schouwenburg et al., 2010a, 2010b). In fact several studies suggested that DLPFC is involved in "attentional selection" of the items that are stored in working memory (Miller, 1999; Passingham & Rowe, 2002). These studies argued that multiple attentional selection processes occur during remembering of an item from working memory and the activations found in the DLPFCmight account for the selection process of this items (Passingham & Rowe, 2002).

Our results also indicate that DLPFC activity might be involved in the response selection process when the outcome of each response is uncertain. Evidence to support this is that the novel stimuli elicited a stronger activation for stimulus-outcome pairs that are uncertain (0.6/0.4 probability pair) compared to stimulus-outcome pairs that are predictable (1/0 probability pair) (See **Figure 8.6**). In fact previous fMRI studies showed that LPFC is involved in exploratory decision-making and the activity in DLPFC is correlated with trial-by-trial estimates of relative uncertainty (Badre et al., 2012). Moreover, several other studies also showed a strong activation in the LPFC when the participants required figuring out complex rules in the Wisconsin Card Sorting Task (WCST) (Nakahara K, et al. 2002; Mansouri et al. 2006). It is possible that the DLPFC activity that is found in the current study might account both for the novelty and attention aspects of conditional stimuli, whereas anterior cingulate gyrus and medial frontal cortex activity might account for the familiarity aspects of conditional stimuli as suggested by previous studies (Chiba et al., 1997; Passingham et al., 2000; Maddock et al., 2001; Inase et al., 2006)

8.2.2 Functional Connectivity of DLPFC

The current results demonstrated that the human lateral PFC is important for producing stimulus novelty effects. Combined with our connectivity analysiswe further identified subregions of a distributed novelty-processing network, including the amygdala, ventral striatum, and insula. Previous neuroanatomical connectivity studies showed that DLPFC is higly connected with the dorsal striatum (Haber and Knutson, 2010) and this might seem to contradict our findings on functional connectivity betweenthe ventral striatum and the DLPFC for novelty. However several studies suggest that these two regions of striatum might be functionally coupled (Gao et al., 2007; Ballard et al., 2011). In fact there are two possible sources of explanations for this. Firstly, previous studies showed that activity in DLPFC could be modulated by the midbrain dopaminergic activity (Durstewitz et al., 2000; Seamans and Yang, 2004; Wang et al., 2004). In this case the basal ganglia might serve to choose which contents should be gated into dorsolateral PFC, to be subsequently maintained in working memory (Braver & Cohen, 2000; Frank et al., 2001; Orielly and Frank; 2006; Hazy et al., 2006; 2007). Recent evidence in neuroimaging (McNab, T. Klingberg, 2008) and PD patients (Moustafa et al, 2008) provide additional evidence on the role of the basal ganglia in gating working memory representations, as well as modulation by DA.

In the second scenario DLPFC might function asan inhibitory motor region but in a selective manner to ensure the correct behavioural choice is made (Chevalier and Deniau, 1990). The selection-related inhibition may constitute the DLPFC activation during response inhibition. Connections to the ventral striatum might mediate such selective inhibitions as well as disinhibitions (Redgrave et al., 1999).

We also found amygdala activation from the functional connectivity analysis. Anatomical studies have shown that the basal nucleus of the amygdala are the main source of inputs to the ventral striatum (Russchen et al., 1985; Fudge et al., 2002) and amygdala has been previously associated with arousal or attention (Murray, 2007). We think thatamygdala is working collaboratively withother circuits for novelty detection, which is also suggested by other studies (Writh et al., 2003; Schwartz et al., 2003; Blackford et al., 2010; Balderston et al., 2011).

Finally, although recently, studieshave proposed that novelty signals originate in the hippocampus and modulate the activity of dopamine neurons in the SN/VTA (Lisman and Grace, 2005). We were able to identify a small cluster of hippocampus activity for the novel CS > familiar CS contrast perhaps because our task is a reinforcement learning task rather than a memory task. Moreover because our functional connectivity analysis doesn't look at the directionality of connections between regions the question of how these regions interact within the network is an important avenue of future research.

8.2.3 Novelty responses in the Striatum and Midbrain

We showed that activity in the midbrain for novel CS is significantly greater than for the familiar CS that has predictable outcomes. Previous studies showed that, SN/VTA showed increased hemodynamic responses to novel stimuli (Bunzeck and Duzel, 2006; Wittmann et al., 2008). In fact there is a wide variety of electrophysiology studies in monkeys and rodents suggest that dopamine neurons code more than reward prediction errors where they are also involved in alerting and novel events (Schultz, 1998; Redgrave et al., 1999; Horvitz, 2000; Lisman and Grace, 2005; Redgrave and Gurney, 2006; Joshua et al., 2009; Schultz, 2010). Moreover when we looked at the basal ganglia separately for The familiar and novel stimulus sets we found significant activity in the ventral striatum for both novel and stimulus types.

8.2.4 Uncertainty Related Responses

We found that the main effect of uncertainty is the activity in the insula and the putamen whereas the interaction between uncertainty and novelty activates inferior frontal gyrus. Previous electrophysiological studies in monkeys indicate that dopaminergic neurons not only code a transient reward prediction error signal but also a sustained signal covarying with reward uncertainty (i.e., reward probability= 0.5) (Fiorillo et al., 2003). This finding also has been replicated in humans and uncertainty related activation was shown in the basal ganglia (Preuschoff et al., 2006) and this is in agreement with the previous findings which showed that putamen is involved in coding motor actions with uncertainty (Deffains et al., 2010; Vincente et al., 2012). Moreover our results showed that expected uncertainty produces significantly increase of activity in cingulate gyrus, bilateral insula and caudate nucleus. Previous studies suggested expected and unexpected uncertainty play complementary but distinct roles in top-down and bottom up attention and both forms of uncertainty are suggested to increase the rate of learning (Yu and dayan, 2003).

8.2.5 Linking Automaticity with Processing Novel and Familiar Stimuli

Our discussion so far has emphasized a role for novelty related activation, which focused on the dorsal fronto-parietal attention network, but these results are also consistent with rostro-caudal shift of activity during automaticity (Graybiel, 2008; Ashby et al., 2010) and can be interpreted in this direction. During the past ten years research in our understanding of how the brain responds to familiar and novel learning situations have reveals a number of important findings. One of these findings is the shift of activation from anterior to posterior regions in the networks linking cerebral cortex to subcortical striatal structures during learning of habits (Salmon & Butters, 1995; Hikosaka et al., 1999; Costa, 2007; Graybiel, 2008; Belin et al., 2009; Ashby et al., 2010). Studies that perform detailed analysis of novelty vs familiarity in learning of motor-movements showed that the brain regions involved in the familiarization procedure brain activation depend not only on the reaction time per se but effected by various factors including task domain (executive, visual or motor) (Jonides, 2004; Seger & Spiering, 2011), amount of practice (Waldschmidt & Aschby, 2011; Bassett et al., 2011), task subcomponents, dual task performance (Poldrack et al., 2005) or the speed of the response times (Kelly & Garavan, 2005; Saling & Phillips 2007, Bor & Owen, 2007; Helie & Cousineau, 2011). Our results showed that the overall effects of practice during the familiarization phase significantly decreased the reaction times of participants compared to scanning phase of the experiment. This behavioural effect was consistent with the existing literature on the development of automaticity suggesting that more automatic responses should be faster and more accurate. Overall, these results are consistent with a series of recent studies showing that portions of the lateral prefrontal cortex is involved in processing of novel stimulus response pair (Assad et al., 1998; Duncan & Owen, 2000; Ranganath and Rainer, 2003).

Chapter 9

Experiment 2: Neural Correlates of Predicted-Value, Prediction Error and Learning Rate for Novel and Familiar Stimuli

9.1 Model-Based Analysis of Novel versus Familiar Stimuli

9.1.1 Introduction

In this chapter the data that was presented in **Chapter 8** is re-analysed in a model-based fashion. The reinforcement-learning model used in this chapter is a modified version of the reinforcement-learning algorithm that was used in **Chapter 6 and Chapter 7**. The difference between the algorithms lies in the way they update the learning rate (see the Methodology section below). The aim of the current chapter is similar to that described in **Chapter 6-7**, namely to identify the neural correlates of predicted-values for familiar and novel chosen stimulus and related prediction errors. An objectives is also to compare the results of the second experiment with the first experiment to check whether the brain regions that are involved in coding reward predicted-values during early versus late learning trials (that refers to first 10 and the last 10 trials) are similar to predicted values for familiar and novel stimuli respectively in the second experiment.

In addition to that, we reviewedseveral mathematical accounts of adaptive learning rate algorithms in **Chapter 4** and suggested that the neural networks that are responsible from coding the learning rate should change their firing characteristics in real circumstances depending on the familiarity and predictability of the environment. It has also been reviewed that this might correspond psychologically to either the level of top-down attention or stimulus novelty (Pearce & Hall, 1980, Schumajuk, 1997 see also the discussion section of **Chapter 8**). Therefore in order to test that possibility we looked at the neural correlates of the learning rate parameter for the novel and the familiar stimuli.

9.1.2 Methodology

9.1.2.1 Reinforcement-Learning Model

The model used in this chapter is the same as the one used in **Chapter 6-7** where the predicted values are updated only for the symbol that is chosen. The updating of the predicted value of the chosen symbol is based on the difference between the outcome and the estimated value. Similarly a softmax action selection rule was used for updating the probability of selected options (High probability "hp", low probability option "lp"). For example if the participant chose the high probability symbol the probability of choosing the symbol is calculated by the following equation.

$$p(hp) = \frac{e^{\beta Q(hp)}}{e^{\beta Q(hp)} + e^{\beta Q(lp)}} = \frac{1}{1 + e^{-\beta(Q(hp) - Q(lp))}}$$

$\beta$is the inverse temperature, which relates to the randomness in selecting between two options. For example, high βmeans higher probability of random action selection (0<β<1). The prediction error was calculated by the difference of actual reward ,received (r), minus the value of choosing that symbol (i.e., high probability option).

$$\delta = r - Q(hp)$$

We set the value of the reward *r* to 1 for positive feedback and 0 for neutral feedback. The predicted values Q (high probability, low probability) were also set to 0 at the beginning of each learning session. When the outcome for the particular symbol was represented, the value of choosing that symbol was updated by the following equation

$$Q(hp) = Q(hp) + \alpha_t \delta_t$$

The learning rate controls the amplitude of change and is determined by a standard recursive procedure as follows:

$$if \delta_t > 0$$

$$\alpha_{t+1} = \alpha_t - \alpha_{a,t} \theta$$

$$if \delta_t < 0$$

$$\alpha_{t+1} = \alpha_t + \alpha_{a,t} \theta$$

$$if \delta_t = 0$$

$$\alpha_{t+1} = \alpha_{a,t}$$

To determine the parameters with which the model best fit with the behavioural data of participants' actual choices, the likelihood function *l*(Q|z) was calculated for each set of parameters (Q=α_0, β, θ)with participants actual choices (z) . The model fitting procedure is as follows: we first calculate the action values sampling from all possible combinations of parameter values (incremental search). Then we estimate the probabilities for all possible parameter values for each trial. Then the probability that a participant can select the symbol *a*in trial *i* is inserted in the likelihood function. The highest likelihood parameters were selected as best fits. Note that for the learning rate parameter, we only determine the best fitted initial learning rate, α_0 . The Matlab algorithm can be found from the **Appendix C.**

9.1.3 Behavioural Results of the Model-Fitting Procedure

The learning rate is a fundamental feature of behavior that determineshow agents should adjust the decisions that they make in the faceof changing circumstances (Behrens et al., 2007). Behavioral analysis of the fitted parameters showed that during learning of high-uncertainty novel stimulus pairsthe mean learning rate is higher than in trials with high-uncertainty familiar pairs. Even in the beginning of reinforcement learning sessions the learning rate for novel stimuli starts higher than the familiar pair which indicates that more weight is given to prediction errors during updating the predicted values of novel pairs (**Figure 9.1**).

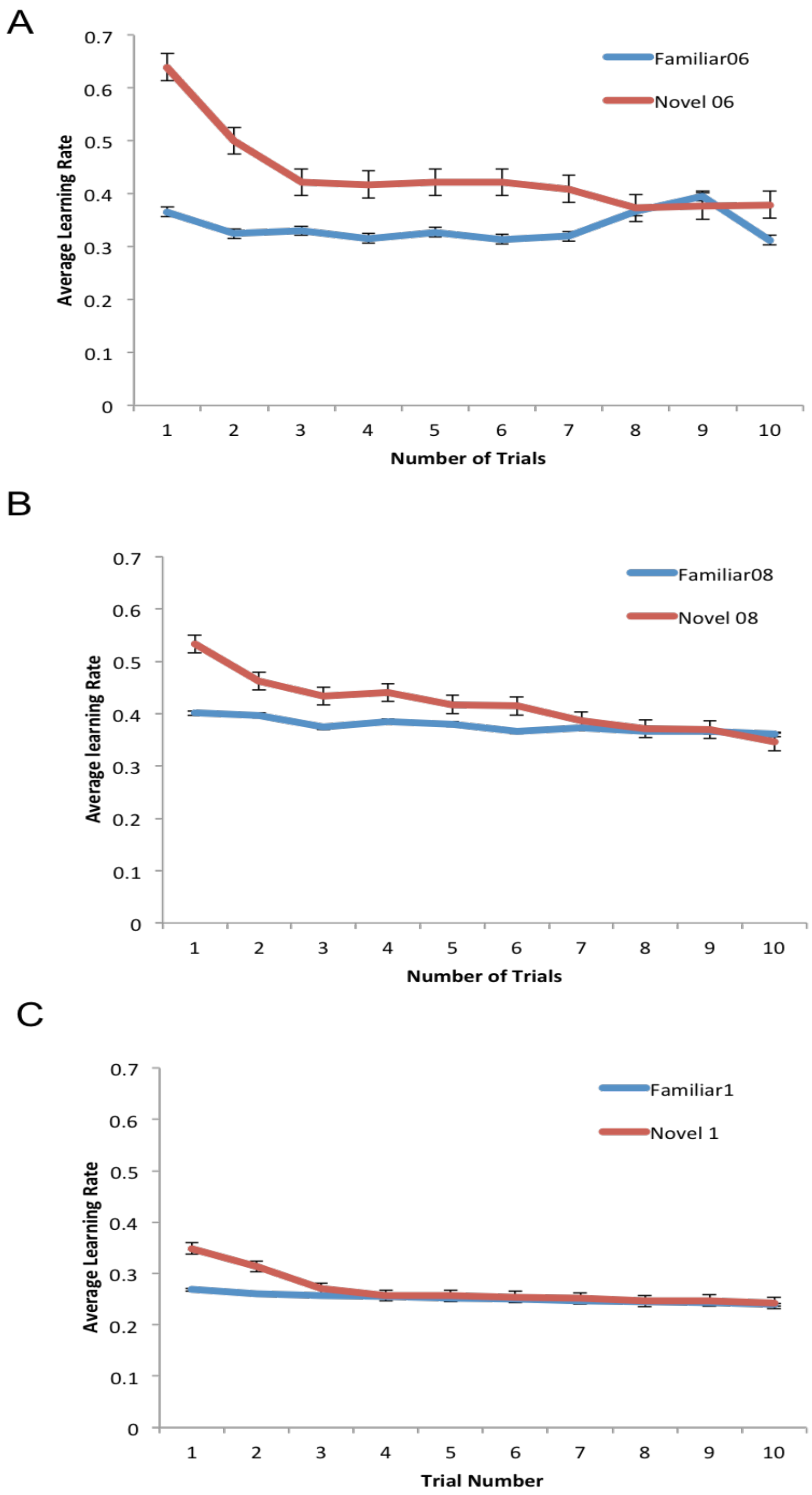


Figure 9.1 Average changes in learning rate across all 18 participants categorized according to reward probability **A)** The learning rate parameter for each trial of each individual is averaged for the high uncertainty familiar and novel stimuli. **B)** Average learning rate change for mid-uncertainty familiar-novel stimuli, **C)** Average learning rate change for deterministic familiar-novel stimuli.

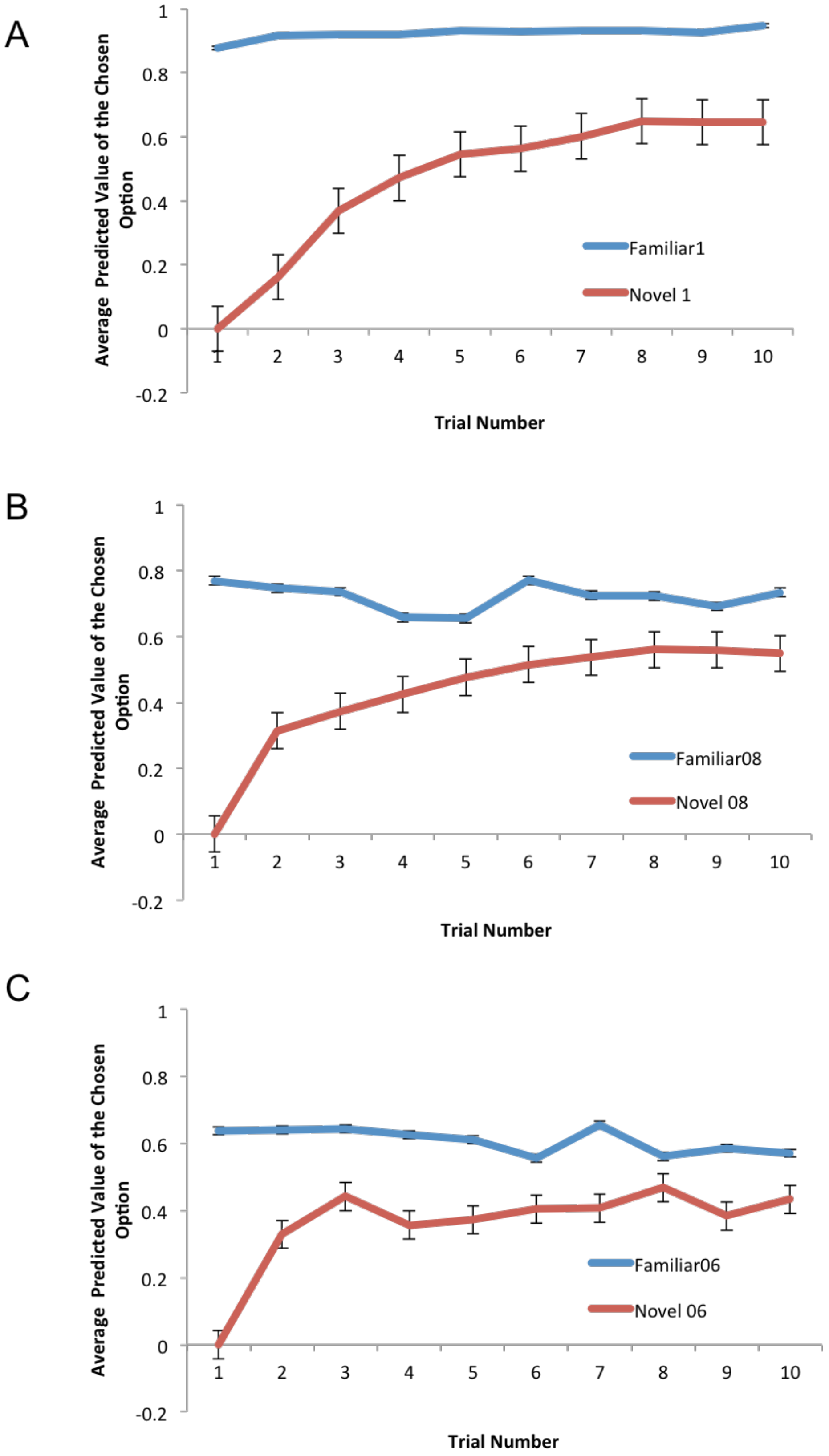


Figure 9.2 Average changes in the predicted value of chosen stimulus across all 18 participants categorized according to reward probability **A)** The predicted value for each chosen symbol of each individual is calculated and averaged for high uncertainty pair of familiar and novel stimuli. **B)** Average predicted-value changes for 0.8 probability

familiar-novel stimuli, **C)** Average predicted-value change for 1 probability familiar-novel stimuli.

9.2 Model-Based fMRI Results

9.2.1 Neural Correlates of Prediction Errors For Novel and Familiar Stimuli

We first determined which brain areas are more strongly activated for the prediction errors with the novel stimuli. The results showed that activity in several regions was significantly positively correlated with the reward prediction error signal,for the novel stimuli including the ventral striatum, medial frontal cortex, posterior cingulate gyrus, right medial frontal gyrus (BA10), left insula and bilateral extra nucleus (full list of activation was presented in **Table 9.1**). Secondly, in order to determine the difference between novel and familiar prediction errors, the prediction errors for the familiar stimuli were regressed against fMRI data. The results of the whole brain analysis revealed that there was no significant prediction error activity for the familiar stimuli with the p value of less than 0.001 (see, **Figure 9.3**).

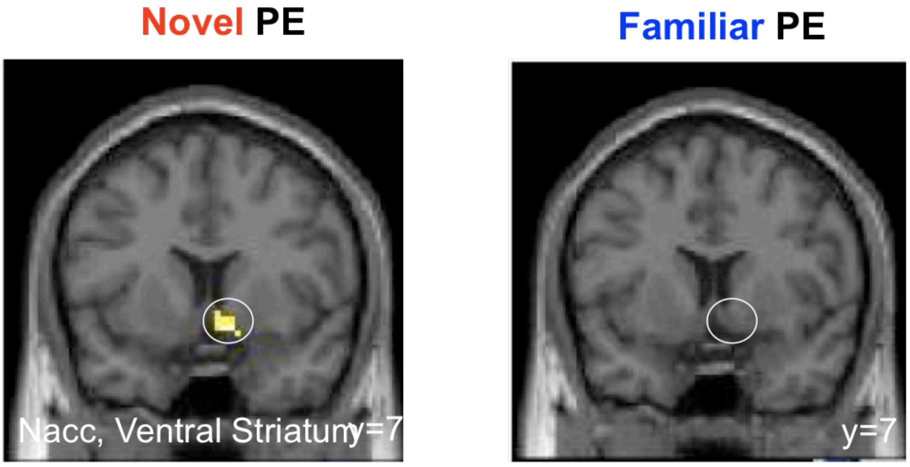


Figure 9.3 Whole brain analysis showed that there is significant prediction error in ventral striatum for novel stimuli on the left panel (p<0.001, uncorrected) but no significant prediction error activity for familiar stimuli (right panel) even with a more liberal threshold of (p<0.0001, uncorrected).

Table 9.1 Activation coordinates for the Novel and Familiar Prediction Errors (p<0.0001, uncorrected)

Novel PE

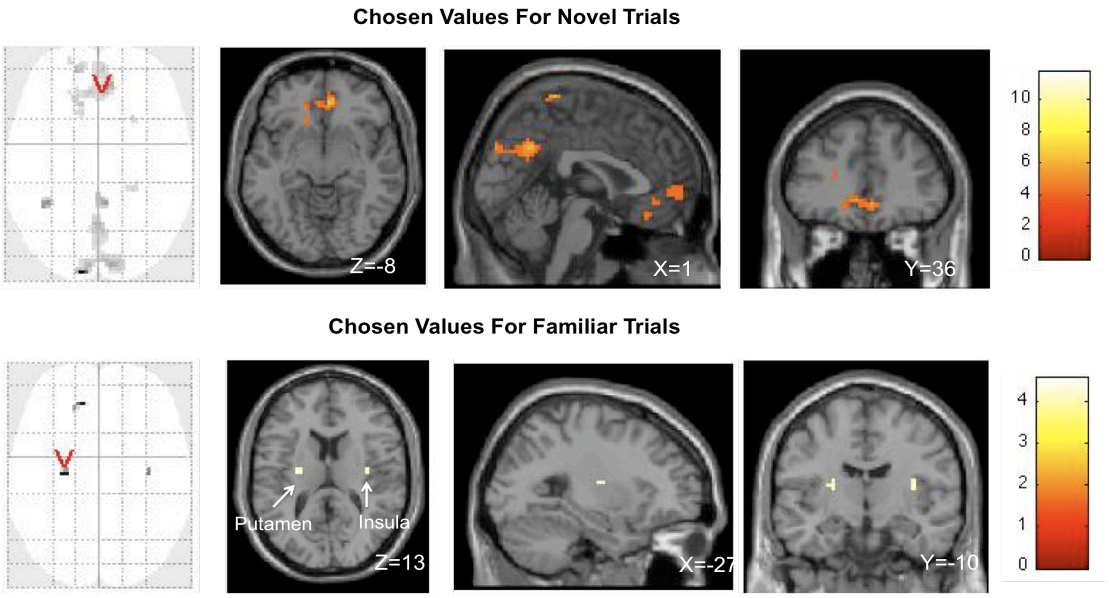
				MNI Coordinates		
Regions of Activation (BA)	Laterality	Cluster Size	x	y	z	T-Value (Peak)
Middle Temporal Gyrus		5	-66	-16	-17	5.78
Inferior Frontal Gyrus		9	-30	29	-17	7.34
Temporal Lobe		6	-51	-37	-14	6.4
Caudate	R	23	12	8	-11	7.38
Lingual	R	5	18	-67	-11	5.46
Lingual Gyrus		5	9	-73	-8	5.24
Anterior Cingulate		10	-6	41	-5	5.44
Middle Frontal Gyrus	R	116	36	47	7	7.9
Middle Frontal Gyrus	R	139	-9	56	7	7.01
Cingulate Gyrus		37	-3	-28	40	7.11

Familiar PE

				MNI Coordinates		
Regions of Activation (BA)	Laterality	Cluster Size	x	y	z	T-Value (Peak)
No Significant Activation						

9.2.2 Neural Correlates of Predicted Values for Chosen Options For Novel and Familiar Stimuli

We next tested one of our core hypotheses, which is whether the activation for predicted values differ for highly trained familiar stimuli and novel stimuli. To identify brain regions whose activation was modulated by predicted value, we regressed our predicted value signal onto the BOLD data. BOLD responses correlating with the model-derived predicted value for novel stimuli were mainly found in the VMPFC (peak at, x=1 y=36, z=-8), and precuneus (x=0, y=-58, z=34). On the other hand, the activation for the predicted value of the chosen option for the familiar stimuli were found in the left putamen (x=-27, y=-7, z= 13) and right insula (x=36, y=-13, z=10) (see **Figure 9.4**).



Chosen Values For Novel Trials

Chosen Values For Familiar Trials

Figure 9.4 Activations correlating with predicted value for the novel and familiar stimulus conditions. Top panel: results of the predicted value signal for the novel stimuli for the chosen actions derived from our computational model correlating with the medial frontal cortex and precuneus (p<0.001, uncorrected). Bottom panel: results of the predicted value signal for the familiar stimuli whose activation show significant changes in putamen and insula (p<0.001, uncorrected).

The findings so far support the conclusion that the BOLD signal in the mOFC is correlated with goal values, but the signal indorsal striatum is correlated with habit values (see **Figure 9.4**). However, when we used a more liberal statistical threshold p<0.005 (uncorrected) we observed that a region in the medial OFC (x=0.6, y=56, z=10) also shows significant activation for the familiar predicted values (see **Figure 9.5a**). In order to verify whether this activity is specifcaly caused by the predicted values of expected outcomes, we looked at the percent signal change graphs. The results revealed that unlike putamen, medial OFC shows negative percent signal changes for predicted values for familiar stimuli. We also looked at novel predicted-values with a more threshold and find a similar activation pattern as the familiar predicted values. For the novel predicted values both bilateral putamen show significant activation but this average percent signal change is negative compared to average signal change in medial OFC (see **Figure 9.5b**).

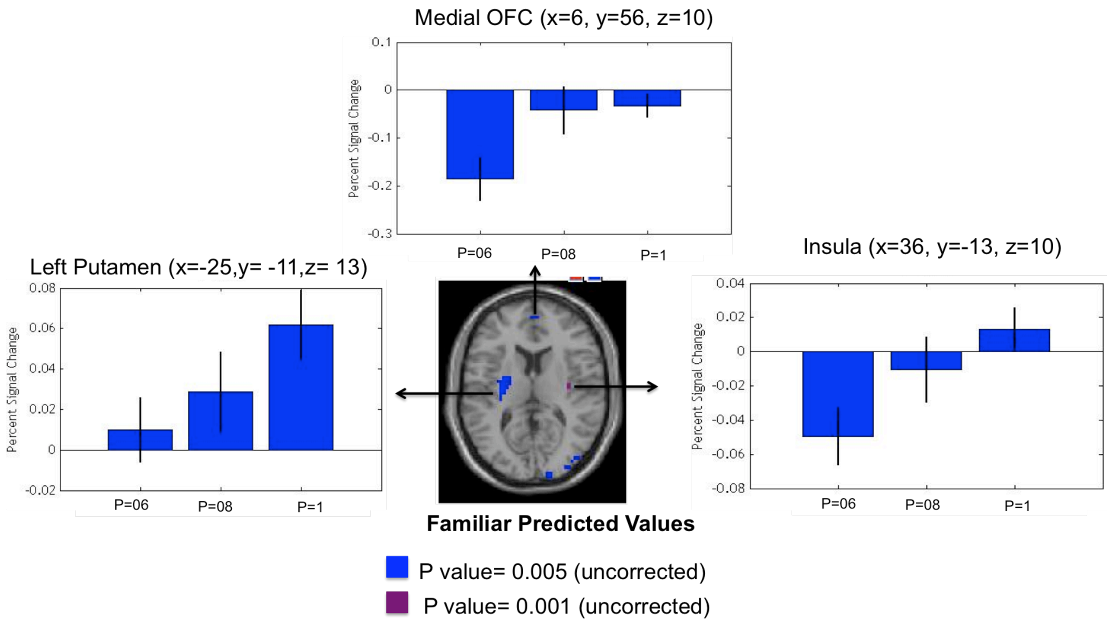


Figure 9.5a The middle picture showsactivations correlating with the predicted value for the familiar stimulus conditions. Results of the predicted value signal for the novel stimuli for the chosen actions derived from the computational model correlating with the medial frontal cortex ($p<0.005$, uncorrected), left putamen ($p<0.001$, uncorrect) and precuneus ($p<0.001$, uncorrected). Bottom panel: results of the predicted value signal for the familiar stimuli whose activation show significant changes in putamen and insula ($p<0.001$, uncorrected).

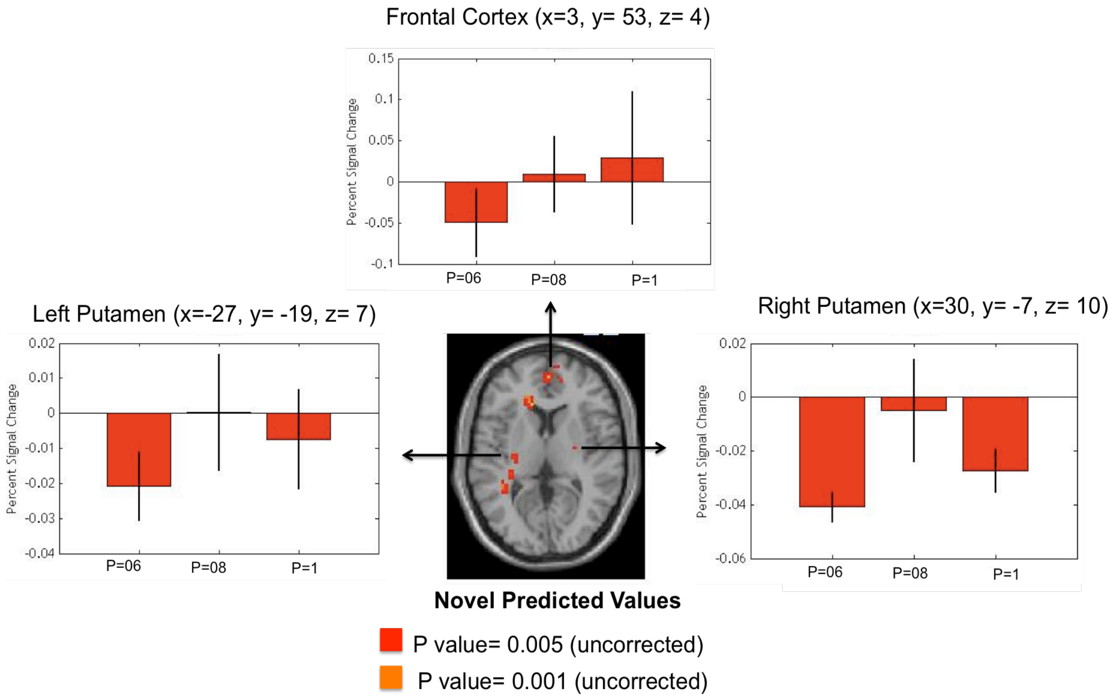
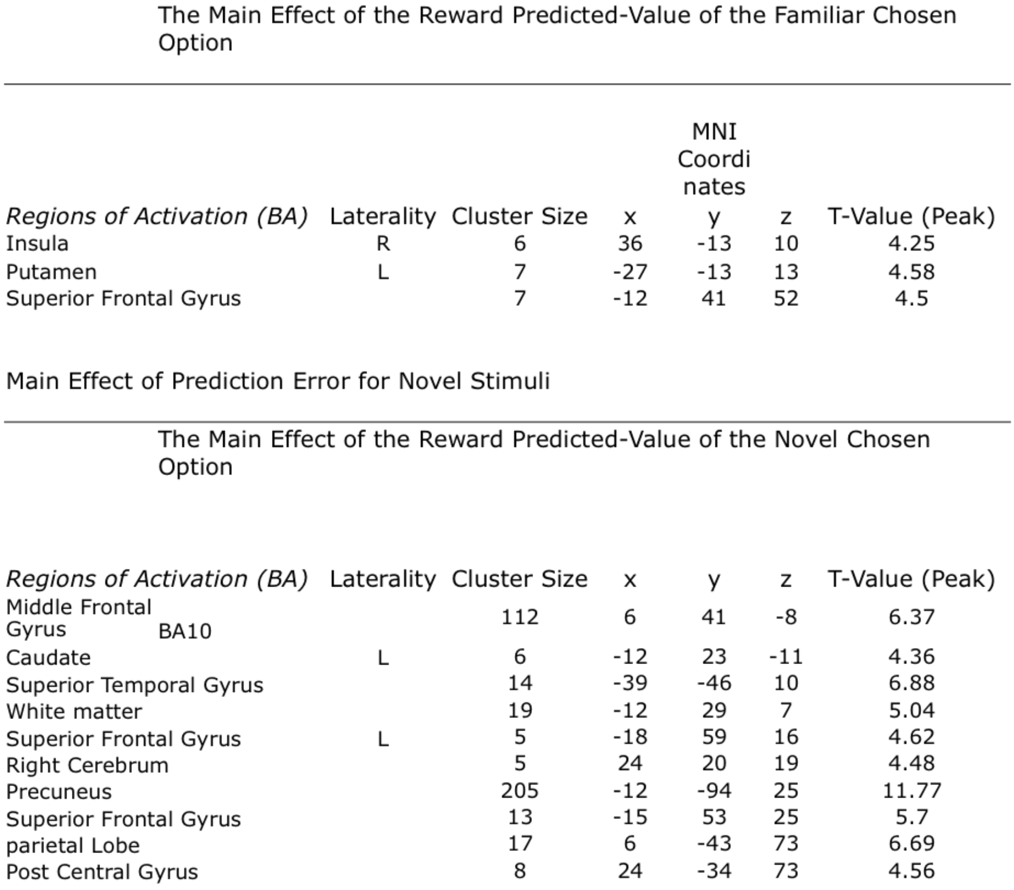


Figure 9.5b The middle picture shows activations correlating with the predicted value

for the novel stimulus conditions. Results of the predicted value signal for the novel stimuli for the chosen actions derived from the computational model correlating with the medial frontal cortex (p<0.005, uncorrected), left putamen (p<0.001, uncorrected) and precuneus (p<0.001, uncorrected). Bottom panel: results of the predicted value signal for the novel stimuli whose activation show significant changes in putamen and insula (p<0.001, uncorrected).

Table 9.2 Activation coordinates for the Novel and Familiar Predicted Values (p<0.001, uncorrected).



The Main Effect of the Reward Predicted-Value of the Familiar Chosen Option

				MNI Coordinates		
Regions of Activation (BA)	Laterality	Cluster Size	x	y	z	T-Value (Peak)
Insula	R	6	36	-13	10	4.25
Putamen	L	7	-27	-13	13	4.58
Superior Frontal Gyrus		7	-12	41	52	4.5

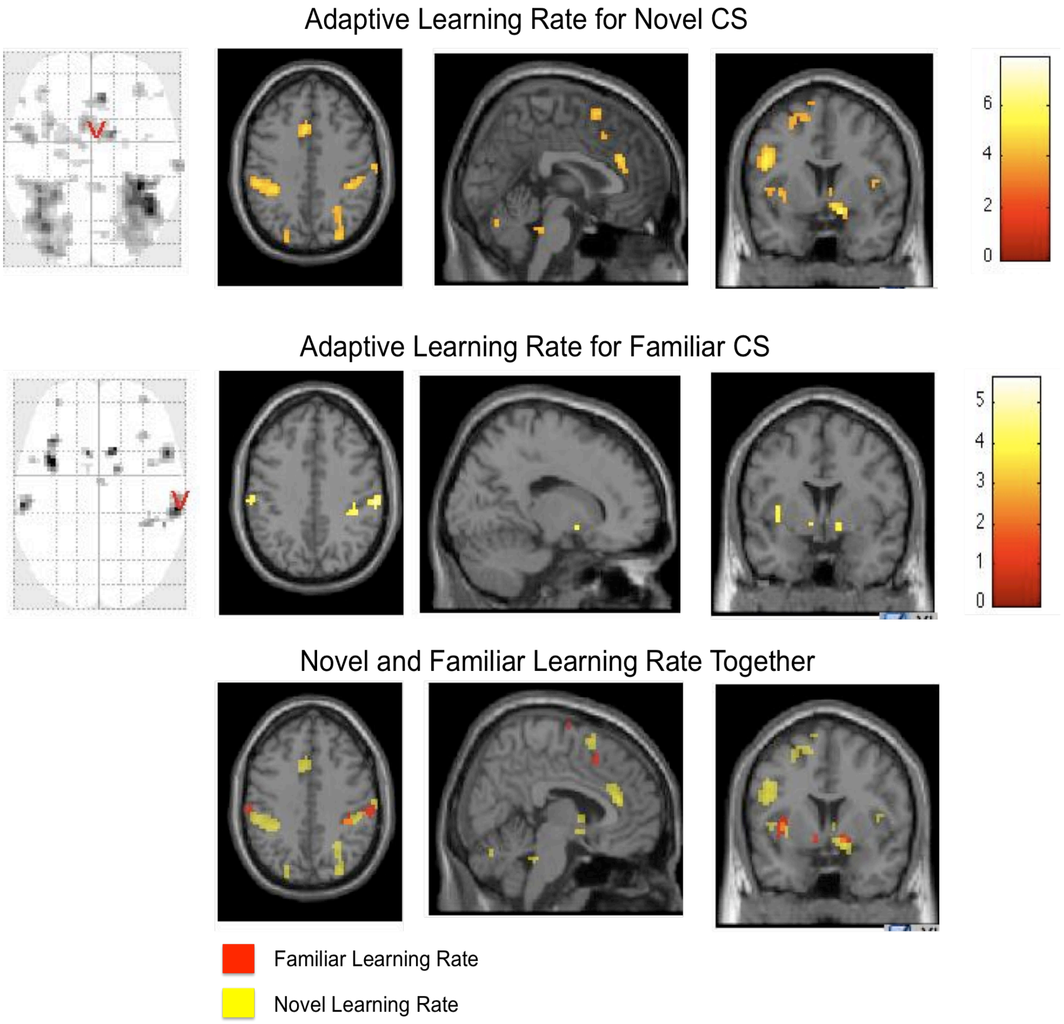
Main Effect of Prediction Error for Novel Stimuli

The Main Effect of the Reward Predicted-Value of the Novel Chosen Option

Regions of Activation (BA)	Laterality	Cluster Size	x	y	z	T-Value (Peak)
Middle Frontal Gyrus BA10		112	6	41	-8	6.37
Caudate	L	6	-12	23	-11	4.36
Superior Temporal Gyrus		14	-39	-46	10	6.88
White matter		19	-12	29	7	5.04
Superior Frontal Gyrus	L	5	-18	59	16	4.62
Right Cerebrum		5	24	20	19	4.48
Precuneus		205	-12	-94	25	11.77
Superior Frontal Gyrus		13	-15	53	25	5.7
parietal Lobe		17	6	-43	73	6.69
Post Central Gyrus		8	24	-34	73	4.56

9.2.3 Neural Correlates of Adaptive Learning Rates

Previous studieswhich utilized adaptive learning rate models in their imaging studies (e.g. Behrens et al., 2007; Haruno & Kawato, 2006). For example, Behrens et al., (2007) correlate model derived learning rateswith BOLD signal at the point of reward outcome and describe the brain activations as indicating the level of environmental volatility. In the current study however, we looked at the neural correlates of learning rates during the presentation of novel and familiar conditional stimulus. We found that during the presentation of novel CS activity in certain regions of the cingulate cortex, dorso-lateral prefrontal cortex, bilateral supplementary motor area as well as in occipital cortex (see **Table 9.3** for a full list of activations). Moreover when we looked at the adaptive learning rates for the familiar stimuli, we found a similar activation pattern in the ventral striatum and the parietal cortex but no occipital cortex activity (see **Figure 9.6**).

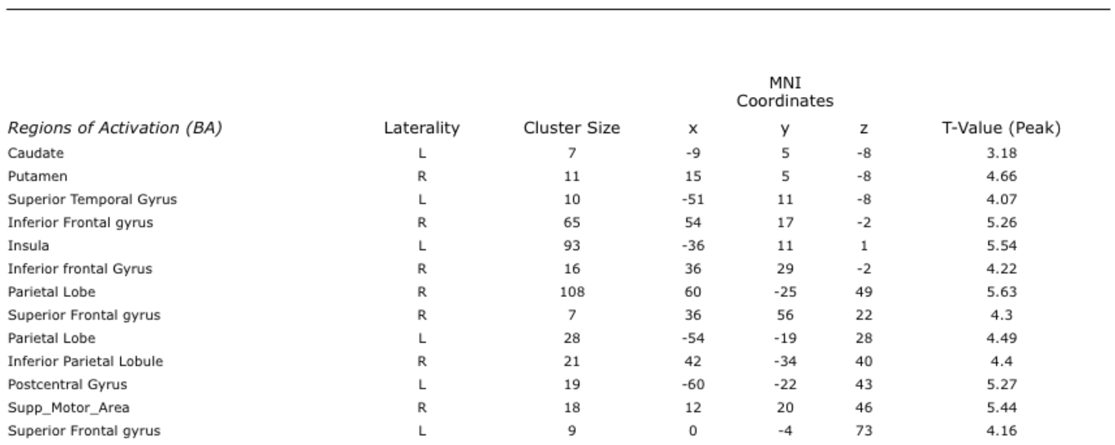

Adaptive Learning Rate for Novel CS

Adaptive Learning Rate for Familiar CS

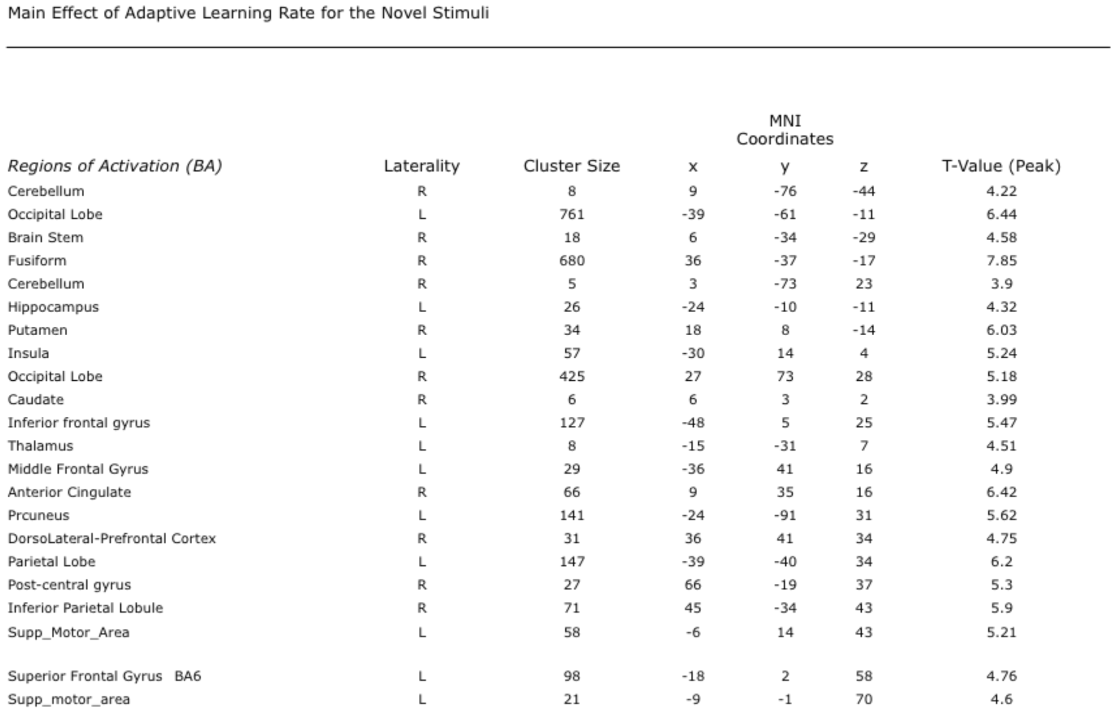
Novel and Familiar Learning Rate Together

Figure 9.6BOLD activation related to learning rate during CS presentation based on stimulus familiarity. On the bottom figureNovel and Familiar contrasts together overlapped on the same single subject anatomical image. Red regions show BOLD activation for adaptive learning rates for Familiar CS. Yellow regions show adaptive learning ratesfor Novel CS.

Table 9.3 Activation coordinates for the Novel and Familiar Adaptive Learning Rates (p<0.001, uncorrected).



Regions of Activation (BA)	Laterality	Cluster Size	MNI Coordinates x	 y	 z	T-Value (Peak)
Caudate	L	7	-9	5	-8	3.18
Putamen	R	11	15	5	-8	4.66
Superior Temporal Gyrus	L	10	-51	11	-8	4.07
Inferior Frontal gyrus	R	65	54	17	-2	5.26
Insula	L	93	-36	11	1	5.54
Inferior frontal Gyrus	R	16	36	29	-2	4.22
Parietal Lobe	R	108	60	-25	49	5.63
Superior Frontal gyrus	R	7	36	56	22	4.3
Parietal Lobe	L	28	-54	-19	28	4.49
Inferior Parietal Lobule	R	21	42	-34	40	4.4
Postcentral Gyrus	L	19	-60	-22	43	5.27
Supp_Motor_Area	R	18	12	20	46	5.44
Superior Frontal gyrus	L	9	0	-4	73	4.16



Main Effect of Adaptive Learning Rate for the Novel Stimuli

Regions of Activation (BA)	Laterality	Cluster Size	MNI Coordinates x	 y	 z	T-Value (Peak)
Cerebellum	R	8	9	-76	-44	4.22
Occipital Lobe	L	761	-39	-61	-11	6.44
Brain Stem	R	18	6	-34	-29	4.58
Fusiform	R	680	36	-37	-17	7.85
Cerebellum	R	5	3	-73	23	3.9
Hippocampus	L	26	-24	-10	-11	4.32
Putamen	R	34	18	8	-14	6.03
Insula	L	57	-30	14	4	5.24
Occipital Lobe	R	425	27	73	28	5.18
Caudate	R	6	6	3	2	3.99
Inferior frontal gyrus	L	127	-48	5	25	5.47
Thalamus	L	8	-15	-31	7	4.51
Middle Frontal Gyrus	L	29	-36	41	16	4.9
Anterior Cingulate	R	66	9	35	16	6.42
Prcuneus	L	141	-24	-91	31	5.62
DorsoLateral-Prefrontal Cortex	R	31	36	41	34	4.75
Parietal Lobe	L	147	-39	-40	34	6.2
Post-central gyrus	R	27	66	-19	37	5.3
Inferior Parietal Lobule	R	71	45	-34	43	5.9
Supp_Motor_Area	L	58	-6	14	43	5.21
Superior Frontal Gyrus BA6	L	98	-18	2	58	4.76
Supp_motor_area	L	21	-9	-1	70	4.6

Furthermore, in order to look for the effect of environmental volatility on the learning rate we analysed the data regardless of stimulus novelty but categorized it according to outcome uncertainty. We found that during the presentation of high uncertainty, stimulus regions of the dorsal cingulate cortex (pre-SMA) are activated and when the outcomes become predictable (i.e. low uncertainty) then regions of anterio-medial cingulate cortex get activated. Accordingly when the environment is more volatile the more caudal regions of the cingulate cortex are involved, and when the outcomes are predictable the activity occurs in the anterior regions of cingulate cortex.

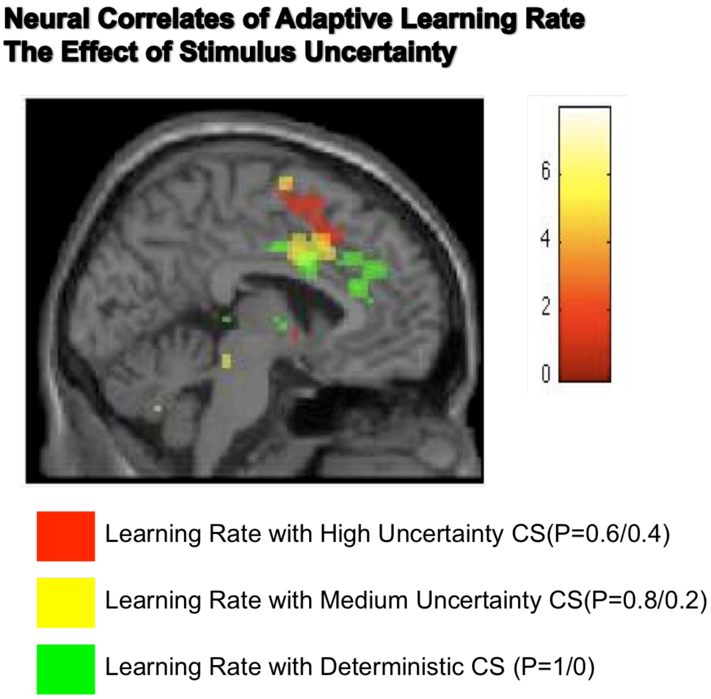

Neural Correlates of Adaptive Learning Rate
The Effect of Stimulus Uncertainty
6
4
2
0
🟥 Learning Rate with High Uncertainty CS(P=0.6/0.4)
🟨 Learning Rate with Medium Uncertainty CS(P=0.8/0.2)
🟩 Learning Rate with Deterministic CS (P=1/0)

Figure 9.7 BOLD activation related to learning rate during CS presentation with different reward probabilities regardless of their familiarty. The red regions show BOLD activation when the probability of receiving a reward is highly uncertain. Yellow regions show learning rate when probability of receiving a reward has mid-uncertainty. Green regions show BOLD activation for learning rate when probability of receiving a reward is deterministic.

9.3.Discussion

9.3.1 Novel versus Familiar Predicted Values

We found that distinct brain regions track distinct aspects of predicted value: VMPFC tracks the predicted value when the stimuli are completely novel that is when the outcome of the CS is still unpredictable. The putamen tracks the predicted value of the outcome when the outcomes of each CS are familiar to the participant who therefore knows which option to choose. Such a result provides strong evidence that VMPFC is involved early in learning but putamen is involved later in learning.

Given the results of this chapter and those of **Chapter 7** these collectively suggest that medial-orbito-frontal cortex is involved in the earlier goal directed phase of the learning where the predicted values of the rewarding outcomes still unknown (novel) but crucial. On the other hand when the participants know the actual predicted value of choosing a particular option and execute their selection in automatic fashion putamen is involved in coding the predicted value perhaps in the form of habits.

9.3.2 Novel versus Familiar Prediction Error

According to reward-prediction error models of learning, when the animal fully learns the predictive relationship between a conditional stimulus (CS) and a reward, the CS becomes the predictor of reward, and the burst of phasic dopaminergic firing for prediction error no longer occurs for the expected reward outcome. In supporting of this theory we could not identify any significant activation for prediction error in the ventral striatum for the familiar stimuli but we did find significant prediction error activation in the ventral striatum for novel stimuli (see, **Figure 9.3**). This result suggests that the ventral striatum has an important role in behavioral learning guided by reward prediction error, when stimuli are novel.

9.3.3 Neural Correlates of Adaptive Learning Rates

Learning rate is a fundamental feature of the behaviour of all organisms and even for artificial agents. In reinforcement learning theory, adaptive learning rates are expected to reflect synaptic plasticity responsible for behavioral change (Schultz, 1998), which is a product of the reward prediction error and learning rate. Learning rate alone determines how fast an individual adapts itself to new behavioural contingencies. The results of our imaging study showed that activity for the adaptive learning rates increased activation in ventral striatum and cingulate cortex for both novel CS and familiar CS. Moreover, motor cortex activation was also found for both familiar and novel learning rates. In addition to that we found activity in DLPFC for the adaptive learning rates for Novel

CS.We think that change in activity in the motor cortex might be due to behavioural learning via changing the sensori-motor cortico-striatal plasticity for actions, which may be linked to the learning rate parameter that determines the impact of the prediction error on the motor structures.

Furthermore, studies have shown that dopamine neurons code for adaptive learning rates, and these in control the gating mechanism of the top-attention in working memory in DLPFC (see for a discussion **Chapter 3**). According to this proposal, novelty signals (in our computation model the novelty signal is modelled by the adaptive learning rate) that rise from DLPFC or the midbrain dopaminergic system might modulate the neural mechanisms in the basal ganglia and anterior cingulate cortex to controlthe speed of learning for the predicted-values. Furthermore, the adaptive learning rate can modulate the attention networks through bidirectional communication and can influence the activity cortico-striatal spiral loop for the anterior-posterior shift.

Finally the neural correlates of adaptive learning rate in our study show similarities with the study of Behrens et al. (2007), who showed that variation in learning rate (or volatility) correlated with fMRI signal in the rostral cingulate cortex. Thus, when we categorized the learning rate according to the level of reward uncertainty rather that stimulus familiarity we found that distinct regions of cingulate cortex are activated for different levels of uncertainty where high level uncertainty is correlated with dorsal part of the posterior medial cingulate cortex (pMCC) and Pre-supplementary motor area (Pre-SMA) whereas medium level of uncertainty correlated with more ventral parts of the posterior medial cingulate cortex and completely determisitic stimuli is correlated with antero-medial cingulate cortex (aMCC). It is important to note that the posterior cingulate structures including pre-SMA are more sensitive to conflict resolution where as aMCC is more involved in outcome related activity such as inference based on previous outcomes (Botvinick et al., 2001; 2004; Pearson et al., 2011). Various lines of evidence suggest that Pre-SMA is also involved in pro-active switching in response to cues in order to facilitate new procedure after the particular action is becomes habitual or automatic (Hikosaka and Isoda, 2010). These results suggest that along the cingulate line caudal structures are involved in high volatile environments and rostral regions involved in deterministic environments.

Chapter 10

10 General Conclusions

10.1 Summary of the General Findings

The initial aim of this thesis was to establish the neural correlates of rewards and punishments in humans and how they influence learning related changes in the brain. Particular attention was given to the potential value of computational models as useful explanatory tools in order to achieve this aim. Whilst collecting and analysing the data from the first experiment it became apparent from my own empirical findings as well as those being published by others that the initial aim required some revision to bear in mind different patterns of involvement of the human valuation system in novel as opposed to familiar situations.

During the last two decades, psychological experiments have provided a wealth of data, which for learning paradigms consist of time-series of behavioural outcomes. In certain circumstances interpretation of these data requires computational models (e.g., Rescorla-Wagner Model), often involving additional, mostly hidden, variables (e.g., Rescorla-Wagner prediction error). Instrumental and Pavlovian learning in humans and animals are two of the most studied, and best understood, processes in experimental psychology and neuroscience and perhaps for this reason they have been used widely in computational modelling studies (see for a review, Montague, Hyman & Cohen, 2004;

Montague, King-Casas & Cohen, 2006) where they provide a framework to computational modellers for studying the changes in the hidden variables during a learning process.

Based on this approach during this PhD, two fMRI experiments were designed and the results of these experiments were reported between the **Chapters 6** and **Chapter 9**. In some of these chapters including **Chapter 7** of Experiment 1 and **Chapter 9** of Experiment 2 computational models were used to further analyse the behavioural data in order to find the estimated values of hidden variables (reviewed as a methodological review in **Chapter 5**). It was hypothesized that those hidden values have important implications forinfluencing the associative learning processes (reviewed in **Chapter 3** and **Chapter 4**).

In **Experiment 1**, the brain activations during the *anticipation* and *outcome* periods for monetary *gains* and *losses* were examined.

In the case of anticipation activation was found mainly in the cingulate cortex, the medial prefrontal cortex and the basal ganglia and in fact all of these structures showed increased BOLD response not only during the anticipation of reward outcomes but also for the anticipation of monetary losses. Moreover subtraction analyses for the anticipation contrast (reward anticipation - neutral anticipation & loss anticipation - neutral anticipation) showed that both reward and loss anticipation cause activity in the striatum and lateral frontal cortex. In the case of outcomes strong bilateral positive activity was found in the insular cortex for loss outcomes only since the difference betweenloss outcomes that are punished (-1 pound) andthe neutral outcomes (0 pound) was significant but a similar contrast between gain and neutral outcomes was not statistically significant in the insula. Activation was also found in amygdala, but this region appears to code not only for the difference between loss outcomesthat are punished (-1 pound outcomes only) but also gain outcomes that are rewarded (+1 pound outcomes only).

Considering both these results for anticipation and outcome it would seem that gains and losses might be processed by the same neuro-circuitry during the anticipation period (where decision making takes place) but processed differentially during the outcome period. However, it is important to note that the neural circuitry that was shared by reward and punishment expectation might be involve in processing action requirements rather than motivational valence. Perhaps this is due to the task design which does not separate action preparation from stimulus evaluation. For example, in a recent study Guitart-Masip et al., (2011) showed that ventral striatum and partially dorsal striatum is involved in coding go responses in a go-nogo task independent of the outcome valence suggesting that these regions might be involved in action opponency rather than motivational opponency. In order to understand these results better, we turned to computational models and calculated model-based analysis, which involved fitting data, derived from computational models.

Model based analysis of expected value and prediction error revealed that both reward expectation and prediction errors (the model used assumed that prediction error is calculated at outcome) are coded in the basal-ganglia specifically in the ventral-striatum but loss prediction error is coded in the dorsal striatum.

The results of Experiment 1 also included findings which suggested that functional organization in human brain differs for novel and familiar stimulus processing wherenovel stimuli use goal-directed or deliberate processes which correlate with the fronto-cortical activation (due to stimulus and action novelty) but the familiar stimuliuse more automated processes which correlate with the sensori-motor regions of the striatum and the cortex. This led to an extensive review of the literature on both the cognitive psychology of goal directed and automatic processing as well as computational models of these different systems. These were reviewed in **Chapter 3**. In Chapter 3 it was argued that the benefit of this shift from controlled to automated processing might be that automated processing frees attentional resources that are allocated to certain stimuli or actions, thus making multi-tasking plausible (Passingham & Rowe, 2002). Following this review the data from Experiment 1 was re-analysed using a model-based analysis (for the neural correlates of predicted value and prediction error) in order to test how the brain activities for associative learning change for the familiar and novel stimulus-reward associations. The method of analysis involved a comparison of the first and the last ten trials since the behavioural data suggested that this might separate goal based learning and automated processing. During the analysis we used separate ROIs, for the ventro-medial frontal cortex and sub compartments of human basal ganglia, which are assumed to serve different cognitive functions. The results showed that predicted value of the reward outcomes is coded in the medial frontal cortex during the first ten trials (where the learning is more goal directed) but during the last 10 trials (where the learning is more habitual) the activation was found in bilateral putamen. In contrast for the predicted value of loss outcomes we could notidentify any significant activation in the predetermined ROIs during early learning but for the last 10 trials we found significant activity in the internal segment of the globus pallidus (Gpi). For the reward prediction error we found significant activity in the ventral striatum during the early trials but we could not find any significant activation during the late learning trials. For the punishment prediction error, we found activity in the head of the caudate nucleus during early in learning but no significant activity during late learning trials. It has been argued that the most important difference between reward and punishment prediction error is that the reward prediction error is involved in ventral striatum where as the punishment prediction error is involved in dorsal striatum. With the model-based analysis to **Experiment 1,** we concluded that there is a shift of activation from anterior to posterior brain regions (from ventro-medial frontal cortex to putamen) for the predicted-values of reward outcomes.

However, even though there is a vast amount of electrophysiological and fMRI evidence (as reviewed in **Chapter 3**) showing correlations of predicted-value signals in cortical and limbic areas it wasn't clear how these values are affected by the stimulus novelty because we only performed model-based predicted value analysis to compare early and late learning trials. Given these findings of Experiment 1, a new experiment is designed with a more powerful experimental design to evaluate brain systems involved in processing novel versus familiar stimuli. **Experiment 2** primarily differed from Experiment 1 by comparing overlearnt stimuli, presumably highly familiar, with stimuli, which were novel for all presentations. This experiment therefore looked at the difference in processing familiar and novel stimulus by pre-training the participants with a set of stimuli until overlearnt and intermixing those with a set of novel stimuli during scanning. Using these results together with connectivity analysis the results of Experiment 2 showed, similar to Experiment 1 dorso-lateral pre-frontal cortex involvement in coding novel-stimuli which appears, according to the connectivity analysis to modulate activity in the ventral striatum, amygdala and insula as revealed by the connectivity analysis. Further model based analyses that were performed in **Chapter 9** also suggest a similar activation pattern that was found in **Chapter 7** which showed that predicted values for novel stimuli activate ventro-medial frontal structures but predicted values for the familiar stimuli activate dorsal parts of the striatum.

10.2 Summary of the Main Contributions

In this thesis, I have particularly focused on the neural correlates of predicted values and learning rates. It was suggested that neural correlates of learning rates might be a proxy measure for dopamine modulation, and it might promote neural transitions between mechanisms regulating goal directed and habitual learning by integrating information from the regions that are specialized in attentional processes. It was suggested that two aspects of reinforcement learning, namely "the adaptive learning rates" and "the predicted-values", are crucial for understanding the transition from momentary simple decisions to long-term habits. The functional imaging data presented here provides further evidence regarding certain aspects of this neural transition. However, one should be cautious that the stated model in this thesis on the interpretation of predicted-values and adaptive learning rates does not provide a comprehensive explanation of extreme habits like addiction, which would require further investigation with a devaluation paradigm.

Main contributions of the findings to Cognitive Neuroscience can be summarized as follows:

- The findings add to the existing pharmacological, electrophysiological, and functional imaging literature regarding the involvement of the striatum in aversive processing during *anticipation* of monetary losses (as reported in **Chapter 6**).

- The findings have important implications for understanding monetary loss *outcomes* becausethe activations found in the insula for monetary losses show overlapping regions with the existing imaging studies that looked at the phenomenological aspects of pain processing (as reported in **Chapter 6**).

- The findings suggest that both avoiding a loss outcome and getting a rewarding outcome activate similar regions in the medial frontal cortex but to a differential degree. On the other hand the general opponency relationship between gains and losses suggests that processing of financial losses need additional activation of bilateral amygdala (as reported in **Chapter 6**).

- The results add to the growing body of neural and psychological data supporting the biological basis of prediction error theory. They showed that reward prediction error is coded in ventral striatum and punishment prediction error is coded in dorsal striatum (as reported in **Chapter 6-7**). Furthermore, the results showed that prediction error is only involved early in learning, which validates that it as a learning signal.

- The results showed that there is a shift of activation for the neural substrates of predicted-value in the rostral-caudal axis during the automatization of the decision process for the predicted values of chosen stimuli (**Chapter 7 and Chapter 9**).

- The results showed that novel stimuli induce significantly greater BOLD activation in the dorso-lateral frontal cortex (DLPFC) and posterior parietal cortex (PPC) compared tofamiliar stimuli. As a further implication, we argued that increased activity in the bi-lateral DLPFC might explain the effects of stimulus novelty and novelty-induced-attentional changes in ventral striatum, amygdala and insula (**Chapter 8**).

- Finally the results showed that cingulate cortex is involved in coding the adaptive learning rates that are based on the sign of prediction error (positive or negative). This result suggests that cingulate cortex might be involved in the acquisition of predicted-values and might control the relative-amount of shift of activation from anterior to posterior regions (**Chapter 9**).

The results reported here might also have implications for other closely-related disciplines such as cognitive robotics and cognitive science in general which are summarized below:

- In cognitive science, there has been a long-going philosophical debate about the content of mental representations (Clark, 1997) including the type of information they store (Putnam, 1988), and the way they are encoded by the mind/brain (Churchland & Sejnowski, 1992). Keijzer (2001, p.2) defines mental representations as "theoretical entities that are the bearers of meaning and the source of intentionality". According to this definition, mental representations carry relevant information for an individual about the outside world they are interacting with. Then, the question becomes what type of information predicted-values carry and how individuals use that information during the process of learning. The first fundamental problem an individual has to solve in an experimental setup like the ones presented in **Chapter 6** and **Chapter 8** is to make an identification of the options to choose from displayed on the screen. Perhaps this question is partially solved by the visuo-sensory system with the guidance of dorsal and ventral visual pathways which process, respectively, what certain objects are and where they are located in space (Cisek & Kalaska, 2010). According to traditional cognitive models, the individual evaluate each option and compare their utilities in a serial manner (Shafir & Tversky 1995). Then, by using these representations, they make decisions on which actions to choose. However this approach has difficulty in explaining the interaction between different mental representations such as the transition between perception and action or between action and value representations. These decision-making models are usually slow due to their serial information processing capacity (Sternberg, 1969, 1977). In the current thesis, the reinforcement learning model and the brain imaging data suggest that participants make a value-based comparison among options and that the neural activity shifts from the brain regions mostly associated with cognitive-emotional processing toward those regions that are associated with motor-related processing. This suggests that the same decision variable elicits the activation in multiple brain regions, which depends on the individuals' familiarity with the stimuli.

- A second implication of the predicted-value coding is related to common representation framework in cognitive science (Hommel, 2001). This conceptual view is called event-coding framework which tries to connect the linkage between (late) perception and (early) action (Hommel, 2004; 2010). In the event–coding framework the perception and action could be based on a common representational domain (e.g, a neuron coding both motor features and perceptual features of a task) and it is different from other representational schemes where it assumes that the mental representations for stimulus and action are not coded separately but together in a common representational code (Hommel, 2001). The common representation view of Hommel et al. (2001) is in fact very similar to the action-value concept in reinforcement learning theory because it assumes that the same neurons (or brain regions) responds both to the visual properties of an object and the action features of an object. However, the most significant difference between the common-coding framework and value-based action coding in reinforcement learning is that in reinforcement learning there is no assumption about how predicted-value coding neurons respond to visual properties of a perceived object but rather the question is about the relationship between the motivational properties of an object (e.g., conditional stimulus) and its' action requirements (e.g., reaching movement). Perhaps common coded representations are used in decision-making but further investigation is needed in order to understand how the values are attributed to these common representations (for a discussion, see Cisek & Kalaska, 2010).

- Another implication of the findings reported in this thesis might be related to advancements in cognitive robotics. A cognitive robot should be fluent in routine operations and it should be capable of adjusting its behavior when it's faced with an unexpected situations (Kawamura et al., 2005; Ratanaswasd et. al., 2005). In order to maintain adaptability in complex environments, robots should be cabaple of controling their actions in order to select and focus important information throughout their task executions. These abilities are known to exist in humans as executive functions, and are usually studied under the title of cognitive control (Cohen et al., 1990). Cognitive control observed in humans is thought to be useful for a cognitive robot during the action-selection process as well as learning as it guides the robot through the search for component behaviors that might be combined and used efficiently to execute routine tasks as well as to behave in novel situations (Ratanaswasd et. al., 2005). Several paradigms in artificial intelligence (AI) and robotics have been developed in order to explain the behavior of an agent in terms of stages of perception and action

(Kadihasanoglu, Erdeniz, Kucuktunca, 2007). In neuroscience and robotics this is commonly referred to as the perception-action cycle (Cutsuridis et al., 2011; Fuster, 2003). According to the reinforcement learning models used in this thesis when the agents learn the predicted values, there is no further need for calculation of the predicted values and a hypothetical robot should select actions based on its policy. During the last decade reinforcement learning models that are similar to the ones used in this thesis have been successfully used in robotic projects (Peters et al., 2003; Lapko, 2007). It is possible that the fMRI results found in this thesis might lead to reinforcement learning models for future robotic applications which consider the involvement of functional neuroanatomy of cortico-striatal loops. Such robotic applications might show big performance increases in situations, where the task requirements for a robot need efficient sensori-motor control for routine task sets and high-level cognitive control for novel task sets.

10.3 Challenges and Limitations

One of the most important limitations is the total number of participants in the first experiment. As reported in **Chapter 6** data was collectedfrom 15 participants and before that we had piloted with two participants. Unfortunately, we could not include 2 participants from the 15 participants because most of the button press onset times were lost for those two participants (a software error). We thought that the problem was related to the synchronization of the scanner with the Superlab software. Considerable effort was put into finding the cause of the problem and how to program the experimental software (superlab) to avoid this problem in subsequent experiments. This involved much support and assistance from the technicians in the School of Psychology.

In the end the technicians contacted the suppliers of Superlab who acknowledged that a feature of the experimental design could not be accommodated. Thus we switched to E-Prime in the second experiment, which resolved the problem.

Secondly in several chapters of this thesis the contributions of computational models to our understanding of the brain function were reviewed. It was suggested computational models of reinforcement learning are good theoretical guides where they provide evidence for "where", "when" and "how" in the brain certain variables (e.g., action-value, prediction error) are represented (O'Doherty et al., 2007; Corroda & Doya, 2007; Mars et al., 2010). This approach might lead psychologists/neuroscientists to study cognitive/psychological phenomena, through specifying essential structures, divisions of modules, and relations between variables in the brain and used very commonly as common approach in cognitive science. As it was reviewed in **Chapter 3** and **Chapter 4** it is easy to find a large body of literature of fMRI studies which used computational models in their data analysis. However, one might ask the question such as which reinforcement algorithm provides the best modelling framework to study the effect of novelty in associative learning? The answer to this question is not easy, and it is possible that there might be more than one computational model that is suitable for modelling a particular task such as the one used in **Chapter 9.** For example, there are alternative adaptive learning rate models than the one used in **Chapter 9**. One such model is Kalman Filter, which captures the essential relationship between the prediction error and learning rate fairly well (Daw et al., 2006). In addition to that several reinforcement learning models might provide solutions as good as adaptive learning rate models, which include the heterarchical reinforcement learning model of Haruno & Kawato (2006), the attention gated reinforcement learning model of Roelfsema & van Ooyen (2005), the model-based framework of Daw et al., (2005, 2011) and Dayan (2008), the hierarchical reinforcement learning model of Botvinick et al., (2009), actor-critic architectures (Barto, 1995; Joel et al., 2002), and reinforcement learning models based neural network (O'Reilly et al., 1999; Frank et al., 2001). Given that how the brain uses such models is still not well understood, where much theoretical and empirical research in this are needs further research.

9.2 Future Research

Some ideas are suggested as extensions to my work are as follows:

Some fMRI chapters in this thesis used statistical parametric analysis based on the principles of computational models as predictors of BOLD signal changes under the framework of reward-related learning. A number of sub-cortical and cortical areas were found to be involved in learning of stimulus-response associations as well as predictions and it was therefore suggested that these activations might be modelling dopaminergic modulation. To our knowledge dopamine is the best candidate for explaining such complex learning processes in humans and animals. However, as we reviewed in **Chapter 5** the evidence that indicates a direct relation between BOLD signal and dopamine is not yet strong enough to make a direct casual link. Therefore, in order to support the dopamine part of our hypothesis, we need direct measures of dopamine activity perhaps by using positron emission tomography (PET).

9.3 List of Publications

During the four years of my PhD, I have attended several conferences; I have presented 3 posters, and two full conference papers.

The following conference papers were presented in the conferences:

Erdeniz B., Done. J., Maex. R (2011) From Simple Decisions to Habits: Adaptive Coding of Action Values Through Multiple Cortico-Striatal Loops. International Cognitive Neuroscience Congress. Turkey/Marmaris

Erdeniz B., Done. J Davey. N Frank. R Maex, R. (2009) Model Based fMRI: A novel technique for combining computational models and Brain Function. International Cognitive Neuroscience Congress. Turkey/Marmaris

The following posters were presented in conferences:

Erdeniz B., Done. J Davey. N Frank. R , (2012)Neural Correlates of Chosen Action Values Early Vs Late Learning SIXTEENTH INTERNATIONAL CONFERENCE ON COGNITIVE AND NEURAL SYSTEMS, Boston University, United States (Poster)

Erdeniz B., Done. J (2012) Neural Correlates of Chosen Action Values Early Vs Late Learning.Second Symposium on "Biology of Decision-Making", Paris, France 10-11 May, 2012 (Poster)

Erdeniz B., Done. J Davey. N Frank. R , Maex R. Annett L. (2009) Modeling Self Control Behavior in a Reinforcement Learning Environment using Quasi Hyperbolic Discounting. Dopamine Agonists. THIRTEENTH INTERNATIONAL CONFERENCE ON COGNITIVE AND NEURAL SYSTEMS, Boston University, United States (Poster)

The following papers are in preparation for submition:

(In Preparation) Erdeniz B., Done. J Davey. N Frank. R, Maex R. (2012) Neural Correlates of Stimulus Novelty During Associative Learning

(In Preparation) Erdeniz B., Done. J Davey. N Frank. R , Maex R. (2012) Neural Correlates of Opponent Processes for Financial Gains and Losses:The Role of Medial Frontal Cortex

(In Preparation) Erdeniz B., Done. J Davey. N Frank. R , Maex R. (2012) Adaptive Coding of Predicted Values, Early versus Late Learning.

Bibliography

Aarsland D., Alves G., Larsen J.P. (2005) Disorders of motivation, sexual conduct, and sleep in Parkinson's diseas*e. Adv Neurol 96:56-64.*

Abler B, Walter H, Erk S, Kammerer H, Spitzer M (2006) Prediction error as a linear function of reward probability is coded in human nucleus accumbens. *Neuroimage.* Jun;31(2):790-5.

Adcock RA, Thangavel A, Whitfield-Gabrieli S, Knutson B, Gabrieli JD (2006) Reward-motivated learning: mesolimbic activation precedes memory formation. *Neuron* 50:507–517.

Albin, R. L., Young, A. B. & Penney, J. B. (1989) The functional anatomy of basal ganglia disorders. *Trends Neurosci.* 12, 366–375.

Albin R.L., Mink J.W. (2006) Recent advances in Tourette syndrome research. *Trends Neurosci.* 29:175—182

Alcaro A, Huber R, Panksepp J (2007) Behavioral functions of the mesolimbic dopaminergic system: an affective neuroethological perspective. Brains Res Rev 56(2):283–321

Alexander, G. E., DeLong, M. R. & Strick, P. L. (1986) Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu. Rev. Neurosci.* 9, 357–381.

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

Amiez, C. Joseph, J.P. Procyk E. (2006) Reward encoding in the monkey anterior cingulate cortex *Cereb Cortex,* 16 pp. 1040–1055

Ande´n, N.E., Fuxe, K., Hamberger, B., and Hokfelt, T. (1966). A quantitative study on the nigro-neostriatal dopamine neurons. *Acta physiol. scand.*67, 306–312.

Anderson, J. R., & Lebiere, C. (1998). The atomic components of thought. Hillsdale, NJ: Lawrence Erlbaum Associates.

Antzoulatos EG, Miller EK. (2011) Differences between neural activity in prefrontal cortex and striatum during learning of novel abstract categories. *Neuron* Jul 28;71(2):243-9.

Apicella P., Ljungberg T., Scarnati E., Schultz W. (1991) Responses to reward in monkey dorsal and ventral striatum. *Exp Brain Res* 85:491–500.

Arkadir, D., Morris, G., Vaadia, E., & Bergman, H. (2004) Independent Coding of Movement Direction and Reward Prediction by Single Pallidal Neurons.Experimental *Brain Research,* 24(45), 10047-10056.

Armel, C. Beaumel, A. Rangel, A. (2008) Biasing simple choices by manipulating relative visual attention. *Judgment and Decision Making,* 3(5):396-403

Aron, A., Fisher, H., Mashek, D. J., Strong, G., Li, H. and Brown, L. L. (2005) Reward, motivation, and emotion systems associated with early-stage intense romantic love. *J. Neurophysiol.* 94, 327–337.

Ashby, F.G., Turner, B.O., Horvitz, J.C. (2010) Cortical and basal ganglia contributions to habit learning and automaticity. *Trends in Cognitive Science.*14(5), 191-232.

Ashby, F.G., & Crossley, M.J. (2010) Interactions between declarative and procedural-learning categorization systems. *Neurobiology of Learning and Memory,* 94, 1-12.

Ashby, F.G., Waldschmidt, J.G. (2008) Fitting computational models to fMRI data. *Behavior Research Methods,* 40, 713-721.

Ashby, F.G., & Valentin, V.V. (2007). Computational cognitive neuroscience: Building and testing biologically plausible computational models of neuroscience, neuroimaging, and behavioral data. In M. J. Wenger & C. Schuster (Eds.), *Statistical and process models for cognitive neuroscience and aging* (pp. 15-58). *Mahwah,* NJ: Erlbaum.

Asaad, W. F., Rainer, G. & Miller, E. K. (1998) Neural activity in the primate prefrontal cortex during associative learning. *Neuron* 21, 1399–1407

Aydinonat, E. & Erdeniz, B. (2008) Evolution of Money: Perspectives from Neurobiology and Economics**,** Neuroeconomics: Hype or Hope? EIPE Conference Report, Rotterdam, The Netherlands

Backman, L., Karlsson, S., Fischer, H., Karlsson, P., Brehmer, Y., Rieckmann, A., MacDonald, S.W.S., Farde, L., Nyberg, L., (2009) Dopamine D1 receptors and age differences in brain activation during working memory. *Neurobiol. Aging,* Oct;32(10):1849-56.

Badre, D., & D'Esposito, M. (2009) Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience,* 10, 659-669.

Balleine, B. W., Delgado, M. R., & Hikosaka, O. (2007). The Role of the Dorsal Striatum in Reward and Decision-Making. *Journal of Neuroscience*, 27(31), 8161- 8165.

Balleine B.W., Dickinson A. (1998) The role of incentive learning in instrumental outcome revaluation by sensory-specific satiety. *Anim. Learn. Behav.*;26:46–59.

Balleine, B.W., O'Doherty, J.P. (2010). Human and rodent homologies in action control: Cortico-striatal determinants of goal-directed and habitual action, *Neuropsychopharmacology,* 35(1):48-69.

Balderston NL, Schultz DH, Helmstetter FJ. (2011) The human amygdala plays a stimulus specific role in the detection of novelty. *Neuroimage.* 2011 Apr 15;55(4):1889-98.

Bayer, H. M. and P. W. Glimcher (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron 47(1):* 129-41.

Bar-Gad, I., Havazelet-Heimer, G., Goldberg, J.A., Ruppin, E., Bergman, H., (2000) Reinforcement-driven dimensionality reduction—a model for information processing in the basal ganglia. *J. Basic Clin. Physiol. Pharmacol.* 11, 305–320.

Bar-Gad I, Morris G, Bergman H (2003) Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. *Prog Neurobiol* 71:439–473.

Bartels A. and Zeki S. (2000) The neural correlates of romantic love. *NeuroReport* Vol. 17(11), p. 3829-3834

Bartels A. and Zeki S. (2004) The neural correlates of maternal and romantic love *NeuroImage* Vol. 21(3), p. 1155-1166

Barto, A.G., (1995) Adaptive critics and the basal ganglia. In: Houk, J.C., Davis, J.L., Beiser, D.G. (Eds.), Models of Information Processing in the Basal Ganglia. MIT Press, Cambridge, pp. 215–232.

Barraclough, D.J., M.L. Conroy & D. Lee. (2004) Prefrontal cortex and decision making in a mixed-strategy game. *Nat. Neurosci.*7: 404–410.

Baxter,M.G. and Murray,E.A. (2002) The amygdala and reward.*Nat.Rev.Neurosci.3*, 563-573.

Becerra V. M. (2008) Scholarpedia, 3(1) :5354.

Becerra, L. Breiter, H.C. Wise, R. Gonzalez, R.G Borsook D. (2001) Reward circuitry activation by noxious thermal stimuli *Neuron,* 32 pp. 927–946

Bechara, A. Damasio, A.R. Damasio, H. Anderson, S.W. (1994) Insensitivity to future consequences following damage to human prefrontal cortex, *Cognition* 50 7–15.

Bechara, A. Damasio, H. Tranel, D. Damasio, A.R. (1997) Deciding advantageously before knowing the advantageous strategy, *Science* 275 1293–1295.

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature* 10(9), 1214-1221.

Belin D, et al. (2009) Parallel and interactive learning processes within the basal ganglia: Relevance for the understanding of addiction. Behav. Brain Res.;199:89–102.

Bellman, R.E. (1957) Dynamic Programming.Princeton University Press, Princeton, NJ. Republished 2003: Dover,

Berger, B. and Gaspar, P. (1994) Comparative anatomy of the catecholaminergic innervation of rat and primate cerebral cortex. In: Phylogeny and Development of Catecholamine Systems in the CNS of Vertebrates, pp. 293-324. Eds. W. J. A. J. Smeets and A. Reiner. Cambridge University Press, Cambridge, UK.

Berns, G.S., McClure, S.M., Pagnoni, G., Montague, P.R. (2001) Predictability modulates human brain response to reward. *Journal of Neuroscience 21:*2793-2798.

Berridge, K.C. & Robinson, T.E. (2003) Parsing reward. *Trends in Neurosciences*, 26(9), 507-513.

Berridge, K.C. (2007) The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology*, 191, 391-431,

Berridge, K.C. & Kringelbach, M.L. (2008) Affective neuroscience of pleasure: reward in humans and animals. *Psychopharmacology*,199, 457-480.

Berridge, K.C., Ho, C.Y., Richard, J.M., DiFeliceantonio, A.G. (2010) The tempted brain eats: Pleasure and desire circuits in obesity and eating disorders. *Brain Research*, 1350: 43-64.

Berridge, K.C., Robinson, T.E. & Aldridge, J.W. (2009) Dissecting components of reward: 'liking', 'wanting', and learning. *Current Opinion in Pharmacology*, 9, 65-73.

Belin, D., Jonkman, S., Dickinson, A., Robbins, T.W., Everitt, B.J., (2009) Parallel and interactive learning processes within the basal ganglia: relevance for the understanding of addiction. *Behav. Brain Res.* 199, 89–102.

Bennett, C. M. Wolford, G. L. Miller M. B. (2009) The principled control of false positives in neuroimaging *Soc Cogn Affect Neurosci4 (4): 417-422.*

Ben-Sreti MM, Gonzalez JP, Sewell RD. (1983) Differential effects of SKF 38393 and

LY 141865 on nociception and morphine analgesia. *Life Sci.*; 33: Suppl. 1, 665-668.

Bevins RA & Besheer J (2005) Novelty reward as a measure of anhedonia. *Neuroscience and Biobehavioral Reviews*, 29, 707-714.

Bhatia KP, Marsden CD. 1994. The behavioural and motor consequences of focal lesions of the basal ganglia in man. *Brain.* 117(Pt 4): 859--876.

Bischoff-Grethe, A., Hazeltine, E., Bergren, L., Ivry, R. B., & Grafton, S. T. (2009) The influence of feedback valence in associative learning. *NeuroImage*, *44*(1), 243-251

Bizot J, Le Bihan C, Puech AJ, Hamon M, Thiebot M *(1999)* Serotonin and tolerance to delay of reward in rats.*Psychopharmacology (Berl) 146:400–412.*

Blood, A. J., R. J. Zatorre, et al. (1999) Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nat Neurosci* 2 (4): 382-7.

Blackford JU, Buckholtz JW, Avery SN, Zald DH. (2010) A unique role for the human amygdala in novelty detection. *Neuroimage.* Apr 15;50(3):1188-93.

Bolam, J.P. and Bennett, B.D. (1995) Microcircuitry of the neostriatum. In Molecular and CellularMechanisms of Neostriatal Function (Ariano, M.A. and Surmeier, D.J., eds), pp. 1–19, R.G. Landes Co

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S. & Cohen, J. D. Conflict monitoring and cognitive control. *Psychol. Rev.* 108, 624–652

Botvinick, M. M., Y Niv & A Barto (2009) - Hierarchically organized behavior and its neural foundations: A reinforcement-learning perspective. *Cognition* 113, 262-280.

Botvinick, M. M. (2008) Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences*, 12, 201-208.

Botvinick, M. & Plaut, D. C. (2006) Such stuff as habits are made on: A reply to Cooper and Shallice (2006), *Psychological Review,* 113, 917-928.

Botvinick, M., Cohen, J. D. & Carter, C. S. (2004) Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences.*8, 539-546.

Botvinick, M., Braver, T., Barch, D. Carter, C. & Cohen, J. (2001) Conflict monitoring and cognitive control. *Psychological Review*, 108 (3), 624-652.

Bouton, M. E. (2007) *Learning and Behavior: A Contemporary Synthesis*, Sunderland, MA: Sinauer

Bower G.H. (1994) A turning point in mathematical learning theory. *Psychol Rev.* Apr; 101(2):290-300.

Bor, D., Owen, A.M. (2006) Working Memory: Linking Capacity with Selectivity, *Current Biology,* 16(4), R136-138.

Braver TS & Cohen JD (2000). On the control of control: The role of dopamine in regulating prefrontal function and working memory. In Monsell S & Driver J (Eds.), Attention and Performance XVIII; Control of cognitive processes, pp.713-737.

Bray S, O'Doherty J. (2007) Neural coding of reward-prediction error signals during classical conditioning with attractive faces. *J Neurophysiol.* 97(4):3036-45

Breiter HC, Aharon I, Kahneman D, Dale A, Shizgal P. (2001) Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron 30:*619–39

Breiter HC, Etcoff NL, Whalen PJ, Kennedy WA, Rauch SL, Buckner RL, Strauss MM, Hyman SE, Rosen BR. (1996) Response and habituation of the human amygdala during visual processing of facial expressions. *Neuron;* 17:875–877.

Brett M, Anton JL, Valbregue R, Poline JB (2002) Region of interest analysis using an SPM toolbox 8th International Conference on Functional Mapping of the Human Brain, June 2–6, Sendai, Japan. Available on CD-ROM in *NeuroImage.* 16.

Brodmann, K (1909) Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues. Leipzig: JA Barth.

Bromberg-Martin, Ethan S. Masayuki Matsumoto, and Okihide Hikosaka (2010) Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron* volume 68, issue 5, pp. 815-834

Brown, J. (1955) Pleasure-seeking behavior and the drive-reduction hypothesis. *Psychological Review,* 62, 169-179.

Buhot MC *(*1997) Serotonin receptors in cognitive behaviors. *Curr Opin Neurobiol 7:243–254.*

Bush*,* R. R. and Mosteller*,* F. (1955) Stochastic models for learning. New York: Wiley

Bush, G., Luu, P. & Posner, M. I. (2000) Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn. Sci.* 4, 215–222

Bunge (2002) How we use rules to select actions: a review of evidence from cognitive neuroscience._*Cogn Affect Behav Neurosci.* 2004 Dec;4(4):564-79.

Bunge, S. A., Hazeltine, E., Scanlon, M. D., Rosen, A. C., & Gabrieli, J. D. (2002). Dissociable contributions of prefrontal and parietal cortices to response selection. *NeuroImage,* 17, 1562-1571

Bunzeck N, Schutze H, Stallforth S, Kaufmann J, Duzel S, Heinze H-J, Duzel E (2007) Mesolimbic Novelty Processing in Older Adults. *Cereb Cortex*. Dec;17(12):2940-8

Cauda, F. et al. (2011) Functional connectivity of the insula in the resting brain. *Neuroimage* 55, 8–23

Campbell-Meiklejohn D, Woolrick MW, Passingham RE, Rogers RD (2008) Knowing when to stop: the brain mechanisms of chasing losses. *Biol Psychiatry* 63:292–300

Campbell-Meiklejohn D, Wakeley J, Herbert V et al (2011) Serotonin and dopamine play complementary roles in gambling to recover losses. *Neuropsychopharmacology* 36:402–410

Camara E, Rodriguez-Fornells A and Münte TF (2009a) Functional connectivity of reward processing in the brain. *Front. Hum. Neurosci.*2:19.

Camara E, Rodriguez-Fornells A, Ye Z and Münte TF (2009b). Reward networks in the brain as captured by connectivity measures. *Front. Neurosci.*3:1.350-362

Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J.R., Sirigu, A. (2004). The involvement of the orbitofrontal cortex in the experience of regret. *Science,* 304, 1167–70

Carlezon W.A. Thomas Jr., M.J. (2009) Biological substrates of reward and aversion: a nucleus accumbens activity hypothesis *Neuropharmacology,* 56 (Suppl 1) pp. 122–132

Cardinal,R.N., Parkinson,J.A., Hall,J., and Everitt,B.J. (2002).Emotion and motivation: the role of the amygdala, ventral striatum,and prefrontal cortex. *Neurosci.Biobehav.Rev. 26*, 321-352.

Carter, R. M. (2009). Activation in the VTA and nucleus accumbens increases in anticipation of both gains and losses. *Frontiers in Behavioral Neuroscience*, *3*(August), 1-15.

Carlsson, A. (1988) The current status of the dopamine hypothesis of schizophrenia. *Neuropsychopharmacology,* Vol 1(3), Sep 179-186.

Cisek, P. and Kalaska, J.F. (2001) Common codes for situated interaction. *Behavioral and Brain Sciences.* 24(5): 883-884.

Cisek, P. (2006) Integrated neural processes for defining potential actions and deciding between them: A computational model. *Journal of Neuroscience.* 26(38): 9761-9770.

Cisek, P. (2007) Cortical mechanisms of action selection: The affordance competition hypothesis*Philosophical Transactions of the Royal Society B.* 362: 1585-1599.

Cisek, P. and Kalaska, J.F. (2010) Neural mechanisms for interacting with a world full of action choices *Annual Review of Neuroscience.*33: 269-298.

Clark, Andy. (1997) Being There: Putting Brain, Body and World Together Again MIT Press, Bradford Books

Clark L, Chamberlain SR, Sahakian BJ. (2009) Neurocognitive mechanisms in depression: implications for treatment. Annu Rev Neurosci. 32:57–74.

Chase T.N. (2011) Apathy in neuropsychiatric disease: diagnosis, pathophysiology, and treatment *Neurotoxicity research,* Feb;19(2):266-78.

Chevalier, G. and Deniau, J. M. (1990) Disinhibition as a basic process in the expression of striatal functions. *Trends Neurosci.* 13, 277-281

Chib, V. S., Rangel, A., Shimojo, S., & O'Doherty, J. P. (2009) Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex.*The Journal of neuroscience: Neuroscience*, *29*(39), 12315-20.

Chiba AA, Kesner RP, Gibson CJ. (1997) Memory for temporal order of new and familiar spatial location sequences: role of the medial prefrontal cortex. *Learn Mem.* Nov-Dec;4(4):311-7.

Churchland, P. S. and T. J. Sejnowski (1992). The computational brain, MIT Press.

Coizet, V. Dommett, E.J. Redgrave, P. Overton P.G. (2006), Nociceptive responses of midbrain dopaminergic neurones are modulated by the superior colliculus in the rat *Neuroscience,* 139 pp. 1479–1493

Cohen J. D., Dunbar K, & McClelland .JL (1990) On the control of automatic processes: A parallel processing model of the Stroop effect. *Psychological Review*, 97(3):332-361.

Cohen JD, Braver TS & Brown JW (2002) Computational perspectives on dopamine function in prefrontal cortex, *Current Opinion in Neurobiology*, 12, 223-229.

Cohen, M.X. (2007) Individual differences in reinforcement learning parameters and the neural representations of value and reward prediction errors. *Social, Cognitive, and Affective Neuroscience* Mar;2(1):20-30.

Cole, M. W., Yeung, N., Freiwald, W. A. & Botvinick, M. (2009) Cingulate cortex:

diverging data from humans and monkeys. *Trends Neurosci.* 32, 566–574

Colwill *RM,* Rescorla *R.A.* (1985) Postconditioning devaluation of a reinforcer affects instrumental responding. *J Exp Psychol Anim Behav Processes* 11*:*120*–132.*

Cools R., Robinson O.J., Sahakian B. (2008a) Acute tryptophan depletion in healthy volunteers enhances punishment prediction but does not affect reward prediction. *Neuropsychopharmacology 33:2291–2299.*

Cools R, Roberts A.C, Robbins T.W. (2008b) Serotoninergic regulation of emotional and behavioural control processes. *Trends Cogn Sci 12:31–40.*

Cools R., Blackwell A, Clark L, Menzies L, Cox S, Robbins T.W. (2005) Tryptophan depletion disrupts the motivational guidance of goal-directed behavior as a function of trait impulsivity. *Neuropsychopharmacology 30:1362–1373*

Cools, R., Rogers, R., Barker, R. a, & Robbins, T. W. (2010) Top-down attentional control in Parkinson's disease: salient considerations. *Journal of cognitive neuroscience,* 22(5), 848-59.

Corbetta, M. & Shulman, G. L. (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nature Rev. Neurosci.* 3, 201–215

Corrado, G., & Doya, K. (2007). Understanding neural coding through the model-based analysis of decision making. *Journal of Neuroscience,* 27, 8178–8180.

Costa, R.M., (2007) Plastic corticostriatal circuits for action learning: what's dopamine got to do with it? *Ann. NY Acad. Sci.* 1104, 172–191.

Cowles, J. T. (1937). Food-tokens as incentives for learning by chimpanzees.Comparative Psychology Monograph, 1937, 14 (5, Serial No. 71).

Craig,A.D. (2002). How do youfeel? Interoception: the sense of the physiological condition of the body. *Nat.Rev.Neurosci.3*, 655-666.

Craig, A.D. (2009) How do you feel—now? The anterior insula and human awareness. *Nat Rev Neurosci.* 10: 59-70.

Crockett, M.J., Clark, L., Roiser, J.P., Robinson, O.J., Cools, R., Chase, H.W., den Ouden, H., Apergis-Schoute, A.M., Campbell-Meikeljohn, D., Seymour, B., Sahakian, B.J., Rogers, R.D., & Robbins, T.W. (2012). Converging evidence for central 5-HT effects in acute tryptophan depletion. *Molecular Psychiatry,* 17(2):121-3.

Crockett M.J., Clark L., & Robbins T.W. (2009). Reconciling the role of serotonin in punishment and inhibition in humans: tryptophan depletion abolishes punishment-induced inhibition. *Journal of Neuroscience*, 29(38), 11993-11999.

Curzon G. (1990) Serotonin and appetite. *Ann NY Acad Sci 600:521–531.*

Cutsuridis, V., Hussain, A., and Taylor, J., (2011) *Perception-Action Cycle: Models, Architecture and Hardware*, 601-636. Springer.

Dayan P (2008) The role of value systems in decision making. In Engel C & Singer W, editors, Better than Conscious? Decision Making, the Human Mind, and Implications for Institutions Frankfurt, Germany: MIT Press, 51-70.

Dayan P, Abbott LF. (2001) Theoretical neuroscience: computational and mathematical modeling of neural systems. Cambridge (MA): The MIT Press.

Dayan P & Huys QJM (2008) Serotonin, inhibition and negative mood. *Public Library of Science: Computational Biology* **4** e4.

Dayan, P., Kakade, S., & Montague, P. R. (2000) Learning and selective attention. *Nature Neurscience* 3, 1218–23

Dayan, P. & Yu, A.J. (2003) Uncertainty and learning. *IETE Journal of Research* 49, 171-182.

Davis J, Choi, DL, and Benoit S. (2009) Insulin, Leptin and Reward. *Trends in Endocrinology and Metabolism*. 21 (2) 68-74.

Daw, N. D Kakade, S Dayan, P. (2002) Opponent interactions between serotonin and dopamine. *Neural Networks* 15:603-616

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

Daw,et al. (2006) Cortical substrates for exploratory decisions in humans. *Nature* 2006 Jun 15;441(7095):876-9.

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

Dawson, M.R.W. (2008). *Connectionism and Classical Conditioning.* A peer reviewed monograph (115 pages) published by *Comparative Cognition and Behavior Reviews* on behalf of the Comparative Cognition Society.

D'Ardenne, K., McClure, S.M., Nystrom, L.E., Cohen, J.D. (2008) BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. *Science* 319: 1264-1267.

de Araujo, M.L. Kringelbach, E.T. Rolls and F. McGlone, Human cortical responses to water in the mouth, and the effects of thirst, J Neurophysiol 90 (2003), pp. 1865–1876.

Deakin, J. F. & Graeff, F. G. (1991) 5-HT and mechanisms of defence. *Journal of Psychopharmacology,* 5, 305 -315

Deffains M, Legallet E, Apicella P. (2010) Modulation of neuronal activity in the monkey putamen associated with changes in the habitual order of sequential movements. *J Neurophysiol.* Sep;104(3):1355-69.

Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA. (2000) Tracking the hemodynamic responses to reward and punishment in the striatum. *J. Neurophysiol.* 84:3072–77

Delgado, M.R., Stenger, V.A., Fiez, J.A. (2004) Motivation-dependent responses in the human caudate nucleus. *Cerebral Cortex, 14*:1022-1030.

Delgado, M.R., Jou, R.L., & Phelps, E.A. (2011). Neural systems underlying aversive conditioning in humans with primary and secondary reinforcers. *Frontiers in Human Neuroscience, 5:* 71.

Delgado José Manuel Rodriguez (1969) *Physical Control of the Mind: Toward a Psychocivilized Society*. Harper and Row.

Deniau, J. M. & Chevalier, G. (1985) Disinhibition as a basic process in the expression of striatal functions. II. The striato-nigral influence on thalamocortical cells of the ventromedial thalamic nucleus. *Brain Res.* 334, 227–233

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18:193-222

Devinsky, O., Morrell, M. J. & Vogt, B. A. (1995) Contributions of anterior cingulate to behaviour. Brain 118, 279–306

Diaconescu AO, Menon M, Jensen J, Kapur S, McIntosh AR Dopamine-induced changes in neural network patterns supporting aversive conditioning. *Brain Res* 1313: 143–161.

Dickinson,A. (1985) Actions and habits: The development of behavioral autonomy. *Philosophical Transactions of the Royal Society (London), SeriesB, 308,* 6778. That learns to count. *ConnectionScience,11,*5–40.

Dickinson, A., & Balleine, B. (2002). The role of learning in motivation. In C. R. Gallistel (Ed.), Learning, motivation, and emotion (pp. 497-533). New York: Wiley.

Dillon DG, Holmes AJ, Jahn AL, Bogdan R, Wald LL, Pizzagalli DA. (2008) Dissociation of neural regions associated with anticipatory versus consummatory phases of incentive processing. Psychophysiology. 45:36–49

Doya K (2007) Reinforcement learning: Computational theory and biological mechanisms. *HFSP Journal*, 1, 30-40.

Doya K (2008) Modulators of Decision Making. *Nature Neuroscience*11, 410 - 416

Doyon, J. Bellec, P. Amsel, R. Penhune, V. Monchi, O. Carrier, J. Lehéricy, S. Benali, H. (2009) Contributions of the basal ganglia and functionally related brain structures to motor learning. Behav Brain Res. Apr 12;199(1):61-75.

Doyon J, Benali H, (2005) Reorganization and plasticity in the adult brain during learning of motor skill*s*. Current Opinion in Neurobiology, 2005, 15 (2): 161-167

Doyon J, Ungerleider L. G. (2002) Functional anatomy of motor skill learning. In:Squire 1032 LR, SchacterDL, editors. *Neuropsychology of memory*.p. 225–38.

Draganski B., Kherif F., Kloppel S., Cook P.A., Deichmann R., Alexander D.C., Parker G.J.M, Ashburner J., and Frackowiak, R.S.J. (2008) Evidence for segregated and integrative connectivity patterns in the human basal ganglia, *J Neurosci.* Jul 9;28(28):7143-52

Dreher J.C., Grafman J. (2003) Dissociating the roles of the rostral anterior cingulate and the lateral prefrontal cortices in performing two tasks simultaneously or successively. *Cereb Cortex* 13:329 –339.

Dum, R. P., Levinthal, D. J. & Strick, P. L. (2009) The spinothalamic system targets motor and sensory areas in the cerebral cortex of monkeys. *J. Neurosci.* 29, 14223–14235

Duncan J, Owen AM (2000) Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci* 23(10):475-83

Duzel E, Habib R, Guderian S, Heinze HJ (2004) Four types of novelty-familiarity responses in associative recognition memory of humans. *Eur J Neurosci* 19:1408-1416.

Duzel E, Bunzeck N, Guitart-Masip M, Wittmann B, Schott BH, Tobler PN (2009) Functional imaging of the human dopaminergic midbrain. *Trends Neurosci* 32:321–328.

Eichenbaum H, Fagan A, Cohen NJ (1986) Normal olfactory discrimination learning set and facilitation of reversal learning after medial temporal damage to rats: implications for an account of preserved learning abilities in amnesia. *J Neurosci* 6:1876–1884.

Eisenberger NI, Lieberman MD. (2004) Why rejection hurts: a common neural alarm system for physical and social pain. *Trends Cogn. Sci.* 8:294–300

Eisenberger NI, Lieberman MD, Williams KD. (2003) Does rejection hurt? An fMRI study of social exclusion. *Science* 302:290–92

Elliott R, Friston KJ, Dolan R. J. (2000) Dissociable neural responses in human reward systems. *J. Neurosci.* 20:6159–65

Elliott, R., Agnew, Z. & Deakin, J (2010) Hedonic and informational functions of the human orbitofrontal cortex. *Cereb Cortex*, 20(1), 198-204.

Elliott R, Newman JL, Longe O.A, Deakin J.F. (2004) Instrumental responding for rewards is associated with enhanced neuronal response in subcortical reward systems. *NeuroImage 21:*984–990.

Erickson KI, Boot WR, Basak C, Neider MB, Prakash RS, et al. (2010) Striatal Volume Predicts Level of Video Game Skill Acquisition. *Cerebral Cortex* 20: 2522–2530.

Erk, S., M. Spitzer, et al. (2002) Cultural objects modulate reward circuitry. *Neuroreport 13(18):* 2499-503.

Esher N, Roiser J. Reward and punishment processing in depression. Biol Psychiatr. 2010;68:118–124.

Estes W.K (1967) Outline of the theory of punishmentIn B. A. Campbell & R.M Church (Eds).Punishment and Aversive behavior (pp.57-82). New York: Appleton-Centuary-Crofts

Evers EA, Cools R, Clark L, van der Veen FM, Jolles J, Sahakian BJ, et al. (2005) Serotonergic modulation of prefrontal cortex during negative feedback in probabilistic reversal learning. Neuropsychopharmacology. 30:1138–1147

Fahrmeir L, Tutz D (2001) Multivariate statistical modelling based on generalized linear models. New York: Springer.

Farrell, M. J., Laird, A. R. & Egan, G. F Brain activity associated with painfully hot stimuli applied to the upper limb: A meta-analysis. *Hum. Brain Mapp.* 25, 129–139 (2005).

Ferry AT, et al. (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 CD, Tobler PN & Schultz W. (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. *Science 299:* 1898-1902

Fladung, A. et al., (2010) A Neural Signature of Anorexia Nervosa in the Ventral Striatal Reward System *Am J Psychiatry* 2010; 167:206-212

Fuster, J.M., (2003) Cortex and Mind: Unifying Cognition. New York: Oxford.

Fudge J.L., Kunishio K., Walsh P, Richard C, Haber SN. (2002) Amygdaloid projections to ventromedial striatal subterritories in the primate. *Neuroscience;* 110(2):257-75.

Fujiwara J, Tobler PN, et al. (2009) Segregated and integrated coding of reward and punishment in the cingulate cortex. *J Neurophysiol.;* 101(6):3284–93.

Frank M.J. (2005) Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism J. Cogn. Neurosci., 17 (2005), pp. 51–72

Frank, M.J., Loughry, B. & O'Reilly, R.C. (2001). Interactions between the frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, and Behavioral Neuroscience,* 1, 137-160.

Frank, M.J. Seeberger, L.C. O'Reilly R.C. (2004) By carrot or by stick: Cognitive reinforcement learning in parkinsonism Science, 306 (2004), pp. 1940–1943

Frank, M.J. & O'Reilly, R.C. (2006). A mechanistic account of striatal dopamine function in human cognition: Psychopharmacological studies with cabergoline and haloperidol. *Behavioral Neuroscience, 120,* 497-517.

Frank, M.J., Cohen, M.X. & Sanfey, A.G. (2009). Multiple systems in decision making: A neurocomputational perspective. *Current Directions in Psychological Science, 18,* 73--77.

Friston K.(2009) The free-energy principle: a rough guide to the brain? *Trends Cogn Sci.* Jul;13(7):293-301.

Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E., Penny, W. D. (Eds.). (2006) Statistical parametric mapping: the analysis of functional brain images Amsterdam: Elsevier.

Fox MT, Barense MD, BaxterMG (2003) Perceptual attentional set-shifting is impaired in rats with neurotoxic lesions of posterior parietal cortex. *J Neurosci* 23:676–681.

Gao Q, Horvath TL. (2007) Neurobiology of feeding and energy *expenditure Annu Rev Neurosci.*; 30:367-98.

Gazzaley, A. and D'Esposito, M. (2007) Unifying prefrontal cortex function: Executive control, neural networks and top-down modulation. In: Miller, B., Cummings, J. (Eds) The Human Fontal Lobes. Guildford Publications

Gerardin, E., Lehericy, S., Pochon, J.B., Tezenas du Montcel, S., Mangin, J.F., Poupon, F., Agid, Y., Le Bihan, D. and Marsault, C., (2003) Foot, hand, face and eye representation in the human striatum. *Cereb. Cortex* 13, pp. 162–169.

Gerfen, C.R.. Engber, T.M Mahan, L.C. Susel, Z. Chase, T.N. Monsma F.J. Jr., D.R. Sibley (1990) D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. *Science*, 250 pp. 1429–1432

Gerfen, C. R., & Surmeier, D. J. (2011). Modulation of striatal projection systems by dopamine. *Annual review of neuroscience*, *34*, 441-66.

Gershman, S.J., Pesaran, B., & Daw, N.D. (2009) Human reinforcement learning subdivides structured action spaces by learning effector-specific values, *Journal of Neuroscience*, 29, 13524-13531.

Ghahremani, D.G. and Poldrack, R.A., (2009) Neuroimaging and Interactive Memory Systems. In: Rösler, F., Ranganath, F., Röder, B., Kluwe, R.H., (Ed.) Neuroimaging of Human Memory. Oxford-University Press

Ghashghaei, H. T., Hilgetag, C. C. & Barbas, H. (2007) Sequence of information processing for emotions based on the anatomic dialogue between prefrontal cortex and amygdala. *Neuroimage* 34, 905–923

Gläscher, J., Hampton, A. N., & O'Doherty, J. P. (2009). Determining a Role for Ventromedial Prefrontal Cortex in Encoding Action-Based Value Signals During Reward-Related Decision Making. *Cerebral Cortex,* 19(2), 483 -495.

Gläscher et al., (2010) States versus Rewards: Dissociable Neural Prediction Error Signals Underlying Model-Based and Model-Free Reinforcement Learning Neuron, Volume 66, Issue 4, 585-595, 27

Glimcher PW, Rustichini A. (2004) Neuroeconomics: the consilience of brain and decision. *Science; 306:*447–452.

Glimcher, P.W. (2009a). Neuroeconomics and the Study of Valuation. In: Gazzaniga , M.S. (ed.) *The Cognitive Neurosciences*, Fourth Edition. Cambridge, MA: The MIT Press, pp. 1085-1092.

Glimcher, P.W. (2009b). Neuroscience, Psychology, and Economic Behavior: The Emerging Field of Neuroeconomics. In: Tommasi, L., Peterson, M.A., and Nadel, L. (eds.) *Cognitive Biology: Evolutionary and Developemental Perspectives on Mind, Brain, and Behavior.* Cambridge, MA: The MIT Press, pp. 261-278.

Gluck, M. A. & Bower, G. H. (1988)From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology:* General,117(3), 227-247.

Gluck, M. A., Mercado, E., & Myers, C. E. (2008). *Learning and Memory: From Brain to Behavior. New York: Worth.*

Gold, JI and Shadlen MN (2007) The Neural Basis of Decision Making.*Annual Review of Neuroscience,* Vol. 30: 535-574

Goto Y., Otani S, and. Grace A. A. (2007) The Yin and Yang of Dopamine Release A New Perspective *Neuropharmacology* October; 53(5): 583–587.

Gottfried, J. a, O'Doherty, J., & Dolan, R. J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science (New York, N.Y.)*, *301*(5636), 1104-7.

Grabenhorst,F. and Rolls,E .T. (2011) Value, pleasure, and choice in the ventral prefrontal cortex. *Trends in Cognitive Sciences* 15: 56-67.

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

Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiol. Learn. Mem.* 70, 119–136.

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

Grill H. J. , S. Karolina P. Hayes M. R. (2007) Imaging Obesity: fMRI, Food Reward, and Feeding *Cell Metabolism* Volume6, Issue6,5, Pages 423-425

Grossberg S. (1984) Some normal and abnormal behavioral syndromes due to transmitter gating of opponent processes. *Biol Psychiatry* 19: 1075–1118.

Guitart-Masip, M., Fuentemilla, L., Bach, D.R., Huys, Q.J., Dayan, P., Dolan, R.J., Duzel, E., (2011) Action dominates valence in anticipatory representations in the human striatum and dopaminergic midbrain. J. Neurosci. 31, 7867–7875.

Haber, S. N., Fudge, J. L., & Mcfarland, N. R. (2000) Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum.*J. Neurosci.*, 20, 2369‑2382.

Haber, S. N. (2003) The primate basal ganglia: parallel and integrative networks. *J. Chem. Neuroanat.*26, 317–330 83.

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

Hagelberg N., Jääskeläinen S.K., Martikainen I.K., Mansikka H., Forssell H., Scheinin H., Hietala J, Pertovaara A. (2004) Striatal dopamine D2 receptors in modulation of pain in humans: a review. *Eur J Pharmacol.* Oct 1;500(1-3):187-92.

Hampton, A.N. Bossaerts, P. O'Doherty J.P. (2006) The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans *J Neurosci,* 26 pp. 8360–8367

Hamann SB, Herman RA, Nolan CL, Wallen K. Men and women differ in amygdala response to visual sexual stimuli. *Nat Neurosci 2004;* 7:411–416.

Harlow, Harry F. (1953) Mice, monkeys, men, and motives. *Psychological Review*, Vol 60(1), Jan 23-32.

Haruno, M., & Kawato, M. (2006). Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus action‑reward association learning. *Neural Network*, 19, 1242-1254.

Haruno, M., Kuroda, T., Doya, K., Toyama, K., Kimura, M., Samejima, K., et al. (2004). A neural correlate of reward-based behavioral learning in caudate nucleus: a functional magnetic resonance imaging study of a stochastic decision task. *J. Neurosci,* 24, 1660-1665.

Hassani, O. K., Cromwell, H. C. & Schultz, W. (2001) Influence of expectation of different rewards on behavior-related neuronal activity in the striatum. *J. Neurophysiol.* 85, 2477–2489.

Hazy, T.E. Frank, M.J. O'Reilly, R.C. (2007) Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 139 105–118.

Holroyd CB, Coles MG (2002) The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol Rev* 109:679 – 709.

Hélie, S., & Cousineau, D. (2011). The cognitive neuroscience of automaticity: Behavioral and brain signatures. *Cognitive Sciences,* 6*, 25-43.*

Henson RNA, Rugg MD (2003) Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia* 41:263- 270.

Herculano-Houzel S (2009). The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in Human Neuroscience* 3: 31.

Hertwig, R. & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists*?. Behavioral and Brain Sciences,* 24, 383–403.

Hester, R., Murphy, K., Brown, F. L., & Skilleter, A. J. (2010). Punishing an error improves learning: the influence of punishment magnitude on error-related neural activity and subsequent learning. *The Journal of Neuroscience 30*(46), 15600-7.

Hikida, T. Kimura, K. Wada, N. Funabiki, K. Nakanishi S. (2010) Distinct roles of synaptic transmission in direct and indirect striatal pathways to reward and aversive behavior Neuron, 66, pp. 896–907

Hikosaka O. (2007) Basal ganglia mechanisms of reward-oriented eye movement Ann. N Y Acad. Sci., 1104, pp. 229–249

Hikosaka, O., Bromberg-Martin, E., Hong, S., Matsumoto, M. (2008) New insights on the subcortical representation of reward *Current Opinion in Neurobiology*, *Volume 18:Issue2 ,203-208*

Hikosaka, O., Isoda, M., (2010) Switching from automatic to controlled behavior: cortico-basal ganglia mechanisms. *Trends Cogn. Sci.* 14, 154–161.

Hikosaka O, Rand MK, Nakamura K, Miyachi S, Kitaguchi K, Sakai K, Lu X, Shimo Y (2002) Long-term retention of motor skill in macaque monkeys and humans. *Exp Brain Res* 147:494-504.

Hikosaka, K; Watanabe, M (2000) Delay activity of orbital and lateral prefrontal neurons of the monkey varying with different rewards. *Cereb Cortex*: 263-71

Hikosaka, O*.,* Nakahara, H*.,* Rand, M.K*.,* Sakai, K*.,* Lu, X*.,* Nakamura, K*.,* Miyachi, S. Doya, K. *(*1999*)* Parallel neural networks for learning sequential procedures. *Trends Neurosci.,*22*,* 464–471.

Hikosaka O, Rand MK, Miyachi S, Miyashita K (1995) Learning of sequential movements in the monkey - Process of learning and retention of memory. *J Neurophysiol* 74: 1652-1661.

Houk,J.C., J.L.AdamsA.G.Barto (1995) A model of how the basal ganglia generate and use neural signals that predict reinforcement. In Models of Information Processing in the Basal Ganglia.J.C.Houk, J.L.Davis B.G. Beiser, Eds.:249–270. MIT Press. Cambridge.

Hollerman, J. R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature neuroscience,* 1(4), 304-9.

Holt, D.J., Graybiel, A.M. and Saper C.B. (1997) Neurochemical architecture of the human striatum. *J. Comp. Neurol.* 384:1-25.

Hommel, B., Müsseler, Aschersleben, G. and Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences,* 24, 849-937.

Hommel, B. (2004). Event files: feature binding in and across perception and action. *Trends in Cognitive Sciences,* 8, 494-500.

Hommel, B. (2010). Grounding attention in action control: The intentional control of selection. In B.J. Bruya (ed.), Effortless attention: A new perspective in the cognitive science of attention and action (pp. 121-140). Cambridge, MA: MIT Press

Hori, Y., Minamimoto, T., and Kimura, M. (2009) Neuronal Encoding of Reward Value and Direction of Actions in the Primate Putamen *J Neurophysiol,* December 1, 2009; 102(6): 3530 – 3543

Horvitz, J. C., Stewart, T., & Jacobs, B. (1997) Burst activity of ventral tegmental dopamine neurons is elicited by sensory stimuli in the awake cat. *Brain Research,* 759, 251 – 258.

Horvitz, J.C. (2000) Mesolimbic and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience*, 96, 651-656.

Horvitz, J. O. N. C., Choi, W. O. N. Y., Morvan, C., Eyny, Y., & Balsam, P. D. (2007) A "Good Parent" Function of Dopamine. *New York Academy of Sciences,* 288, 270-288.

Horvitz, J. C. (2009) Stimulus-response and response-outcome learning mechanism in the striatum. *Behavioural Brain Research,* 199, 129-140.

Howes O.D., Kapur S. (2009) The Dopamine Hypothesis of Schizophrenia: Version III—The Final Common Pathway *Schizophrenia Bulletin* Volume 35,Issue 3 Pp. 549-562.

Huettel, S.A., Song, A.W., & McCarthy, G. (2004) Functional Magnetic Resonance Imaging. Sunderland, Massachusetts: Sinauer Associates

Hull, Clark L. (*1943*) *Principles of behavior*. New York: Appleton-Century

Humphries, M. D., and Prescott, T. J. (2010), The ventral basal ganglia, a selection mechanism at the crossroads of space, strategy, and reward, *Progress in Neurobiology*, 90:385-417.

Hutchison, W. D., Davis, K. D., Lozano, a M., Tasker, R. R., & Dostrovsky, J. O. (1999).Pain-related neurons in the human cingulate cortex. *Nature neuroscience*, *2*(5), 403-5.

Hyman, S. E., Malenka, R. C. & Nestler, E. J. (2006) Neural mechanisms of addiction: the role of reward–related learning and memory. *Annu. Rev. Neurosci.* 29, 565–598

Ide, J. S., & Li, C.-S. R. (2011). Error-related functional connectivity of the habenula in humans. *Frontiers in human neuroscience*, *5*(March), 25.

Ihalainen JA, Riekkinen P, Jr., Feenstra MG (1999) Comparison of dopamine and noradrenaline release in mouse prefrontal cortex, striatum and hippocampus using microdialysis. *Neurosci Lett* 277:71-74.

Isella V, Melzi P, Grimaldi M, Iurlaro S, Piolti R, Ferrarese C, et al (2002) Clinical, neuropsychological, and morphometric correlates of apathy in Parkinson's disease. *Mov Disord 17:366-71.*

Inase M, Li BM, Takashima I, Iijima T. (2006) Cue familiarity is represented in monkey medialprefrontal cortex during visuomotor association learning. *Exp Brain Res.* Jan;168(1-2):281-6.

Izuma, K., Saito, D.N., and Sadato, N. (2008) Processing of social and monetary rewards in the human striatum. *Neuron* 58, 284–294

Jacobs, R.A. (1988) Increased rates of convergence through learning rate adaptation. Neural Networks 1, 295–307

Jansma JM, Ramsey NF, Slagter HA, Kahn RS. (2001) Functional anatomical correlates of controlled and automatic processing. J Cogn Neurosci. Aug 15;13(6):730-43.

Jenkins et al. (1994) Motor sequence learning: a study with positron emission tomography *J. Neurosci.,* 14 pp. 3775–3790

Jensen, J., McIntosh, A. R., Crawley, A. P., Mikulis, D. J., Remington, G., & Kapur, S. (2003). Direct activation of the ventral striatum in anticipation of aversive stimuli. *Neuron*, *40*(6), 1251-7.

Jiang Y (2004) Resolving dual-task interference: an fMRI study. *NeuroImage* **22**: 748-754.

Jocham, G., Neumann, J., Klein, T. A., Danielmeier, C., & Ullsperger, M. (2009). Adaptive Coding of Action Values in the Human Rostral Cingulate Zone. *Journal of Neuroscience* 29(23), 7489 -7496.

Joel D, Niv Y & Ruppin E (2002) Actor-critic models of the basal ganglia: New anatomical and computational perspectives - Neural Networks 15, 535-547.

Jones, A.K.P., Friston, K., and Frackowiak, R.S.J. (1992). Localization of Responses to Pain in Human Cerebral-Cortex. *Science 255*, 215.

Joshua, M. Adler, A. Mitelman, R. Vaadia, E. Bergman H. (2008) Midbrain dopaminergic neurons and striatal cholinergic interneurons encode the difference between reward and aversive events at different epochs of probabilistic classical conditioning trials *J. Neurosci.,* 28 pp. 11673–11684

Jonides, J. (2004) How does practice make perfect?. *Nature Neuroscience,* 7, 10-11.

Jueptner, M., Frith, C. D., Brooks, D. J., Frackowiak, R. S. & Passingham, R. E. (1997). Anatomy of motor learning. II. Subcortical structures and learning by trial and error. *J. Neurophysiol.* 77, 1325–1337 87.

Jueptner M. et al. (1997), Anatomy of motor learning. I. Frontal cortex and attention to action *J. Neurophysiol.,*77 pp. 1313–1324

Kable, J.W. and Glimcher, P.W. (2009). The Neurobiology of Decision: Consensus and Controversy. *Neuron.* 63(6): 733-745.

Kadıhasanoğlu, D. Erdeniz, B. Kucuktunca, M (2007) Robots and Their Brains. Cognitive Neuroscience Forum Journal, Vol 3,1

Kagan, Jerome. (2009) Categories of novelty and states of *uncertainty*. Review of General Psychology, Vol 13(4), 290-301.

Kaelbling, Leslie P.; Michael L. Littman; Andrew W. Moore (1996). Reinforcement Learning: A Survey. Journal of Artificial Intelligence Research 4: 237–285.

Kaelbling, Leslie P Learning in Embedded Systems, The MIT Press, 1993

Kakade, S., & Dayan, P. (2002) Dopamine: generalization and bonuses. *Neural Netw.,* 15, 549-559.

Kapur S. (2003) Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. Am *J Psychiatry; 160:13-23.*

Kapur S, Mizrahi R, Li M (2005) From dopamine to salience to psychosis–linking biology, pharmacology and phenomenology of psychosis. *Schizophr Res;79:59-68.*

Kawabata, H. and S. Zeki (2004) Neural correlates of beauty. *J Neurophysiol91(4):* 1699-705.

Kawamura, K., Dodd, W., Ratanaswasd, P. & Gutierrez, R. A. (2005) Development of a Robot with a Sense of Self. 6th CIRA Symposium, Espoo, Finland.

Kawagoe, R., Takikawa, Y., Hikosaka, O., (1998) Expectation of reward modulates cognitive signals in the basal ganglia. *Nat Neurosci* 1:411– 416.

Kawato, M., & Samejima, K. (2007). Efficient reinforcement learning: computational theories, neuroscience and robotics. *Current opinion in neurobiology, 17*(2), 205-12.

Kaye, Walter H. Fudge Julie L. & Paulus Martin (2009) New insights into symptoms and neurocircuit function of anorexia nervosa *Nature Reviews Neuroscience* 10, 573-584

Keijzer, Fred. (2001) Representation and behavior. Cambridge, MA: MIT Press

Kelly A.M., Garavan, H. (2005) Human functional neuroimaging of brain changes associated with practice, *Cereb. Cortex* 15 pp. 1089–1102.

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

Kim, H., Shimojo, S., & O'Doherty, J. P. (2006). Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. PLoS Biology, 4(8), e233.

Kim, H., Shimojo, S., & O'Doherty, J. P. (2011). Overlapping responses for the expectation of juice and money rewards in human ventromedial prefrontal cortex. *Cerebral Cortex* (New York, N.Y.: 1991), 21(4), 769-776.

Kim, H., Sul, J.H., Huh, N., Lee, D., Jung, M.W., (2009) Role of striatum in updating values of chosen actions. J. Neurosci. 29, 14701–14712.

Kirchhoff BA, Wagner AD, Maril A, Stern CE (2000) Prefrontal-temporal circuitry for episodic encoding and subsequent memory. *J Neurosci* 20:6173-6180.

Kirsch-Darrow L, Fernandez HH, Marsiske M, Okun MS, Bowers D (2006) Dissociating apathy and depression in Parkinson disease. *Neurology; 67:33-8.*

Klingberg, T. Roland P.E (1998) Right prefrontal activation during encoding, but not during retrieval, in a non-verbal paired-associates task *Cereb. Cortex,* 8 pp. 73–79

Knudsen, E.I. (2007) Fundamental components of attention. *Annual Review of Neuroscience* 30:57–78.

Knowlton, B. J., Mangels, J. A. & Squire, L. R. (1996) A neostriatal habit learning system in humans. *Science* 273, 1399–1402 .

Knight RT (1984) Decreased responses to novel stimuli after prefrontal lesions in man. *Electroencephalogr Clin Neurophysiol* 59:9 –20.

Knight RT (1996) Contribution of the human hippocampal region to novelty detection. Nature 383:256 –259.encoding networks in the human brain: Positron emission tomography data. *Neuroreport* 5:2525–2528.

Knight, R.T. Scabini D. (1998) Anatomic bases of event-related potentials and their relationship to novelty detection in humans *J. Clin. Neurophysiol.,* 15 pp. 3–13

Knutson, B. & Cooper, J. C. (2005). Functional magnetic resonance imaging of reward prediction.*Current Opinion in Neurology, 18*, 411-417.

Knutson, B., Gibbs, S. E. B. (2007). Linking nucleus accumbens dopamine and blood oxygenation. *Psychopharmacology,* 191, 813-822.

Knutson B, Fong GW, Adams CM, Varner JL, Hommer D. (2001) Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport* 12:3683–87

Knutson, B., Rick, S., Wimmer, G. E., Prelec, D., Loewenstein, G. (2007) Neural predictors of purchases. *Neuron,* 53, 147-157.

Kobayashi, S. et al. (2006) Influences of rewarding and aversive outcomes on activity in macaque lateral prefrontal cortex. *Neuron* 51: 861–870.

Koob, G.F., Drugs of abuse: anatomy, pharmacology and function of reward pathways, *Trends in Pharmacological Sciences,* 13 (1992) 177-184.

Konorski J (1967). Integrative Activity of the Brain: An Interdisciplinary Approach. University of Chicago Press: Chicago, IL.

Krajbich, C. Armel, A. Rangel, (2010) Visual fixations and comparison of value in simple choice. *Nature Neuroscience,* 13:1292-1298.

Krakauer JW, Shadmehr R. (2007) Towards a computational neuropsychology of action. *ProgBrainRes* ;165:383–94.1119

Kravitz, A.V. et al., (2010) Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry *Nature,* 466, pp. 622–626

Krebs, R. M., Heipertz, D., Schuetze, H., Duzel, E. (2011). Novelty increases the mesolimbic functional connectivity of the substantia nigra/ventral tegmental area (SN/VTA) during reward anticipation: Evidence from high-resolution fMRI. *Neuroimage* 58(2), 647-655

Kreitzer, A. (2009) Physiology and Pharmacology of Striatal Neurons Annual Review of Neuroscience Vol. 32: 127-147

Krimer, L. S.et al. (1998) Dopaminergic regulation of cerebral cortical microcirculation. Nature Neuroscience **1**, 286-289.

Kringelbach M.L. & Rolls E.T. (2004) The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology *, Progress in Neurobiology, 72:341-72.*

Kringelbach M.L. (2005) The human orbitofrontal cortex: linking reward to hedonic experience. *Nature Reviews Neuroscience, 6:691-702.*

Kringelbach, M.L. & Berridge, K.C. (2009) Toward a functional neuroanatomy of pleasure and happiness. *Trends in Cognitive Sciences,* 13(11), 479-487,

Kringelbach, M.L. & Berridge, K.C. (2011) The neurobiology of pleasure and happiness. In The Oxford Handbook of Neuroethics. J. Illes & B.J. Sahakian (Eds.) Oxford University Press, pp. 15-32.

Krugel LK, Biele G, Mohr PN, Li SC, Heekeren HR. (2009) Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. *Proc Natl Acad Sci USA.* 106:17951–17956.

Kubota K, Komatsu H (1985) Neuron activities of monkey prefrontal cortex during the learning of visual discrimination tasks with GO/NO-GO performances. *Neurosci Res* 3:106

Kumaran D, Summerfield JJ, Hassabis D, Maguire EA (2009). Tracking the emergence of conceptual knowledge during human decision making.. *Neuron*, 63(6), 889 – 901

Lapko, M. (2007) Control of Four-wheeled Vehicle on Ice Surface Using Attention-gated Reinforcement Learning (AGREL). *Cybernetics,* 487-497.

Laplane D, Baulac M, Widlocher D, Dubois B (1984) Pure psychic akinesia with bilateral lesions of basal ganglia. *J Neurol Neurosurg Psychiatry* 47:377–385.

Laplane D, et al. (1989) Obsessive-compulsive and other behavioural changes with bilateral basal ganglia lesions. A neuropsychological, magnetic resonance imaging and positron tomography study. *Brain; 112(Pt 3):699-725.*

Laplane D, Dubois B (2001) Auto-activation deficit: a basal ganglia related syndrome. *Mov Disord 2001;16:810-4.*

Lau, B., Glimcher, P.W. (2007) Action and outcome encoding in the primate caudate nucleus. *J Neurosci* 27:14502–14514.

Lau, B., Glimcher, P.W. (2008) Value representations in the primate caudate nucleus during matching behavior. *Neuron* 58: 451–463,

Lauwereyns, J., Takikawa, Y., Kawagoe, R., Kobayashi, S., Koizumi, M., Coe, B., Sakagami, M., Hikosaka, O., (2002) Feature-based anticipation of cues that predict reward in monkey caudate nucleus. *Neuron* 33:463– 473.

Lea, S.E.G. & Webley, P. (2006) Money as tool, money as drug. *Behavioral and Brain Sciences* 29:2:161-209

Legault M, Wise RA (2001) Novelty-evoked elevations of nucleus accumbens dopamine: dependence on impulse flow from the ventral subiculum and glutamatergic neurotransmission in the ventral tegmental area. *Eur J Neurosci* 13:819-828.

Lehéricy S, Benali H, Van de Moortele PF, Pélégrini-Issac M, Waechter T, Ugurbil K, Doyon J, (2005) Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. *Proceedings of the National Academy of Sciences of the U.S.A.*, 102 (35): 12566-12571

Leknes, S., & Tracey, I. (2008) A common neurobiology for pain and pleasure. *Nature reviews Neuroscience*, *9*(4), 314-20.

Leon, M.I. and Shadlen, MN (1999) Effect of Expected Reward Magnitude on the Response of Neurons in the Dorsolateral Prefrontal Cortex of the Macaque .*Neuron,* Vol. 24, 415–425.

Levita L, Hare TA, Voss HU, Glover G, Ballon DJ, Casey BJ. (2009) The bivalent side of the nucleus accumbens. Neuroimage. Feb 1; 44(3):1178-87

Levy R. and Dubois B. (2006) Apathy and the Functional Anatomy of the Prefrontal Cortex–Basal Ganglia Circuits. *Cerebral Cortex Volume* 16,Issue 7 Pp. 916-928.

Li, J., Daw, N.D. (2011) Signals in human striatum are appropriate for policy update rather than value prediction. *Journal of Neuroscience* 31:5504-11

Li, X., Li, W., Wang, H., Cao, J., Maehashi, K., Huang, L., Bachmanov, A. A., Reed, D. R., Legrand- Defretin, V., Beauchamp, G. K. and Brand, J. G. (2005) Pseudogenization of a sweet receptor gene accounts for cats' indifference toward sugar. *PLoS Genet.* 1, e3.

Lieberman, M., & Cunningham, W. A. (2009). Type I and Type II error concerns in fMRI research: Re-balancing the scale. *Social Cognitive and Affective Neuroscience, 4*, 423-428.

Lin MT, Wu JJ, Chandra A, Tsay BL. (1981) Activation of striatal dopamine receptors induces pain inhibition in rats. *J. Neural Transm.* 5: 213-222.

Lisman JE, Grace AA. (2005) The hippocampal-VTA loop: controlling the entry of information into long-term memory. *Neuron,* 46(5): 703-13.

Liu Z, and Ikemoto S. (2007) The midbrain raphe nuclei mediate primary reinforcement via GABA_A receptors. *Eur J Neurosci.* February; 25(3): 735–743.

Liu, Z.H Shin, R. Ikemoto S. (2008) Dual role of medial A10 dopamine neurons in affective encoding *Neuropsychopharmacology,* 33 pp. 3010–3020

Liu, X., Hairston, J., Schrier, M., & Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: A meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews.* Elsevier Ltd.

Liu, X., Powell, D. K., Wang, H., Gold, B. T., Corbly, C. R., & Joseph, J. E. (2007) Functional Dissociation in Frontal and Striatal Areas for Processing of Positive and Negative Reward Information. *J. Neurosci.* ;27:4587–4597.

Ljungberg T, Apicella P, Schultz W (1992) Responses of monkey dopamine neurons during learning of behavioral reactions. *J Neurophysiol* 67:145-163.

Logan GD (1979) On the use of a concurrent memory load to measure attention and automaticity. *J Exp Psychol Hum Percept Perform* 5 189-207

Logothetis NK (2008) What we can do and what we cannot do with fMRI *Nature* 453(7197) 869-878.

Logothetis NK , Pauls J, Augath MA, Trinath T, and Oeltermann A, (2001) Neurophysiological investigation of the basis of the fMRI signal *Nature* 412(6843)

Logothetis NK and Pfeuffer J (2004) On the Nature of the BOLD fMRI Contrast Mechanism *Magnetic Resonance Imaging* 22 (10) 1517-1531.

Luders HO, et al. (1995) Cortical electrical stimulation in humans. The negative motor areas. *Adv Neurol.;* 67:115–129.

MacDonald G, Leary MR. (2005) Why does social exclusion hurt? The relationship between social and physical pain. *Psychol. Bull.* 131:202–2300

Maddock; R. J. Garrett A. S.; Buonocore. M. H. (2001) Remembering familiar people: the posterior cingulate cortex and autobiographical memory retrieval. *Neuroscience* 104(3):667-76,

Magnusson JE , Fisher K. (2000) The involvement of dopamine in nociception: the role of D1 and D2 receptors in the dorsolateral striatum. *Brain Res.* 855: 260-266.

Mansouri FA, et al. (2006) Prefrontal cell activities related to monkeys' success and failure in adapting to rule changes in a Wisconsin Card Sorting Test analog. *J Neurosci.;* 26:2745–2756

Marin RS. (1991) Apathy: a neuropsychiatric syndrome. *J Neuropsychiatry Clin Neurosci.* Summer;3(3):243-54.

Mars, R.B., Shea, N.J, Kolling, N. and Rushworth, M.F.S. (2010) Model-based analyses: Promises, pitfalls, and example applications to the study of cognitive control. *Quarterly Journal of Experimental Psychology* 1-16

Mason, G. J., Cooper, J. and Clarebrough, C. (2001) Frustrations of fur-farmed mink. *Nature* 410, 35–36.

Matsuzaka Y, Tanji J. (1996) Changing directions of forthcoming arm movements: Neuronal activity in the presupplementary and supplementary motor area of monkey cerebral cortex. *Journal of Neurophysiology.*;76:2327–2342.

Matsumoto, M., Matsumoto, K., Abe, H., and Tanaka, K. (2007) Medial prefrontal cell activity signaling prediction errors of action values. *Nat. Neurosci.*10, 647–656.

Matsumoto, M., & Hikosaka, O. (2007). Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature*, *447*(7148), 1111-5.

McClure, S.M., Berns, G.S., Montague, P.R. (2003) Temporal prediction errors in a passive learning task activate human striatum. *Neuron, 38:* 339-346.

McClure, S. M., Laibson, D. I. et al. (2004) Separate Neural Systems Value Immediate and Delayed Monetary Rewards. *Science 306(*5695): 503-507.

McClure, S.M., Li, J., Tomlin, D., Cypert, K.S., Montague, L.M., Montague, P.R. (2004) Neural correlates of behavioral preference for culturally familiar drinks. *Neuron, 44:* 379-87.

McClure SM, York MK, Montague PR. (2004) The neural substrates of reward processing in humans: the modern role of fMRI. *Neuroscientist, 10:* 260–268.

McHaffie, J. G., Stanford, T. R., Stein, B. E., Coizet, W. and Redgrave, P. (2005) Subcortical loops through the basal ganglia. *Trends Neurosci* 28, 401-407

McNab, F. Klingberg, T. (2008) Prefrontal cortex and basal ganglia control access to working memory, *Nat. Neurosci.* 11 103–107.

McNab, F., Varrone, A., Farde, L., Jucaite, A., Bystritsky, P., Forssberg, H., Klingberg, T., (2009) Changes in cortical dopamine binding associated with cognitive training. *Science* 323, 800–802.

Mesulam, M. M. & Mufson, E. J. (1982) Insula of the old world monkey. III: Efferent cortical output and comments on function. *J. Comp. Neurol.* 212, 38–52

Mitchell P, Moyle J (1967) Chemiosmotic hypothesis of oxidative phosphorylation. *Nature* 213:137–139,

Middleton, F. A. & Strick, P. L. (2000) Basal ganglia and cerebellar loops: motor and cognitive circuits. *Brain Res. Rev.* 31, 236–250

Miller EK & Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience,* 24:167-202.

Miliaressis E, Bouchard A, Jacobowitz DM. (1975) Strong positive reward in median raphe: specific inhibition by para-chlorophenylalanine. *Brain Res.*;98:194–201.

Mink J.W., Thach W.T. (1993) Basal ganglia intrinsic circuits and their role in behavior. *Curr Opin Neurobiol* 13:3

Mink, J.W. (1996) The basal ganglia: Focused selection and inhibition of competing motor programs. *Prog Neurobiol* 50, 381-425

Minsky, M.L. (1963) Steps toward artificial intelligence. In: Feigenbaum EA, Feldman J, editors. Computers and thought. New York: McGraw-Hill.

Miyachi, S. et al. (2002) Differential activation of monkey striatal neurons in the early and late stages of procedural learning. *Exp. Brain Res.* 146, 122 – 126

Mizuno, K., Tanaka, M., Ishii, A., Tanabe, H. C., Onoe, H., Sadato, N., et al. (2008).The neural basis of academic achievement motivation. *NeuroImage, 42*(1), 369-378.

Mobbs D, Greicius MD, Abdel-Azim E, Menon V, Reiss AL. (2003) Humor modulates the mesolimbic reward centers. *Neuron, Vol. 40*, 1041–1048

Mogenson GJ, Jones DL, Yim CY (1980). From motivation to action: functional interface between the limbic system and the motor system. *Prog Neurobiol* 14: 69–97.

Montague, P.R. King-Casas, B. Cohen, J.D. (2006) Imaging valuation models in human choice. *Annual Review of Neuroscience* 29:417-448.

Morecraft, R. J. et al. (2007) Amygdala interconnections with the cingulate motor cortex in the rhesus monkey. *J. Comp. Neurol.* 500, 134–165

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

Montague, PR, Hyman, SE, Cohen, JD (2004) Computational roles for dopamine in behavioural control. *Nature 431:*760-767.

Montague, P. R., B. King-Casas, et al. (2006) Imaging valuation models in human choice. *Annu Rev Neurosci 29:* 417-48.

Moore, C.D., Cohen, M.X., and Ranganath, C. (2006) Neural mechanisms of expert skills in visual working memory. *J. Neurosci.* 26, 11187–11196.

Morel A (2007) Stereotactic Atlas of the Human Thalamus and Basal Ganglia. Informa Healthcare USA, Inc.

Morel A, Loup F, Magnin M, Jeanmonod D. (2002) Neurochemical organization of the human basal ganglia: anatomofunctional territories defined by the distributions of calcium-binding proteins and SMI-32. *J Comp Neurol;* 443(1):86−103.

Moustafa, A.A. Sherman, S.J. Frank, M.J. (2008) A dopaminergic basis for working memory, learning, and attentional shifting in Parkinson's disease, *Neuropsychologia* 46 3144–3156.

Murray, G.K., Corlett,P.R., Clark, L., Pessiglione. M., Blackwell A.D, Honey, G.' Jones P. B., Bullmore E.T., Robbins T.W, Fletcher P. Substantia nigra/ventral tegmental reward prediction error disruption in psychosis(2008) *Molecular Psychiatry*13, 267–276

Murray E.A. (2007) The amygdala, reward and emotion. *Trends in Cognitive Sciences*, 11: 489-497.

Murray EA, O'Doherty JP, and Schoenbaum G (2007) What we know (and don't know) after 20 years of investigating orbitofrontal function across species. *Journal of Neuroscience*, 27: 8166-8169.

Nadjar, A. et al., (2006) Phenotype of striatofugal medium spiny neurons in parkinsonian and dyskinetic nonhuman primates: a call for a reappraisal of the functional organization of the basal ganglia. *Journal of Neuroscience* Volume: 26, Issue: 34, Pages: 8653-8661

Nakahara K, et al. (2002) Functional MRI of macaque monkeys performing a cognitive set-shifting task. *Science.* 295:1532–1536.

Nakamura K, et al. (1998) Neuronal activity in medial frontal cortex during learning of sequential procedures. *Journal of Neurophysiology.*80:2671–2687

Nakamura K, Matsumoto M, Hikosaka O (2008) Reward-dependent modulation of neuronal activity in the primate dorsal raphe nucleus. *J Neurosci* 28:5331-5343.

Nakamura K, Hikosaka O (2006) Role of dopamine in the primate caudate nucleus in reward modulation of saccades. *J Neurosci* 26:5360-5369.

Nicola, S. M., Taha, S. A., Kim, S. W., & Fields, H. L. (2005). Nucleus Accumbens Dopamine release is Necessary and Sufficient To Promote The Behavioral Response To Reward Predictive Cues. *Drugs*, *135*, 1025-1033.

Nicola, S. M. (2007) The nucleus accumbens as part of a basal ganglia action selection circuit. Psychopharmacology (Berl). 2007 Apr;191(3):521-50.

Nissen MJ, Bullemer P (1987) Attentional requirements of learning: evidence from performance measures. *Cognit Psychol* **19**:1–32

Niv, N (2009) Reinforcement learning in the brain. *The Journal of Mathematical Psychology.*Volume 53, Issue 3, Pages 139–154

Niv, Y Daw ND & Dayan P (2006) Choice values. *Nature Neuroscience* 9(8), 987-988.

Niv, Y. Duff, M.O. & Dayan P. (2005) Dopamine, Uncertainty and TD Learning - *Behavioral and Brain Functions* 1:6 (4 May 2005), doi:10.1186/1744-9081-1-6.

Niv Y. & Schoenbaum G (2008) Dialogues on prediction errors, Trends in Cognitive Sciences, 12(7), 265-272, 2008.

Ng P., Blair R.J.R. (2011) Neural correlates of frustration: Beyond not getting what you want *66th Society of Biological Psychiatry Annual Meeting 2011*, 11, 1028

Nyberg L. (2005) Any novelty in hippocampal formation and memory? *Curr Opin Neurol.* Aug;18(4):424-8.

O'Doherty JP. (2004) Reward representations and reward-related learning in the human brain: Insights from neuroimaging. *Curr Opin Neurobiol 14:*769–766.

O'Doherty, J. P. (2007). Lights, camembert, action! The role of human orbitofrontal cortex in encoding stimuli, rewards, and choices.Annals of the New York Academy of Sciences, 1121, 254-72.

O'Doherty, J. P. (2011). Contributions of the ventromedial prefrontal cortex to goal-directed action selection. Annals of the New York Academy of Sciences, 1239, 118-129.

O'Doherty, J., Dayan, P. Friston, K.J., Critchley, H.D., Dolan, R.J. (2003) Temporal difference models and reward-related learning in the human brain.*Neuron*, 38(2), 329-337.

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ. (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, 304:452-4.

O'Doherty JP, Hampton A, Kim H. (2007) Model-based fMRI and its application to reward learning and decision making. *Ann N Y Acad Sci. May;1104*:35-53.

O'Doherty J, Kringelbach M.L., Rolls E.T., Hornak J., Andrews C. (2001) Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience* Jan;4(1):95-102.

O'Doherty J, Rolls ET, Francis S, Bowtell R, McGlone F, Kobal G, Renner B, Ahne G. (2000) Sensory-specific satiety-related olfactory activation of the human orbitofrontal cortex. *Neuroreport.* 11:399–403.

Ogawa, S., Lee, T. M., Kay, A. R. and Tank, D. W. (1990) Brain Magnetic Resonance Imaging with Contrast dependent on Blood Oxygenation. *Proc. Natl. Acad. Sci.* USA 87, 9868-9872.

Olds, J., Milner, P. (1954) Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *J.Comp. Physiolo. Psycholo.* 47, 419–427

Olds, J. (1956) Pleasure center in the brain. *Sci. Am.* 195: 105-16.

Olds, J. (1958) Self-stimulation of the brain. *Science* 127:315-24.

O'Reilly, R.C., Braver, T.S. & Cohen, J.D. (1999) A Biologically Based Computational Model of Working Memory. A. Miyake & P. Shah (Eds) Models of Working Memory: Mechanisms of Active Maintenance and Executive Control., 375-411, New York: Cambridge University Press.

Öngür, D., Price, J.L., (2000) The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb. Cortex* 10, 206–219.

Öngür, D., Ferry, A. T. & Price, J. L. (2003) Architectonic subdivision of the human orbital and medial prefrontal cortex.*J. Comp.Neurol* .460, 425–449

Packard, M.G. Knowlton, B.J. (2002) Learning and memory functions of the Basal Ganglia. Annu Rev Neurosci 25:563-593.

Padoa-Schioppa C (2011) Neurobiology of economic choice: a good-based model *Annual Review of Neuroscience* 34, 333:59

Padoa-Schioppa C and Assad JA (2006) Neurons in orbitofrontal cortex encode economic value. *Nature* 441, 223-226.

Pagnoni G, Zink CF, Montague PR, Berns GS. (2002) Activity in human ventral striatum locked to errors of reward prediction. *Nat. Neurosci.* 5:97–98

Palminteri, S. Boraud, T. Lafargue, G. Dubois, B. and Pessiglione, M. (2009) Brain Hemispheres Selectively Track the Expected Value of Contralateral Options *J. Neurosci.,* October 28, 29(43): 13465 - 13472.

Palmiter, R.D., (2007) Is dopamine a physiologically relevant mediator of feeding behavior? *Trends Neurosci.* 30, 375–381.

Parker A, Wilding E, Ackerman C (1998) The von Restorff effect in visual object recognition memory in humans and monkeys: the role of frontal/perirhinal interaction. J Cog Neurosci 10:691–703.

Parylak SL, Koob GF, Zorrilla EP. (2011) The dark side of food addiction. *Physiology & Behavior* Volume 104, Issue 1, 25 July, Pages 149-156

Pasupathy, A. Miller, E.K. (2005) Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature* 433:873– 876.

Pasquereau, B. Nadjar, A. Arkadir, D. Bezard, E. Goillandeau, M. Bioulac, B. Gross, C. E. Boraud, T. (2007) Shaping of motor responses by incentive values through the basal ganglia. *J Neurosci* 27: 1176–1183.

Pasquereau B, Nadjar A, Arkadir D, Bezard E, Goillandeau M, Bioulac B, Gross CE, Boraud T (2007) Shaping of motor responses by incentive values through the basal ganglia. *J Neurosci, 27:*1176-1183.

Passingham RE, Rowe JB, Sakai K (2005).Prefrontal cortex and attention to action. In Humphreys G, Riddoch MJ (Eds) Attention in Action, Psychology Press, 263-286

Passingham RE, Rowe JB (2002) Dorsal prefrontal cortex: maintenance in memory or attentional selection. In Principles of Frontal Lobe Function (Stuss DT, Knight RT), pp. 221-232, Oxford University Press, Oxford

Passingham RE, Toni I, Rushworth MF. (2000) Specialisation within the prefrontal cortex: the ventral prefrontal cortex and associative learning. *Exp Brain Res.* Jul;133(1):103-13.

Pauling L. and Coryell C. D., (1936) The magnetic properties and structure of hemoglobin, oxyhemoglobin and carbonmonoxy hemoglobin. Proc. Nayl. Acad. Sci. (USA) 22, 210-216

Pavlov, I. (1927). Conditioned reflexes. London: Oxford University Press.

Pearce, J. M., & Bouton, M. E. (2001). Theories of associative learning in animals. *Annual Review of Psychology, 52,* 111-139.

Pearce J.M. and Hall G. (1980) A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review,* 87:532-552,

Pearson JM, Heilbronner SR, Barack DL, Hayden BY, Platt ML (2011) Posterior cingulate cortex: Adapting behavior to a changing world. *Trends Cogn Sci* 15:143–151.

Peciña, S., Smith, K.S., & Berridge, K.C.(2006) Hedonic hotspots in the brain. *The Neuroscientist,* 12(6), 500-511.

Pelchat Ma. L. et al., (2004) Images of desire: food-craving activation during fMRI*NeuroImage* Volume 23, Issue 4 December, Pages 1486-1493

Pessiglione, M., Petrovic, P., Daunizeau, J., Palminteri, S., Dolan, R. J., & Frith, C. D. (2008). Subliminal instrumental conditioning demonstrated in the human brain. *Neuron,* 59(4), 561-7.

Pessiglione M., Seymour B., Flandin G., Dolan R.J. and Frith C.D. (2006) Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* . Aug 31;442 (7106):1042-5.

Pessoa L. & Engelmann J.B. (2010) Embedding reward signals into perception and cognition. Frontiers in Neuroscience. Front. Neurosci. 4:17.

Peters, J. Sethu, V.; Stefan S. (2003) Reinforcement Learning for Humanoid Robotics IEEE-RAS International Conference on Humanoid Robots.

Petersson K.M. et al. (1999) Dynamic changes in the functional anatomy of the human brain during recall of abstract designs related to practice *Neuropsychologia,* 37, pp. 567–587

Petrides M, Pandya DN (1994) Comparative cytoarchitectonic analysis of the human and the macaque frontal cortex. In: Handbook of neuropsychology, Vol. 9 (Boller F, Grafman J, eds), pp. 17–58. Amsterdam: Elsevier.

Pertovaara A, Martikainen IK, Hagelberg N, Mansikka H, Någren K, Hietala J, Scheinin H. (2004) Striatal dopamine D2/D3 receptor availability correlates with individual response characteristics to pain. *Eur J Neurosci.* Sep;20(6):1587-92.

Peyron, R., Laurent, B. & Garcia-Larrea, L. (2000). Functional imaging of brain responses to pain. A review and meta-analysis. *Neurophysiol. Clin.* 30, 263–288

Pitt, M A & Myung, I J. (2002). When a good fit can be bad . *Trends in Cognitive Sciences , 6(10)* , 421-425.

Platt, M.L. and Glimcher, P.W. (1999) Neural correlates of decision variables in parietal cortex. *Nature.* 400: 233-238.

Poldrack, R. A. & Packard, M. G. (2003) Competition among multiple memory systems: converging evidence from animal and human brain studies. *Neuropsychologia* 41, 245–251.

Poldrack, R. A., Sabb, F. W., Foerde, K., Tom, S. M., Asarnow, R. F., Bookheimer, S. Y., & Knowlton, B. J. (2005). The Neural Correlates of Motor Skill Automaticity.*The Journal of Neuroscience, 25, 22*, 5356-5364.

Posner, M. I., & Snyder, C. R. R. (1975) Attention and cognitive control. In R. Solso (Ed.*), Information processing and cognition:* The Loyola Symposium. Potomac, Md.: Erlbaum.

Postuma, R. B., & Dagher, A. (2006) Basal ganglia functional connectivity based on a meta-analysis of 126 positron emission tomography and functional magnetic resonance imaging publications. Cerebral cortex (New York, N.Y. : 1991), 16(10), 1508-21

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual review of neuroscience*, 13, 25-42.

Pluck GC, Brown RG (2002) Apathy in Parkinson's disease. *J Neurol Neurosurg Psychiatry;73:636-42.*

Prévost, C., McCabe, J. A., Jessup, R. K., Bossaerts, P., & O'Doherty, J. P. (2011). Differentiable contributions of human amygdalar subregions in the computations underlying reward and avoidance learning. *The European Journal of Neuroscience.* pp, 1-12

Prensa L, Parent A (2001) The nigrostriatal pathway in the rat: a single-axon study of the relationship between dorsal and ventral tier nigral neurons and the striosome/matrix striatal compartments. *J Neurosci* **21**:7247–7260.

Preuschoff, K., Quartz, S. R., & Bossaerts, P. (2008). Human insula activation reflects risk prediction errors as well as risk. *The Journal of Neuroscience*, *28*(11), 2745-52.

Preuschoff, K Bossaerts, P Quartz SR: (2006) Neural Differentiation of Expected Reward and Risk in Human Subcortical Structures. *Neuron.* Aug 3; 51(3):381-90.

Price, D. D. (2000) Psychological and Neural Mechanisms of the Affective Dimension of Pain.*Science*, *288*(5472), 1769-1772.

Prodoehl J, Yu H, Little DM, Abraham I, Vaillancourt DE. (2008) Region of interest template for the human basal ganglia: comparing EPI and standardized space approaches. *Neuroimage* 1;39(3):956-65.

Quintana J, Fuster JM. (1999) From perception to action: temporal integrative functions of prefrontal and parietal neurons. *Cereb Cortex.* Apr-May;9(3):213-21.

Raichle M.E. et al.(1999) Practice-related changes in human brain functional anatomy during non-motor learning *Cereb. Cortex*, 4 pp. 8–26

Rainville, P., Duncan, G. H., Price, D. D., Carrier, B. & Bushnell, M. C. (2010) Pain affect encoded in human anterior cingulate but not somatosensory cortex. *Science* 277,
968–971 (1997). 191–198

Ranganath, C. & D'Esposito, M. (2001) Medial temporal lobeactivity associated with active maintenance of novelinformation.*Neuron* 31, 865–873

Ranganath C. & Rainer G. (2003) Neural mechanisms for detecting and remembering novel events. *Nature Reviews Neuroscience* 4, 193-202

Rangel, A. Camerer, C. and Montague, R. (2008) A framework for studying the neurobiology of value-based decision-making. *Nature Reviews Neuroscience*, , 9:545-556

Rangel, A. Hare, T.A. (2010) Neural computations associated with goal-directed choice. *Current Opinion in Neurobiology,* 20:1-9.

Ratanaswasd, P., Dodd, W., Kawamura, K. & Noelle, D. C., (2005). Modular Behavior Control for a Cognitive Robot.12th International Conference on Advanced Robotics, Seattle, Washington, USA.

Redgrave, P., Prescott, T. J. & Gurney, K. The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience* 89, 1009–1023 (1999).

Redgrave, P. & Gurney, K. (2006) The short-latency dopamine signal: a role in discovering novel actions? , *Nature Reviews Neuroscience,* 7:967-975.

Redgrave P, Gurney K, Reynolds J (2007) What is reinforced by phasic dopamine signals? *Brain Res Rev.* ;58(2):322-39.

Reed, P. Mitchell, C. Nokes T. (1996) Intrinsic reinforcing properties of putatively neutral stimuli in an instrumental two-lever discrimination task. *Animal Learning and Behavior,* 24 , pp. 38–45

Rescorla,R.A.Wagner. A.R.(1972) A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. InClassical Conditioning II: Current Research and Theory. A.H.BlackW.F.Prokasy, Eds.:64–99. Appleton Crofts. New York.

Rescorla, R. A. (1972). Informational variables in Pavlovian conditioning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 6) (pp. 1-46). New York: Academic Press.

Rescorla, R. (2008), Scholarpedia, 3(3):2237.

Robinson TE, Berridge KC.(2003) Addiction. *Annu. Rev. Psychol.* 54:25–53

Robbins TW (2000) From arousal to cognition: the integrative position of the prefrontal cortex.*Prog Brain Res 126:469–483.*

Rodriguez, P., Aron, A. R., & Poldrack, R. A. (2005). Ventral striatal/nucleus-accumbens sensitivity to prediction errors during classification learning. *Human Brain Mapping, 27,* 306-13.

Romanelli P, Esposito V, Schaal DW, Heit G. (2005) Somatotopy in the basal ganglia: experimental and clinical evidence for segregated sensorimotor channels. *Brain Res Brain Res Rev.* Feb;48(1):112-28.

Rompré PP, Boye S. (1989) Localization of reward-relevant neurons in the pontine tegmentum: a movable electrode mapping study. Brain Res. ;496:295–302.

Rompré PP, Miliaressis E. (1985) Pontine and mesencephalic substrates of self-stimulation.Brain Res. ;359:246–259.

Roesch M.R., Singh T., Brown P.L., Mullins S.E., Schoenbaum G. (2009) Ventral striatal neurons encode the value of the chosen action in rats deciding between differently delayed or sized rewards. *J Neurosci.* Oct 21;29(42):13365-76.

Roelfsema, P. R., and Van Ooyen, A. (2005). Attention-gated reinforcement learning of internal representations for classification. Neural Computation 17: 2176-2214.

Routtenberg, (1978) The reward system of the brain. *Scientific American.* 154-164).

Rolls, E. T. (2005) Emotion explained. Oxford, Oxford University Press.

Rolls, E. T. (1999) The brain and emotion.Oxford, Oxford University Press.

Rolls, E. T., McCabe C. (2007) Enhanced affective brain representations of chocolate in cravers vs. non-cravers Volume 26, Issue 4, pages 1067–1076

Rolls E, Rolls B, Rowe E. (1983) Sensory-specific and motivation-specific satiety for the sight and taste of food and water in man. *Physiol. Behav.* 30:85–92.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986) Learning representations by back-propagating errors. *Nature,* 323, 533--536.

Rowe J, Frackowiak RSJ, Friston K, Passingham RE (2002) Attention to action: specific modulation of cortico-cortical interactions in humans. *NeuroImage* 17, 988-998

Rowe, J. Passingham RE (2001) Working memory for location and time: activity in prefrontal area 46 relates to selection rather than to maintenance in memory. *NeuroImage,* 14, 77-86

Rushworth, M. (2003). The left parietal and premotor cortices: motor attention and selection. *NeuroImage,* 20, S89-S100

Rushworth MFS, Buckley MJ, Gough PM, Alexander IH, Kyriazis D, McDonald KR, Passingham RE (2005) Attentional selection and action selection in the ventral and orbital prefrontal cortex. *J Neurosci* 25, 1128-1136

Rushworth, M. F., Paus, T. & Sipila, P. K. (2001) Attention systemsand the organization of the human parietal cortex. *J. Neurosci.*21, 5262–5271 (2001).

Russchen FT, Bakst I, Amaral DG, Price JL (1985) The amygdalostriatal projections in the monkey. An anterograde tracing study. *Brain Research* 329(1-2):241-257

Rutledge, R. B., Lazzaro, S. C., Lau, B., Myers, C. E., Gluck, M. A., & Glimcher, P. W. (2009). Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. *Journal of Neuroscience.* 29(48).

Sakagami M, et al. (2001) A code for behavioral inhibition on the basis of color, but not motion, in ventrolateral prefrontal cortex of macaque monkey. *Journal of Neuroscience.*;21:4801–4808.

Salamone JD, Correa M, Farrar A, Mingote SM. (2007) Effort-related functions of nucleus accumbens dopamine and associated forebrain circuits. Psychopharmacology. 191:461–482.

Salas, R., Baldwin, P., de Biasi, M., & Montague, P. R. (2010). BOLD Responses to Negative Reward Prediction Errors in Human Habenula. *Frontiers in human neuroscience*, *4*(May), 36.

Saling, L.L., & Phillips,J.G. (2007) Automatic behaviour: Efficient not mindless. *Brain Research Bulletin*, vol 73, Issues 1-3, pp 1-20.

Salmon, D.P. Butters, N. (1995) Neurobiology of skill and habit learning. *Curr. Opin. Neurobiol.,5,* 184–190

Samejima, K., & Doya, K. (2007). Multiple Representations of Belief States and Action Values in Corticobasal Ganglia Loops. *Annals Of The New York Academy Of Sciences,* 228, 213-228.

Samejima K, Doya K, Ueda Y, Kimura M. (2004). Estimating internal variables and parameters of a learning agent by a particle filter. Advances in Neural Information Processing Systems, 16, MIT Press.

Samejima K, Ueda Y, Doya K, Kimura M. (2005) Representation of action-specific reward values in the striatum. *Science. 25;310(5752):*1337-40.

Santesso, D. L., Evins, A. E., Frank, M. J., Schetter, E. C., Bogdan, R., & Pizzagalli, D. A. (2009). Single dose of a dopamine agonist impairs reinforcement learning in humans: evidence from event-related potentials and computational modeling of striatal-cortical function. *Hum Brain Mapp, 30*(7), 1963-1976.

Sawynok J, Esser MJ, Reid AR. (2001) Antidepressants as analgesics: an overview of central and peripheral mechanisms of action. J Psychiatry Neurosci 26: 21–29, Sato M., Hikosaka O. (2002) Role of primate substantia nigra pars reticulata in reward-oriented saccadic eye movement. *J Neurosci* 22:2363–2373.

Seger CA and Spiering BJ (2011) A critical review of habit learning and the basal ganglia. *Front. Syst. Neurosci.* 5:66.

Seymour B., (2003) Aversive Reinforcement Learning. Unpublished Phd Thesis. University College London

Seymour B., Singer, T., Dolan. R. (2007) The neurobiology of punishment. *Nature Reviews: Neuroscience.* Vol 8; pp300-311

Seymour B., O'Doherty J., Dayan P., Koltzenburg M., Jones A.K., Dolan RJ,. Friston KJ, Frackowiak RS. (2004). Temporal difference models describe higher-order learning in humans. *Nature,* 429, 664-7.

Schmidt L., et al., (2008) Disconnecting force from money: effects of basal ganglia damage on incentive motivation. *Brain*Volume 131,Issue 5 Pp. 1303-1310.

Schott B.H., et al., (2008) Mesolimbic functional magnetic resonance imaging activations during reward anticipation correlate with reward-related ventral striatal dopamine release. *J Neurosci.* Dec 24;28(52):14311-9.

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

Schultz, W. (2000) Multiple reward signals in the brain *Nat Rev Neurosci 1(3):* 199-207.

Schultz W. (2004) Neural coding of basic reward terms of animal learning theory, game theory, microeconomics, and behavioral ecology. *Curr Opin Neurobiol; 14:*139–147.

Schultz, W. (2006) Behavioral theories and the neurophysiology of reward. *Annu Rev Psychol* 57: 87-115.

Schultz, W. (2007a) Behavioral dopamine signals. *Trends in neurosciences,* 30(5), 203-10.

Schultz, W. (2007b) Multiple Dopamine Functions at Different Time Courses. *Annual Review of Neuroscience* Vol. 30: 259-288

Schultz, W. (2007c) Reward signals. Scholarpedia, 2(6):2184.

Schultz, W. et al. (2003) Changes in behavior-related neuronal activity in the striatum during learning. *Trends Neurosci.* 26, 321–328

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

Schultz, W. and Dickinson A. (2000) Neuronal coding of prediction errors. *Annu Rev Neurosci* 23: 473-500.

Schultz, W.; Romo, R; Ljungberg, T; Mirenowicz, J; Hollerman, J. R.; Dickinson, A. (1995) Reward-related signals carried by dopamine neurons. In Houk, James C. (Ed); Davis, Joel L. (Ed); Beiser, David G. (Ed) Models of information processing in the basal ganglia. Computational neuroscience. (pp. 233-248). Cambridge, MA, US: The MIT Press. xii, 382 pp.

Schwartz CE, Wright CI, Shin LM, Kagan J, Rauch SL. (2003) Inhibited and uninhibited infants "grown up": Adult amygdalar response to novelty. Science. 300:1952–1953.

Schweimer JV; Ungless MA. (2010). Phasic responses in dorsal-raphe serotonin neurons to noxious stimuli. *Neuroscience.* 171:1209-1215.

Schweighofer N., Tanaka S.C., Doya K. (2007) Serotonin and the evaluation of future rewards: Theory, experiments, and possible neural mechanisms. *Ann N Y Acad Sci* 1104: 289–300.

Siegrist J., Menrath I., Stöcker, T; Klein M., Kellermann T., Shah N.J., Zilles K., Schneider F. (2005) Differential brain activation according to chronic social reward frustration Neuroreport: Volume 16 - Issue 17 - pp 1899-1903

Siep N., et al., (2009) Hunger is the best spice: An fMRI study of the effects of attention, hunger and calorie content on food reward processing in the amygdala and orbitofrontal cortex *Behavioural Brain Research* Volume 198, Issue 1, 2 March Pages 149-158

Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J. & Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nature Reviews* Neuroscience, 12, 154-167.

Shafir, E., & Tversky, A. (1995) Decision making. In E.E. Smith and D.N. Osherson. (Eds.), An Invitation to Cognitive Science, Second Edition (Volume 3)

Shamay-Tsoory SG, Tibi-Elhanany Y, Aharon-Peretz J (2007) The green-eyed monster and malicious joy: the neuroanatomical bases of envy and gloating (schadenfreude*) Brain* 130*:1663–1678.*

Shen, W. Flajolet, M. Greengard, P. Surmeier D.J. (2008) Dichotomous dopaminergic control of striatal synaptic plasticity. *Science,* 321, pp. 848–851

Shiffrin, R. M., and Schneider, W. (1977) Controlled and automatic human information processing. II. Perceptual learning, automatic attending, and a general theory. *Psychol.* Rev. 84, 127–190.

Shima K, Tanji J (1998) Role for cingulate motor area cells in voluntary movement selection based on reward. S*cience* 282:1335–1338.

Shinar D, Meir M, Ben-Shoham I. (1998) How automatic is manual gear-shifting? *Hum Factor*s 40:647-654.

Smith AC, Frank LM, Wirth S, Yanike M, Hu D, Kubota Y, Graybiel AM, Suzuki WA, Brown EN. (2004) Dynamic analysis of learning in behavioral experiments. *J Neurosci* 24: 447–461.

Soubrié P. *(1986*) Reconciling the role of central serotonin neurones in human and animal behaviour. *Behav Brain Res 9:319–364.*

Sommer, C. (2004). Serotonin in pain and analgesia: actions in the periphery. *Mol Neurobiol*, 30(2):117–125.

Sugrue LP, Corrado GS, Newsome WT (2004) Matching behavior and the representation of value in the parietal cortex. *Science* 304:1782–1787.

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2005) Choosing the greater of two goods: neural currencies for valuation and decision making. *Neuroscience,* 6(May), 363-375.

Sutton, R.S. (1992). Gain adaptation beats least squares? Proceedings of the Seventh Yale Workshop on Adaptive and Learning Systems, pp. 161-166, Yale University, New Haven, CT.

Sutton, R.S. (1982 - unpublished draft). A theory of salience change dependent on the relationship between discrepancies on successive trials on which the stimulus is present.

Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: An introduction. Cambridge, MA: MIT Press.

Suri, R. E., & Schultz, W. (1998). Learning of sequential movements by neural network model with dopamine-like reinforcement signal. *Experimental Brain Research*, 121(3), 350-354.

Suri, R. E., & Schultz, W. (1999). A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. *Science,* 91(3), 871-890.

Stalnaker, T.A., Calhoon, G.G., Ogawa, M., Roesch, M.R., Schoenbaum, G., (2010) Neural correlates of stimulus–response and response–outcome associations in dorsolateral versus dorsomedial striatum. *Front. Integr. Neurosci.* 19, 4–12.

Starkstein SE, Mayberg HS, Preziosi TJ, Andrezejewski P, Leiguarda R, Robinson RG (1992) Reliability, validity, and clinical correlates of apathy in Parkinson's disease. *J Neuropsychiatry Clin Neurosci;4:134-9.*

Stefani MR, Groth K, Moghaddam B (2003) Glutamate receptors in the rat prefrontal cortex regulate set-shifting ability. *Behav Neurosci* 117:728–737.

Stephens, D. W. & Krebs, J. R. 1986 Foraging theory. Princeton, NJ: Princeton University Press.

Sternberg, S. (1969) The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), Attention and performance II. Acta Psychologica, 30, 276-315.

Sternberg, R. J. (1977) Intelligence, information processing,and analogical reasoning: The componential analysis of human abilities. Hillsdale, NJ: Erlbaum.

Stuphorn V. (2006) Neuroeconomics: Cardinal Utility in the Orbitofrontal Cortex? *Current Biology* 16 (15):R591-R593

Strick, P.L. Dum, R.P. Mushiake, H. (1995) Basal ganglia 'loops' with the .cerebral cortex, in: M. Kimura, A.M. Graybiel Eds. , Functions of the cortico-basal ganglia loop, Springer-Verlag, Tokyo, 1995, pp. 106 – 124.

Szameitat, A.J., Schubert, T., Muller, K., von Cramon, D.Y., (2002) Localization of executive functions in dual-task performance with fMRI. *J. Cogn. Neurosci.* 14, 1184 – 1199.

Takada, ·M.Tokuno, H. Nambu, A. Inase M. (1998) Corticostriatal projections from the somatic motor areas of the frontal cortex in the macaque monkey: segregation versus overlap of input zones from the primary motor cortex, the supplementary motor area, and the premotor cortex *Exp. Brain Res.,* 120 pp. 114–128

Takayuki Hosokawa, Keichiro Kato, Masato Inoue, Akichika MikamiNeurons in the macaque orbitofrontal cortex code relativepreference of both rewarding and aversive outcomes. *Neuroscience Research* 57 (2007) 434–445

Tanaka, S., Doya, K., Okada, G., Ueda, K., Okamoto, Y., Yamawaki, S. (2004). Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nature Neuroscience, 7(8),* 887-893.

Tanaka SC, Balleine BW, O'Doherty JP (2008) Calculating consequences: Brain systems that encode the causal effects of actions. *J Neurosci* 28: 6750–6755.

Thorn C. A., Atallah H., Howe M., Graybiel A. M. (2010) Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning. *Neuron* 66, 781–795

Thorndike, Edward Lee (1911), Animal Intelligence, Macmillan.

Thorndike, (1927) The Law of Effect The American Journal of Psychology Vol. 39, No. 1/4, Dec., 1927

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

Toni I, Krams M, Turner R, Passingham RE. (1998) The time course of changes during motor sequence learning: a whole-brain fMRI study. *Neuroimage*; 8: 50-61.

Toni I, Passingham RE. (1999) Prefrontal-basal ganglia pathways are involved in the learning of arbitrary visuomotor associations: a PET study. *Exp BrainRes;* 127: 19-32.

Tremblay, L. Hollerman, J.R. Schultz, W. (1998) Modifications of reward expectation-related neuronal activity during learning in primate striatum. *J Neurophysiol* 80:964 –977.

Tricomi, E., Balleine, B. & O'Doherty, J. (2009) A specific role for posterior dorsolateral striatum in human habit learning. Eur. J. Neurosci., 29, 2225–2232.

Tulving E, Markowitsch HJ, Kapur S, Habib R, Houle S (1994) Novelty encoding networks in the human brain: positron emission tomography data. *Neuroreport* 5:2525-2528.

Tulving E, Markowitsch HJ, Craik FE, Habib R, Houle S (1996) Novelty and familiarity activations in PET studies of memory encoding and retrieval. *Cereb Cortex* 6:71-79.

Turner, R., Le Bihan, D., Moonen, C.T., Despres, D., Frank, J., (1991) Echo-planar time course MRI of cat brain oxygenation changes. Magn. Reson. Med. 22, 159–166.

Tzourio-Mazoyer, N. Landeau, D. Papathanassiou, B. Crivello,F. Etard, O. Delcroix, N. Mazoyer B. and Joliot M. (2002) Automated Anatomical Labeling of activations in SPM using a Macroscopic Anatomical Parcellation of the MNI MRI single-subject brain *Neuroimage* 15 (1): 273–289.

Valentin, V.V., Dickinson, A. & O'Doherty, J.P. (2007) Determining the neural substrates of goal-directed learning in the human brain. J. Neurosci., 27, 4019–4026.

Van der Meer, M. A. A.. Redish A. D. (2010) Expectancies in decision making, reinforcement learning, and ventral striatum *Frontiers in Neuroscience* 4:6

Van der Meer, M. A. A.. Redish A. D. (2011) Ventral striatum: a critical look at models of learning and evaluation *Current Opinion in Neurobiology* 21(3):387-392.

Van Eimeren T, Ballanger B, Pellecchia G et al (2009) Dopamine agonists diminish value sensitivity of the orbitofrontal cortex: a trigger for pathological gambling in Parkinson's disease. *Neuropsychopharmacology* 34:2758–2766

Van Schouwenburg M, Aarts E, Cools R: (2010a) Dopaminergic modulation of cognitive control: distinct roles for the prefrontal cortex and the basal ganglia. Current pharmaceutical design 16(18):2026-2032.

Van Schouwenburg M. R., den Ouden H. E. M., Cools R. (2010b). The human basal ganglia modulate frontal-posterior connectivity during attention shifting. J. Neurosci. 30, 9910–9918.

Van Veen, V., Cohen, J. D., Botvinick, M. M., Stenger, V. A. & Carter, C. S. (2001). Anterior cingulate cortex, conflict monitoring, and levels of processing.NeuroImage, 14, 1302-1308.

Vicente AF, Bermudez MA, Romero Mdel C, Perez R, Gonzalez F. (2012) Putamen neurons process both sensory and motor information during a complex task. *Brain Res.* Jul 23;1466:70-81

Vo LTK, Walther DB, Kramer AF, Erickson KI, Boot WR, et al. (2011) Predicting Individuals' Learning Success from Patterns of Pre-Learning MRI Activity. PLoS ONE 6(1): e16093.

Vogt, B. A. in Cingulate neurobiology and disease (ed. Vogt, B. A.) Oxford Univ. Press, New York, 2009.

Vogt, B. A. (2005) Pain and emotion interactions in subregions of the cingulate gyrus. *Nature Rev. Neurosci.* 6, 533–544

Volkow ND, Wang G, Kollins SH, et al. (2009) Evaluating dopamine reward pathway in ADHD clinical implications. *JAMA*. 302 (10):1084-1091.

Volz, K. G., Schubotz, R. I., & von Cramon, D. Y. (2003). Predicting events of varying probability: Uncertainty investigated by fMRI. *NeuroImage*, *19*, 271–280.

Voon, V. Pessiglione, M. Brezing, C. Gallea, C. Fernandez, H.H. Dolan, R.J. Hallett M (2010) Mechanisms underlying dopamine-mediated reward ias in compulsive behaviors Neuron, 65 pp. 135–142

Waelti P, Dickinson A, Schultz W. (2001) Dopamine responses comply with basic assumptions of formal learning theory. *Nature* 412: 43-48

Wagner, A., et al., (2007) Altered Reward Processing in Women Recovered From Anorexia Nervosa *Am J Psychiatry* 164:1842-1849,

Waldschmidt, J.G, & Ashby, F.G. (2011). Cortical and striatal contributions to automaticity in information-integration categorization. *Neuroimage, 56,* 1791-1802.

Wallis, J. D. (2007a). Neuronal mechanisms in prefrontal cortex underlying adaptive choice behavior. *Annals of the New York Academy of Sciences*, 1121, 447-60.

Wallis, J.D. (2007b) Orbitofrontal cortex and its contribution to decision-making. *Annual Review of Neuroscience,* 30, 31-56.

Wallis, J. D., & Miller, E. K. (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *Neuroscience,* 18, 2069-2081.

Watanabe M. (1996) Reward expectancy in primate prefrontal neurons. *Nature* 382:629–632.

Watanabe, M. et al. (2002) Coding and monitoring of motivational context in the primate prefrontal cortex. *J. Neurosci.* 22: 2391–2400. 27.

Watanabe, K. and Hikosaka, O. (2005) Immediate changes in anticipatory activity of caudate neurons associated with reversal of position-reward contingency. *J Neurophysiol* 94:1879 –1887.

Watkins, C. J. C. H. (1989) Learning from delayed rewards. PhD thesis, University of Cambridge, Cambridge, England.

Watson, K.K., Shepherd, S.V., Platt, M.L. (2009) Neuroethology of Pleasure. MlKringelbach-Chapter 5 5:85-95.

Watson K.K, Platt ML, (2008) Neuroethology of reward and decision making. *Philos Trans R Soc Lond B Biol Sci.* Dec 12;363(1511):3825-35.

Watkins, C.J.C.H. & Dayan, P. (1992). Q-learning. Machine Learning, 8, 279-292.

Wickens, J. R., Horvitz, J. C., Costa, R. M., & Killcross, S. (2007) Dopaminergic Mechanisms in Actions and Habits. *Annals Of The New York Academy Of Sciences*, 27(31), 8181- 8183.

Wickens, J.R., Reynolds, J.N Hyland, B.I. (2003) Neural mechanisms of reward-related motor learning .*Curr. Opin. Neurobiol.*,13*,* 685–690.

Widrow, B. and Hoff, M. (1960) Adaptive switching circuits. In Western Electronic Show and Convention, Volume 4, pages 96-104

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology,* 8, 227–233.

Wilson FA, Rolls ET. (1993) The effects of stimulus novelty and familiarity on neuronal activity in the amygdala of monkeys performing recognition memory tasks. Experimental *Brain Research. ;*93:367–382.

Williams, S. M. & Goldman-Rakic, P. S. (1998) Widespread origin of the primate mesofrontal dopamine system. *Cereb. Cortex* 8, 321–345

Wittmann BC, Bunzeck N, Dolan RJ, Duzel E (2007) Anticipation of novelty recruits reward system and hippocampus while promoting recollection. *Neuroimage* 38:194-202.

Wunderlich, K. Rangel, A. O'Doherty J.P. (2009), Neural computations underlying action-based decision making in the human brain. *PNAS*, 106(40):17199:17204.

Wunderlich, K. Rangel, A. O'Doherty, J. (2010) Economic choices can be made using only stimulus values. *PNAS,* 107:10505:10510.

Wogar MA, Bradshaw CM, Szabadi E (1993)Effect of lesions of the ascending 5-hydroxytryptaminergic pathways on choice between delayed reinforcers. *Psychopharmacology* (Berl) *111:239–243.*

Wolfe, J. B. (1936). Effectiveness of token rewards for chimpanzees. *Comparative Psychology Monograph,* 1936, 11 (5, Series No. 60).

Wolpe J. (1950) Need-reduction, drive-reduction, and reinforcement: A neurophysiological view. *Psychological Review*, 57, 19-26.

Wood P.B., Schweinhardt P., Jaeger E, Dagher A, Hakyemez H, Rabiner EA, Bushnell MC, Chizh BA. (2007) Fibromyalgia patients show an abnormal dopamine response to pain. *Eur J Neurosci.* Jun;25(12):3576-82.

Wood P.B., Glabus M.F., Simpson R., Patterson J.C. (2009) Changes in gray matter density in fibromyalgia: correlation with dopamine metabolism**.** *J. Pain.* 2009 Jun;10(6):609-18.

Wörgötter, F. and Porr, B. (2008), Scholarpedia, 3(3):1448.

Wörgötter, F. and Porr, B. (2005) Temporal sequence learning, prediction and control, A review of different models and their relation to biological mechanisms. *Neural Comp.* 17:245-319

Wrase, J., Kahnt, T., Schlagenhauf, F., Beck, A., Cohen, M. X., Knutson, B., Heinz, A. (2007). Different neural systems adjust motor behavior in response to reward and punishment. *NeuroImage,* 36, 1253-1262.

Wright CI, Martis B, Schwartz CE, Shin LM, Fischer HH, McMullin K, Rauch SL. (2003) Novelty responses and differential effects of order in the amygdala, substantia innominata, and inferior temporal cortex. *Neuroimage.;*18:660–669.

Wright, C.I., Wedig, M.M., Williams, D., Rauch, S.L., Albert, M.S., (2006) Novel fearful faces activate the amygdala in healthy young and elderly adults. Neurobiol. Aging 27, 361–374.

Yacubian, J., Gläscher, J., Schroeder, K., Sommer, T., Braus, D. F., & Büchel, C. (2006).Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain.*The Journal of Neuroscience:26*(37), 9530-7

Yarrow K., Brown P, Krakauer J.W. (2009) Inside the brain of an elite athlete: the neural processes that support high achievement in sports *Nature Reviews Neuroscience.*10: 585-596.

Yeung, N., Botvinick, M. M. & Cohen, J. D. (2004). The neural basis of error-detection: Conflict monitoring and the error-related negativity. *Psychological Review,* 111,931-959.

Yin, H.H., and Knowlton, B.J. (2006) The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience* 7:464–476.

Yin, H. H., Knowlton, B. J. & Balleine, B. W. (2004) Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.* 19, 181–189.

Yin, H. H., Knowlton, B. J. & Balleine, B. W. (2006) Inactivation of dorsolateral striatum enhances sensitivity to changes in the action–outcome contingency in instrumental conditioning. *Behav. Brain Res.* 166, 189–196

Yin, H. H., Mulcare, S. P., Hilário, M. R., Clouse, E., Holloway, T., Davis, M. I., Hansson, A. C., Lovinger, D. M., and Costa, R. M. (2009). Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. *Nat Neurosci.* 12, 333–341.

Yin, H. H., Ostlund, S. B., Knowlton, B. J. & Balleine, B. W. (2005) The role of the dorsomedial striatum in instrumental conditioning. *Eur. J. Neurosci.* 22, 513–523

Yu, AJ & Dayan, P (2005) Uncertainty, neuromodulation, and attention. *Neuron* 46 681-692.

Yuhong J. (2004) Resolving dual-task interference: an fMRI study. *NeuroImage* Volume 22, Issue 2, Pages 748–754

Zak, P. J. (2007) The neuroeconomics of trust. Ch 2. in Renaissance in Behavioral Economics. R. Frantz, Ed. Routledge

Zald, D. H., Lee, J. T., Fluegel, K., & Pardo, J. V. (1998) Aversive gustatory stimulation activates limbic circuits in humans. *Brain.* 121, 1143-1154.

Zink CF, Pagnoni G, Martin ME, Dhamala M, Berns GS. (2003) Human striatal response to salient nonrewarding stimuli. *J Neurosci* 23:8092– 8097.

Zink CF, Pagnoni G, Martin-Skurski ME, Chappelow JC, Berns GS (2004): Human striatal responses to monetary reward depend on saliency. *Neuron* 42:509 –517.

Appendix A

Cliff Walking Task in Q-learning MATLAB Code

```
clear;

stateNum = 40;

actionNum = 5;

      up    = 1;

      down  = 2;

      left  = 3;

      right = 4;

      stay = 5; %stay if you can,

Q = zeros(40,5);

r = zeros(40,5);

      r(1,up) = -1;  r(1,down) =   -1;   r(1,left) =   -1; r(1,right) =   -100;  r(1,stay) =
      -1;

      r(2,up) = -1;   r(2,down) =  -1;   r(2,left) =   -1; r(2,right) =   -100;  r(2,stay) =
      -100;

      r(3,up) = -1;  r(3,down) =   -1;   r(3,left) =   -1; r(3,right) =   -100;  r(3,stay) =
      -100;

      r(4,up) = -1;   r(4,down) =  -1;   r(4,left) =   -1; r(4,right) =   -100;  r(4,stay) =
      -100;

      r(5,up) = -1;   r(5,down) =  -1;   r(5,left) =   -1; r(5,right) =   -100;  r(5,stay) =
      -100;

      r(6,up) = -1;   r(6,down) =  -1;   r(6,left) =   -1; r(6,right) =   -100;  r(6,stay) =
      -100;

      r(7,up) = -1;   r(7,down) =  -1;   r(7,left) =   -1; r(7,right) =   -100;   r(7,stay) =
      -100;

      r(8,up) = -1;   r(8,down) =  -1;   r(8,left) =   -1; r(8,right) =   100;   r(8,stay) =
      100;

      r(9,up) = -1;   r(9,down) =  -1;   r(9,left) =   -1; r(9,right) =   -1;    r(9,stay) =
      -1;

      r(10,up) = -1;   r(10,down) =  -1; r(10,left) =   -1; r(10,right) =   -1;  r(10,stay) =
      -1;

      r(11,up) = -1;   r(11,down) =  -1; r(11,left) =   -1; r(11,right) =   -1;  r(11,stay) =
      -1;

      r(12,up) = -1;   r(12,down) =  -1; r(12,left) =   -1; r(12,right) =   -1;  r(12,stay) =
      -1;

      r(13,up) = -1;   r(13,down) =  -1; r(13,left) =   -1; r(13,right) =   -1;  r(13,stay) =
      -1;
```

```
r(14,up) = -1;   r(14,down) =  -1; r(14,left) =   -1; r(14,right) =   -1;  r(14,stay) =
-1;

r(15,up) = -1;   r(15,down) =  -1; r(15,left) =   -1; r(15,right) =   -1;  r(15,stay) =
-1;

r(16,up) = -1;   r(16,down) =  100; r(16,left) =  -1; r(16,right) =   -1;  r(16,stay) =
-1;

r(17,up) = -1;   r(17,down) =  -1; r(17,left) =   -1; r(17,right) =   -1;  r(17,stay) =
-1;

r(18,up) = -1;   r(18,down) =  -1; r(18,left) =   -1; r(18,right) =   -1;  r(18,stay) =
-1;

r(19,up) = -1;   r(19,down) =  -1; r(19,left) =   -1; r(19,right) =   -1;  r(19,stay) =
-1;

r(20,up) = -1;   r(20,down) =  -1; r(20,left) =   -1; r(20,right) =   -1;  r(20,stay) =
-1;

r(21,up) = -1;   r(21,down) =  -1; r(21,left) =   -1; r(21,right) =   -1;  r(21,stay) =
-1;

r(22,up) = -1;   r(22,down) =  -1; r(22,left) =   -1; r(22,right) =   -1;  r(22,stay) =
-1;

r(23,up) = -1;   r(23,down) =  -1; r(23,left) =   -1; r(23,right) =   -1;  r(23,stay) =
-1;

r(24,up) = -1;   r(24,down) =  -1; r(24,left) =   -1; r(24,right) =   -1;  r(24,stay) =
-1;

r(25,up) = -1;   r(25,down) =  -1; r(25,left) =   -1; r(25,right) =   -1;  r(25,stay) =
-1;

r(26,up) = -1;   r(26,down) =  -1; r(26,left) =   -1; r(26,right) =   -1;  r(26,stay) =
-1;

r(27,up) = -1;   r(27,down) =  -1; r(27,left) =   -1; r(27,right) =   -1;  r(27,stay) =
-1;

r(28,up) = -1;   r(28,down) =  -1; r(28,left) =   -1; r(28,right) =   -1;  r(28,stay) =
-1;

r(29,up) = -1;   r(29,down) =  -1; r(29,left) =   -1; r(29,right) =   -1;  r(29,stay) =
-1;

r(30,up) = -1;   r(30,down) =  -1; r(30,left) =   -1; r(30,right) =   -1;  r(30,stay) =
-1;

r(31,up) = -1;   r(31,down) =  -1; r(31,left) =   -1; r(31,right) =   -1;  r(31,stay) =
-1;

r(32,up) = -1;   r(32,down) =  -1; r(32,left) =   -1; r(32,right) =   -1;  r(32,stay) =
-1;

r(33,up) = -1;   r(33,down) =  -1; r(33,left) =   -1; r(33,right) =   -1;  r(33,stay) =
-1;

r(34,up) = -1;   r(34,down) =  -1; r(34,left) =   -1; r(34,right) =   -1;  r(34,stay) =
-1;

r(35,up) = -1;   r(35,down) =  -1; r(35,left) =   -1; r(35,right) =   -1;  r(35,stay) =
-1;

r(36,up) = -1;   r(36,down) =  -1; r(36,left) =   -1; r(36,right) =   -1;  r(36,stay) =
-1;

r(37,up) = -1;   r(37,down) =  -1; r(37,left) =   -1; r(37,right) =   -1;  r(37,stay) =
-1;
```

```
r(38,up) = -1;   r(38,down) =  -1; r(38,left) =   -1; r(38,right) =   -1;  r(38,stay) =
-1;

r(39,up) = -1;   r(39,down) =  -1; r(39,left) =   -1; r(39,right) =   -1;  r(39,stay) =
-1;

r(40,up) = -1;   r(40,down) =  -1; r(40,left) =   -1; r(40,right) =   -1;  r(40,stay) =
-1;
```

Set up transition functions from state to action

```
%%% For Example if I am in state 1 and

d(1,up) = 9;    d(1,down) = 1;  d(1,left) = 1;   d(1,right) = 2;   d(1,stay) = 1;

d(2,up) = 10;   d(2,down) = 2;  d(2,left) = 1;   d(2,right) = 3;   d(2,stay) = 2;

d(3,up) = 11;   d(3,down) = 3;  d(3,left) = 2;   d(3,right) = 4;   d(3,stay) = 3;

d(4,up) = 12;   d(4,down) = 4;  d(4,left) = 3;   d(4,right) = 5;   d(4,stay) = 4;

d(5,up) = 13;   d(5,down) = 5;  d(5,left) = 4;   d(5,right) = 6;   d(5,stay) = 5;

d(6,up) = 14;   d(6,down) = 6;  d(6,left) = 5;   d(6,right) = 7;   d(6,stay) = 6;

d(7,up) = 15;   d(7,down) = 7;  d(7,left) = 6;   d(7,right) = 8;   d(7,stay) = 7;

d(8,up) = 16;   d(8,down) = 8;  d(8,left) = 7;   d(8,right) = 8;   d(8,stay) = 8;


d(9,up)  = 17;   d(9,down)  = 1;  d(9,left)  = 9;    d(9,right)  = 10;   d(9,stay) = 9;

d(10,up) = 18;   d(10,down) = 2;  d(10,left) = 9;    d(10,right) = 11;   d(10,stay) =
10;

d(11,up) = 19;   d(11,down) = 3;  d(11,left) = 10;   d(11,right) = 12;   d(11,stay) =
11;

d(12,up) = 20;   d(12,down) = 4;  d(12,left) = 11;   d(12,right) = 13;   d(12,stay) =
12;

d(13,up) = 21;   d(13,down) = 5;  d(13,left) = 12;   d(13,right) = 14;   d(13,stay) =
13;

d(14,up) = 22;   d(14,down) = 6;  d(14,left) = 13;   d(14,right) = 15;   d(14,stay) =
14;

d(15,up) = 23;   d(15,down) = 7;  d(15,left) = 14;   d(15,right) = 16;   d(15,stay) =
15;

d(16,up) = 24;   d(16,down) = 8;  d(16,left) = 15;   d(16,right) = 16;   d(16,stay) =
16;


d(17,up) = 25;   d(17,down) = 9;   d(17,left) = 17;   d(17,right) = 18;  d(17,stay) =
17;

d(18,up) = 26;   d(18,down) = 10;  d(18,left) = 17;   d(18,right) = 19;  d(18,stay) =
18;

d(19,up) = 27;   d(19,down) = 11;  d(19,left) = 18;   d(19,right) = 20;  d(19,stay) =
19;

d(20,up) = 28;   d(20,down) = 12;  d(20,left) = 19;   d(20,right) = 21;   d(20,stay) =
20;

d(21,up) = 29;   d(21,down) = 13;  d(21,left) = 20;   d(21,right) = 22;   d(21,stay) =
21;
```

```
d(22,up) = 30;   d(22,down) = 14;  d(22,left) = 21;   d(22,right) = 23;   d(22,stay) =
22;

d(23,up) = 31;   d(23,down) = 15;  d(23,left) = 22;   d(23,right) = 24;   d(23,stay) =
23;

d(24,up) = 32;   d(24,down) = 16;  d(24,left) = 23;   d(24,right) = 24;   d(24,stay) =
24;



d(25,up) = 33;   d(25,down) = 17;  d(25,left) = 25;   d(25,right) = 26;   d(25,stay) =
25;

d(26,up) = 34;   d(26,down) = 18;  d(26,left) = 25;   d(26,right) = 27;   d(26,stay) =
26;

d(27,up) = 35;   d(27,down) = 19;  d(27,left) = 26;   d(27,right) = 28;   d(27,stay) =
27;

d(28,up) = 36;   d(28,down) = 20;  d(28,left) = 27;   d(28,right) = 29;   d(28,stay) =
28;

d(29,up) = 37;   d(29,down) = 21;  d(29,left) = 38;   d(29,right) = 30;   d(29,stay) =
29;

d(30,up) = 38;   d(30,down) = 22;  d(30,left) = 29;   d(30,right) = 31;   d(30,stay) =
30;

d(31,up) = 39;   d(31,down) = 23;  d(31,left) = 30;   d(31,right) = 32;   d(31,stay) =
31;

d(32,up) = 40;   d(32,down) = 24;   d(32,left) = 31;  d(32,right) = 32;   d(32,stay) =
32;



d(33,up) = 33;   d(33,down) = 25;  d(33,left) = 33;   d(33,right) = 34;   d(33,stay) =
33;

d(34,up) = 34;   d(34,down) = 26;  d(34,left) = 33;   d(34,right) = 35;   d(34,stay) =
34;

d(35,up) = 35;   d(35,down) = 27;  d(35,left) = 34;   d(35,right) = 36;   d(35,stay) =
35;

d(36,up) = 36;   d(36,down) = 28;  d(36,left) = 35;   d(36,right) = 37;   d(36,stay) =
36;

d(37,up) = 37;   d(37,down) = 29;  d(37,left) = 36;   d(37,right) = 38;   d(37,stay) =
37;

d(38,up) = 38;   d(38,down) = 30;  d(38,left) = 37;   d(38,right) = 39;   d(38,stay) =
38;

d(39,up) = 39;   d(39,down) = 31;  d(39,left) = 38;   d(39,right) = 40;   d(39,stay) =
39;

d(40,up) = 40;   d(40,down) = 32;  d(40,left) = 39;   d(40,right) = 40;   d(40,stay) =
40;





%%%how many learning episodes it take to fully learn the reward matrix

episodes = 10000;
```

```
%%%discounting parameter
gamma = 0.99;

%%%learning rate
alpha = 0.9;

%%%random action selection parameter, exploration vs exploitation
epsilon= 0.1;

%%%%%define initial counting parameter for each state
statecounting=zeros(40,1);

%%%%%define initial counting parameter for sum reward collection
sumreward=zeros(10000,1);

for i=1:episodes,

   Qold = Q;
   % Start from initial state

   s = 1;


   while s~=8, % Repeat until terminal state is reached.
          randomnumber1= unifrnd(0, 1);

        if (randomnumber1<epsilon) %
                a1= RandomPermutation([up down left right stay]);    % randomize the
possible action

                  a=a1(1);        % select the initial randomized action
          else


                  a1=find(max(Q(s,1:5))==Q(s,1:5));
                   a=a1(1);
```

```
         end;
```

Receive reward

```
      reward = r(s,a);
```

Determine new s' (sNew) state depending on the previous action

```
      sPrime = d(s,a);
```

Update Q(s,a) function

```
     Q(s,a) = Q(s,a) + alpha*(reward + gamma*max(Q(sPrime,[up down left right stay])) -
Q(s,a));


     %p(i)=reward + gamma*max(Q(sPrime,[up down left right stay])) - Q(s,a);


      s = sPrime;



      %%counts how many times each state is visited

   statecounting(s,1)= statecounting(s,1)+1;

   if s == 2


     elseif s == 3

         elseif s == 4


   elseif s == 5


   elseif s == 6


     elseif s == 7

     break;

   end;


   end;


      %counts average reward collected by the agent
      sumreward(i,1)= sum(sum(Q));
end;
   rew1=sumreward;
%%%%% plot learned Q table
figure(1)
imagesc(Q);
   colormap(hot);
 hold

%%%%plot how many times each state is visited
figure(2)
statecount(5,:)= statecounting(1:8)';
  statecount(4,:)= statecounting(9:16)';
  statecount(3,:)= statecounting(17:24)';
  statecount(2,:)= statecounting(25:32)';
  statecount(1,:)= statecounting(33:40)';
  imagesc(statecount);
  colormap(hot);
  hold
  %%%%plot sum of reward
%   %%%%plot how many times each state is visited
figure(3)
plot(sumreward);
hold



     savefile = 'dataqlearning.mat';


    save(savefile, 'rew1');
%


%   figure(3)
%
%     qvalue1=Q(1:8,1:5)';
%       qvalue2=Q(9:16,1:5)';
%   qvalue3=Q(17:24,1:5)';
%     qvalue4= Q(25:32,1:5)';
%    qvalue5= Q(33:40,1:5)';
%
%
%   greatq=cat(1,qvalue5,qvalue4,qvalue3,qvalue2,qvalue1);
%   imagesc(greatq);
%   colormap(hot);
%   hold
```

Appendix B

Parameter Estimation for Experiment 1, MATLAB Code

```
% Parameter estimation for Experiment 1



LL=zeros(100,100);  % initialise vectors for the loglikelihood of alpha x beta


% loop through all combinations of alpha and beta
for i=1:100
    alpha=i/100;  % thus 0 <= alpha <= 1
    for j=1:100
        beta=j/100; % thus 0 <= beta <= 1

        selectdata=data(data(:,3)==1,:); % only look at gain trials
        %selectdata=data(data(:,3)==3,:); % only look at loss trials

        qA=0;  % intialise q values to zero
        qB=0;

        proba=[];  % record the probabilities corresponding to the chosen actions
        error=[]; % record prediction errors (outcome - expectation)

        % loop through trials
        for t=1:length(selectdata)

            % calculate probabilities of choice A and choice B using softmax function
            pA=exp((qA/beta))/(exp((qA/beta))+exp((qB/beta)));
            pB=exp((qB/beta))/(exp((qA/beta))+exp((qB/beta)));

            reward=(selectdata(t,8)==1); % for gain trials
            %reward=-(selectdata(t,8)==-1); % for loss trials
            if selectdata(t,7)==1 % correct choice, chose A
                proba(t)=log(pA);  % note p(chosen action)
                % update q values, arbitrary reward value set at 1
                error(t)=reward-qA;
                qA=qA+alpha*error(t);

            else % incorrect choice, chose B
                proba(t)=log(pB); % note p(chosen action)
                % update q values, arbitrary reward value set at 1
                error(t)=reward-qB;
                qB=qB+alpha*error(t);
            end
        end

        % update likelihhod array
        LL(i,j)=LL(i,j)+sum(proba);
    end
end

[alpha,beta]=find(LL==max(max(LL))); % find optimal values of alpha and beta
alpha=alpha/100; % thus 0 <= alpha <= 1
beta=beta/100; % thus 0 <= beta <= 1
imagesc(LL),colorbar; % display log likelihood array
```

Appendix C

Parameter Estimation for Experiment 2, Adaptive Learning Rate MATLAB Code

```
function[alpha1,beta,teta]=model_estimate(choice,r)


%LL=zeros(100,100);  % initialise vectors for the loglikelihood of alpha x beta

LL=zeros(100,100,100);

% loop through all combinations of alpha and beta

% for i=1:100

%     alpha1=i/100;  % thus 0 <= alpha <= 1

%     for j=1:100

%         beta1=j/100; % thus 0 <= beta <= 1

for i=1:100

    alpha1=i/100  % thus 0 <= alpha <= 1

    for j=1:100

        beta=j/100; % thus 0 <= beta <= 1

         for k =1:100

         teta =  k/100;



        qA=0;  % intialise q values to zero

        qB=0;


        proba=[];  % record the probabilities corresponding to the chosen actions

        delta=[]; % record prediction errors (outcome - expectation)

       learning_rate(1)= alpha1;

        % loop through trials

        for t=1:length(choice)


            % calculate probabilities of choice A and choice B using softmax function

            pA=exp((qA/beta))/(exp((qA/beta))+exp((qB/beta)));
```

```
 pB=1-pA;
 reward=r(t); % for gain trials
 %reward=-(selectdata(t,8)==-1); % for loss trials

 if choice(t)==1 % correct choice, chose A
     proba(t)=pA;  % note p(chosen action)
     % update q values, arbitrary reward value set at 1
    delta(t)=reward-qA;
     qA=qA+learning_rate(t)*delta(t);

if delta(t) > 0
      learning_rate(t+1)= learning_rate(t)-learning_rate(t)*teta;
      end;
      if delta(t) < 0
       learning_rate(t+1)= learning_rate(t)+learning_rate(t)*teta;
      end;
      if delta(t) == 0
      learning_rate(t+1)= learning_rate(t);
      end
 else % incorrect choice, chose B
     proba(t)=pB; % note p(chosen action)
     % update q values, arbitrary reward value set at 1
     delta(t)=reward-qB;
     qB=qB+alpha1*delta(t);

if delta(t) > 0
      learning_rate(t+1)= learning_rate(t)-learning_rate(t)*teta;
      end;
      if delta(t) < 0
       learning_rate(t+1)= learning_rate(t)+learning_rate(t)*teta;
      end;
      if delta(t) == 0
      learning_rate(t+1)= learning_rate(t);
      end
 end
```

```
        end


        % update likelihhod array
      %  LL(i,j)= prod(proba);
  LL(i,j,k)=prod(proba);
    end


end
end
[maxLL, index] = max (LL(:))
[alpha1, beta, teta] = ind2sub (size(LL), index) %


%[alpha1,beta1]=find(LL==max(max(LL))); % find optimal values of alpha and beta
alpha1=alpha1/100; % thus 0 <= alpha <= 1
beta=beta/100; % thus 0 <= beta <= 1
teta=teta/100;
```